

## ARTICLE

# Evaluation of model-based bioequivalence approach for single sample pharmacokinetic studies

Coralie Tardivon<sup>1,2</sup>  | Florence Loingeville<sup>1,3</sup>  | Mark Donnelly<sup>4</sup>  | Kairui Feng<sup>4</sup> |  
Wanjie Sun<sup>5</sup> | Guoying Sun<sup>5</sup> | Stella Grosser<sup>5</sup> | Liang Zhao<sup>4</sup>  | Lanyan Fang<sup>4</sup> |  
France Mentré<sup>1</sup>  | Julie Bertrand<sup>1</sup> 

<sup>1</sup>INSERM, IAME, Université de Paris, Paris, France

<sup>2</sup>Département Epidémiologie Biostatistiques et Recherche Clinique, AP-HP, Hôpital Bichat, Paris, France

<sup>3</sup>METRICS: Evaluation of Health Technologies and Medical Practices, University of Lille, CHU Lille, ULR 2694, Lille, France

<sup>4</sup>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, Maryland 20993, USA

<sup>5</sup>Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland 20993, USA

## Correspondence

Julie Bertrand, INSERM, IAME, Université de Paris, Paris F-75018, France.

Email: [julie.bertrand@inserm.fr](mailto:julie.bertrand@inserm.fr)

## Abstract

In a traditional pharmacokinetic (PK) bioequivalence (BE) study, a two-way crossover study is conducted, PK parameters (namely the area under the time-concentration curve [AUC] and the maximal concentration [ $C_{\max}$ ]) are obtained by noncompartmental analysis (NCA), and the BE analysis is performed using the two one-sided test (TOST) method. For ophthalmic drugs, however, only one sample of aqueous humor, in one eye, per eye can be obtained in each patient, which precludes the traditional BE analysis. To circumvent this issue, the U.S. Food and Drug Administration (FDA) has proposed an approach coupling NCA with either parametric or nonparametric bootstrap (NCA bootstrap). The model-based TOST (MB-TOST) has previously been proposed and evaluated successfully for various settings of sparse PK BE studies. In this paper, we evaluate, via simulations, MB-TOST in the specific setting of single sample PK BE study and compare its performance to NCA bootstrap. We performed BE study simulations using a published PK model and parameter values and evaluated multiple scenarios, including study design (parallel or crossover), sampling times (5 or 10 spread across the dosing interval), and geometric mean ratio (of 0.8, 0.9, 1, and 1.25). Using the simulated structural PK model, MB-TOST performed similarly to NCA bootstrap for AUC. For  $C_{\max}$ , the latter tended to be conservative and less powerful. Our research suggests that MB-TOST may be considered as an alternative BE approach for single sample PK studies, provided that the PK model is correctly specified and the test drug has the same structural model as the reference drug.

## Study Highlights

### WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

In bioequivalence (BE) studies for generic drug development of ophthalmic products, one sample of aqueous humor is taken at a certain timepoint per eye for measurement of drug concentration. Currently, the U.S. Food and Drug Administration (FDA) recommends a nonparametric or parametric bootstrap

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

noncompartmental (NCA) approach in the BE evaluation of these products. Recent research has focused on nonlinear mixed effect models which have been evaluated for sparse pharmacokinetic (PK) BE studies.

#### **WHAT QUESTION DID THIS STUDY ADDRESS?**

The present work compares a model-based (MB) BE approach to the nonparametric NCA for PK BE studies with one sampling timepoint per subject and per occasion.

#### **WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?**

Our simulations show that the MB BE approach controlled type I errors, and demonstrated a higher power for maximal concentration ( $C_{\max}$ ) compared to nonparametric NCA with parallel study designs.

#### **HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?**

Results of our research suggest that an MB approach may potentially serve as an alternative method to evaluate BE for generic ophthalmic drug products. However, further research and assessment are needed.

## **INTRODUCTION**

Pharmacokinetic (PK) bioequivalence (BE) studies are key to the development of generic (or test) drugs. Traditionally, a two-treatment, two-sequence, two-period (two-way) crossover study is conducted and the two one-sided test (TOST) is performed on estimates of the area under the concentration-time curve (AUC) and the maximal concentration ( $C_{\max}$ ) obtained by noncompartmental analysis (NCA). In these BE studies, the AUC can be computed using the trapezoidal method and linear extrapolation, if 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.<sup>1</sup>

In PK BE studies for ophthalmic drug products, typically only one single sample of aqueous humor is collected from one eye of each patient, at one assigned sampling timepoint, to measure drug exposure. Recommendations from the U.S. Food and Drug Administration (FDA) on BE study designs for ophthalmic drug products have been previously described.<sup>2,3</sup> A parallel design is typically used when only one sample is collected per patient. A crossover design may be used when using patients with bilateral disease (i.e., on both eyes), enabling the collection of two samples (i.e., one eye per period).

As such, PK BE studies of ophthalmic drug products present unique statistical challenges. Certainly, classical NCA is not feasible because it is not possible to calculate individual AUC values from a single concentration. To circumvent this issue, a nonparametric bootstrap NCA has been proposed by Shen et al.<sup>4</sup> This method involves resampling of the subjects, at each sampling time in both reference and test treatment groups, to generate nonparametric bootstrap datasets. Next, the arithmetic mean

concentration is computed at each sampling time.<sup>3</sup> Then, AUC and  $C_{\max}$  are computed by NCA from the arithmetic mean concentration in each treatment group for each bootstrap data set to assess the relevant BE criteria. The product-specific guidance (PSG) for ophthalmic drug products generally recommends to use this method or a parametric method for BE analysis in PK studies.<sup>5</sup>

Nonlinear mixed effects modeling (NLMEM) is a powerful tool used to describe the PK properties of a drug and may be used to support a lower number of subjects or reduced sampling in PK studies.<sup>6</sup> In the last 15 years, NLMEM has been demonstrated to be a useful tool in the design and analysis of PK BE studies.<sup>7</sup> For example, Panhard and Mentré<sup>8</sup> proposed to perform a TOST on empirical Bayes estimates (EBEs) of individual AUC and  $C_{\max}$  or to analyze the data globally and perform two Wald tests on the model-based treatment effect estimates (MB-TOST). However, NLMEM requires specific estimation algorithms which may present certain challenges and concerns. For instance, Panhard and Mentré used the first-order conditional estimates (FOCE) algorithm<sup>9</sup> which presents convergence issues that can be avoided with the more exact stochastic approximation expectation maximization (SAEM) algorithm.<sup>10</sup> Therefore, Dubois et al.<sup>11</sup> evaluated the tests on EBE using SAEM, exploring various scenarios, and found the latter are not recommended when the shrinkage is above 20%. Furthermore, they evaluated the MB-TOST using SAEM and identified an inflation of the test type I error rate due to departure from the asymptotic conditions.<sup>12</sup> Consequently, Loingeville et al.<sup>13</sup> proposed alternative calculations of the standard errors (SEs) and showed that MB-TOST using non-asymptotic SE controls type I error rate on sparse PK BE studies with small sample sizes. In PK BE studies with large sample sizes, MB-TOST using asymptotic SE derived from

the observed Fisher information matrix (FIM) performed satisfactorily.<sup>14</sup> Yet, Möllenhoff et al. (2020) proposed the bioequivalence optimal test (BOT) which was shown to be more powerful while controlling type I error using either a model or NCA method to fit the concentrations.<sup>15</sup> Further, their simulation study highlighted the overly conservative type I error rate of the TOST method for drugs with large PK variability. Nonetheless, currently, the Agency recommends the reference-scaled average BE approach for highly variable drugs.<sup>1,16</sup>

Our objective in the present work is to compare an MB approach to the FDA recommended NCA nonparametric bootstrap approach for single point designs.

## METHODS

### Two one-sided tests

For most drug products, BE studies are focused on the release of the drug substance from the drug product into the systemic circulation. In these BE studies, the rate and extent of drug absorption of the test (*T*) product is compared to the reference (*R*) product using AUC and  $C_{\max}$  metrics, respectively. For the drug exposure, one can estimate the AUC from time zero until the last sampling time ( $AUC_{0-t_{\text{last}}}$ )  $t_{\text{last}}$  or the AUC from time zero until infinity ( $AUC_{0-\infty}$ ).

Let  $\beta^T$  be the treatment effect (i.e., the difference between  $\mu_T$  and  $\mu_R$  [ $\beta^T = \mu_T - \mu_R$ ]) the average means of the test and reference products for  $\log(\text{AUC})$  or  $\log(C_{\max})$ , then the null hypothesis for BE testing is:

$$H_0: \beta^T \geq \delta \text{ or } \beta^T \leq -\delta \quad (1)$$

with  $\delta$  a given threshold

The TOST relies on the decomposition of the null hypothesis stated in Equation 1 in two one-sided hypotheses  $H_{0,-\delta}: \{\beta^T \leq -\delta\}$  and  $H_{0,\delta}: \{\beta^T \geq \delta\}$ .<sup>17</sup> Thereby the null hypothesis is rejected if

$$\frac{\hat{\beta}^T + \delta}{\text{SE}(\hat{\beta}^T)} \geq u_{1-\alpha} \quad \text{and} \quad \frac{\hat{\beta}^T - \delta}{\text{SE}(\hat{\beta}^T)} \leq -u_{1-\alpha} \quad (2)$$

where  $\hat{\beta}^T$  and  $\text{SE}(\hat{\beta}^T)$  are the  $\beta^T$  estimate and its standard error, and  $u_{1-\alpha}$  is the  $1 - \alpha$  quantile of a reference distribution.

This method is equivalent to constructing a  $(1 - 2\alpha)$ -confidence interval (CI) for  $\hat{\beta}^T$  and concluding BE if it is completely contained in the equivalence interval  $[-\delta; \delta]$ . In conventional BE analysis, regulatory authorities set the threshold  $\delta = \log(1.25) = -\log(0.8)$  and the

significance level  $\alpha = 0.05$ .<sup>18</sup> The BE acceptance criteria is for the 90% CI around the geometric mean ratio (GMR) of AUC or  $C_{\max}$  to be included in the [80–125]% interval.

### Single point study design

In BE studies for generic drug development of ophthalmic drug products, one sample of aqueous humor is taken at a certain timepoint per eye for measurement of drug concentration. A parallel study design involves sampling from only one eye, whereas crossover study designs involve sampling from both eyes, with test product in one eye and reference product in the other eye.

In both treatment arms of parallel studies, the subjects will be assigned to each of several prespecified sampling times  $t_j$  with  $j = 1, \dots, J$ . Let  $C_{ij}$  be the concentration of subject  $i = 1, \dots, N_j$  at the sampling time  $t_j$ . The total number of samples is  $n_{\text{tot}} = \sum_{j=1}^J N_j$  which is the same as the total number of subjects.

In the crossover studies, each subject with bilateral cataracts is randomly assigned for treatment with the test product in one eye and the reference product in the other eye.

A single sample of aqueous humor is collected from each eye at the same assigned sampling time  $t_j$ . Let  $C_{ijk}$  be the concentration of subject  $i = 1, \dots, N_{kj}$  at the same assigned sampling time  $t_j$  at each period/in each eye  $k = 1, 2$ . So, the total number of samples in the study is  $n_{\text{tot}} = \sum_{k=1}^2 \sum_{j=1}^J N_{kj}$  for crossover. The total number of subjects  $N = n_{\text{tot}} / 2$ .

### MB-TOST method

This approach uses NLMEM that account for the nonlinearity of the concentration versus time profile, as well as between- and within-subject variability (BSV and WSV).<sup>19</sup>

For crossover designs, the concentration  $C_{ijk}$  is described by a nonlinear function  $f$  depending on the vector of individual parameters  $\phi_{ijk}$  of subject  $i$  sampled at time  $t_j$  for period/eye  $k$ :

$$C_{ijk} = f(t_j, \phi_{ijk}) + g(t_j, \phi_{ijk}) \varepsilon_{ijk} \quad (3)$$

The  $l^{\text{th}}$  element of the vector  $\phi_{ijk}$  (i.e., the  $l^{\text{th}}$  individual parameter  $\phi_{ijkl}$   $l = 1, \dots, n_p$  where  $n_p$  is the number of PK parameters) can be decomposed as in the following equation:

$$\log(\phi_{ijkl}) = \log(\lambda_l) + \beta_l^T T_{ijk} + \beta_l^P P_k + \beta_l^S S_{ij} + \eta_{ijl} + \kappa_{ijkl}, \quad (4)$$

with  $\lambda_l$  the  $l^{\text{th}}$  element of the vector of fixed effects for the covariate reference class.  $T_{ijk}$ ,  $P_k$ , and  $S_{ij}$  are known

vectors of treatment, period and sequence covariates, respectively.  $\beta_l^T$ ,  $\beta_l^P$ , and  $\beta_l^S$  are the vectors of coefficients of treatment, period and sequence effects, respectively, for the  $l^{\text{th}}$  individual parameter.  $\eta_{ijl}$  is the  $l^{\text{th}}$  element of the vector of random effects  $\eta_{ij}$  for subject  $i$  at time  $t_j$  capturing the BSV.  $\kappa_{ijk}$  is the  $l^{\text{th}}$  element of the vector of random effects  $\kappa_{ijk}$  for subject  $i$  at time  $t_j$  and period  $k$ , capturing WSV.

The random effect vectors  $\eta_{ij}$  and  $\kappa_{ijk}$  are independently and normally distributed, with mean zero and variance-covariance matrix  $\Omega$  and  $\Gamma$ , respectively, of size  $n_p \times n_p$ . Let  $\omega_l^2$  and  $\gamma_l^2$  denote the  $l^{\text{th}}$  diagonal element of  $\Omega$  and  $\Gamma$ , corresponding to the BSVs and WSVs of the  $l^{\text{th}}$  parameter, respectively.

The residual errors  $\varepsilon_{ijk}$  are independently and identically distributed according to a normal centered distribution of variance  $\sigma^2$ . The associated error model can be additive  $g(t_j, \phi_{ijk}) = a$ , proportional  $g(t_j, \phi_{ijk}) = bf(t_j, \phi_{ijk})$ , or combined  $g(t_j, \phi_{ijk}) = a + bf(t_j, \phi_{ijk})$ .

For parallel designs, both Equations 1 and 2 simplify as follows (i.e., no period or sequence effect because there is only one period/eye and one sequence per subject):

$$\begin{aligned} C_{ij} &= f(t_j, \phi_{ij}) + g(t_j, \phi_{ij})\varepsilon_{ij} \\ \log(\phi_{ij}) &= \log(\lambda_l) + \beta_l^T T_{ij} + \eta_{ijl}, \end{aligned} \quad (5)$$

The full vector of parameters to be estimated in the NLMEM is  $\theta = (\lambda, \beta^T, \beta^P, \beta^S, \Omega, \Gamma, a, b)$  or  $(\lambda, \beta^T, \Omega, a, b)$  depending on the design. Let  $\text{VAR}(\hat{\theta})$  denote the asymptotic variance-covariance matrix of estimation, with the square root of its diagonal elements being the parameter asymptotic SEs.  $\text{VAR}(\hat{\theta})$  is derived as the inverse of the observed FIM.

The treatment effects on AUC and  $C_{\max}$  ( $\beta_{\text{AUC}}^T$  and  $\beta_{C_{\max}}^T$ , respectively) are functions of the reference effects  $\lambda$  and the treatment effects  $\beta^T$  on the PK model parameters. The delta-method is used to obtain estimates of  $\text{SE}(\hat{\beta}_{\text{AUC}}^T)$  and  $\text{SE}(\hat{\beta}_{C_{\max}}^T)$ .<sup>20</sup> These estimates and their SE are used to calculate the TOST statistics stated in Equation 2 to be compared to  $z_{1-\alpha}$  the  $(1 - \alpha)$ -quantile of a normal distribution.

In this approach, the exposure can be captured by  $\text{AUC}_{0-t_{\text{last}}}$  and  $\text{AUC}_{0-\infty}$ , both being different functions of the reference effects  $\lambda$  and the treatment effects  $\beta^T$  on the PK model parameters.

## Nonparametric bootstrap NCA

In recent years, the FDA has recommended parametric or nonparametric bootstrap NCA methods for the BE analysis of generic ophthalmic drug products.<sup>3-5</sup>

For crossover designs, at each time  $t_j$ , the mean of the concentrations  $C_{ijk}$  are calculated, for each period  $k$ . Here,

two versions of the nonparametric bootstrap NCA are evaluated (i) using the geometric mean of the concentrations or (ii) using the arithmetic mean of the concentrations.<sup>3</sup> In both versions, one value of  $C_{\max}^k$  and  $\text{AUC}^k$  is derived from the  $J^k$  (geometric or arithmetic) means at each timepoint, for each period (e.g.,  $k=1$  for treatment  $T$  and 2 for treatment  $R$ ).

Let  $C_{hij}$  be the  $h^{\text{th}}$  product concentration of the subject  $i$  and sampling timepoint  $t_j$ ,  $h = T, R$ ,  $i = 1, \dots, N_{hj}$ , and  $j = 1, \dots, J$ . Let  $\bar{C}_{hj}$  be the mean concentration of the  $h^{\text{th}}$  product at the  $j^{\text{th}}$  sampling timepoint, defined as follows:

$$\bar{C}_{hj} = \begin{cases} \sum_{i=N_{hj}} C_{hij}/N_{hj} & \text{if arithmetic mean} \\ \exp\left(\sum_{i=N_{hj}} \log(C_{hij})/N_{hj}\right) & \text{if geometric mean} \end{cases} \quad (6)$$

Thus, the AUC of the  $h^{\text{th}}$  product from the mean profile of drug concentration is calculated, using the trapezoidal rule, by:

$$\text{AUC}^h = t_1 \times \bar{C}_{hj}/2 + \sum_{j=1}^{J-1} (\bar{C}_{hj} + \bar{C}_{h,j+1}) \cdot (t_{j+1} - t_j)/2 \quad (7)$$

$C_{\max}$  is also obtained from the mean profile by:

$$C_{\max}^h = \max_{j \in (1, \dots, J)} (\bar{C}_{hj}) \quad (8)$$

Then, the ratios  $\frac{\text{AUC}^T}{\text{AUC}^R}$  and  $\frac{C_{\max}^T}{C_{\max}^R}$  are calculated and the 90% CIs around these ratio estimates are obtained using nonparametric bootstrap as follows.

Let us consider  $b$  ( $= 1, \dots, B$ ) bootstrap samples. For each sampling time  $t_j$ , the concentrations of the  $N_j$  individuals at both occasions  $k$  (i.e., for treatments  $R$  and  $T$ ) are re-sampled with replacement. For each bootstrap sample, at each time  $t_j$  the geometric and arithmetic mean of the concentrations  $C_{ij}$  are calculated for each treatment group, before deriving the respective treatment AUC,  $C_{\max}$  and the corresponding ratios. B ratio estimates  $\frac{\text{AUC}_b^T}{\text{AUC}_b^R}$  and  $\frac{C_{\max b}^T}{C_{\max b}^R}$  are obtained and the fifth and

95<sup>th</sup> percentiles of these distributions are used as the lower and upper bounds of the 90% CIs, respectively. Then, BE is concluded if the latter intervals are included in the [80–125]% interval.

To calculate the 90% bootstrap CIs for parallel designs, the concentrations of the  $N_j$  individuals are bootstrapped in each treatment group separately.

In this approach, the exposure is captured by  $\text{AUC}_{0-t_{\text{last}}}$  (i.e., the trapezoidal rule is used to obtain the AUC and no extrapolation to infinity is made from the last sampling

time). In the content of the following tables and figures, we refer to this approach as “NCA bootstrap Geometric and Arithmetic.”

## Simulation study

### Simulation settings

The PK data describing the PK of theophylline drug were simulated using the following one-compartment PK model with first-order absorption and elimination as in ref. 13, thereby assuming the test drug has the same PK structural model as the reference drug:

$$C(t) = f(t, k_a, Cl/F, V/F) = \frac{D}{V/F} \frac{k_a}{\frac{Cl/F}{V/F} - k_a} \left( \exp(-k_a t) - \exp\left(-\frac{Cl/F}{V/F} t\right) \right) \quad (9)$$

where  $F$  is the bioavailability,  $D$  the dose,  $V$  the distribution volume,  $k_a$  the absorption rate constant, and  $Cl$  the clearance of the drug. This PK model describes an exponential rise followed by a mono-exponential decrease in the concentrations which would have fitted the data of the ophthalmic product described in ref. 4.

Parallel and two-way crossover designs were considered, both with one sampling timepoint per subject. For each subject, the one sampling time was drawn from a rich or sparse set containing 10 or 5 sampling times, respectively, for selection. A total of 500 subjects was simulated with balanced sample sizes per sampling time and treatment group. For parallel designs, there were 25 subjects per sampling time in each treatment group for the rich set, and 50 subjects per sampling time in each treatment group for the sparse set. For crossover designs, there were 50 subjects per sampling time and per period/eye for the rich set, and 100 subjects per sampling time and period/eye for the sparse set.

The set of 10 sampling times was selected from:  $t \in \{0.25, 0.5, 1, 2, 3.5, 5, 7, 9, 12, 24\}$  h, and the set of 5 was selected from:  $t \in \{0.25, 1.5, 3.35, 12, 24\}$  h. The latter was obtained by maximization of the determinant of the FIM with fixed sampling times at  $\{1.5, 12\}$  h.<sup>8</sup> For the two-way crossover design, the same timepoint after drug administration is assumed in both periods/eyes within each individual.

Each subject received  $D = 4$  mg. The fixed effects for the reference treatment group of subjects were  $\lambda_{k_a} = 1.5$  /h,  $\lambda_{V/F} = 0.5$  L and  $\lambda_{Cl/F} = 40$  mL/h. Diagonal matrices were used for both BSV and WSV (only used when crossover design) with  $\omega_{k_a} = \omega_{V/F} = \omega_{Cl/F} = 52\%$  for parallel designs, and  $\omega_{k_a} = \omega_{V/F} = \omega_{Cl/F} = 50\%$  and  $\gamma_{k_a} = \gamma_{V/F} = \gamma_{Cl/F} = 15\%$  for crossover designs. We used

a combined error model for the residual variance with  $a = 0.1$  mg/L and  $b = 10\%$ .

The limit of quantification (LOQ) for simulated concentrations was set at 0.2 mg/L. Simulated concentrations below 0.2 mg/L, including negative concentrations ( $< 0$ ) and concentrations between 0 and 0.2 (0–0.2) mg/L were identified as below the limit of quantification (BLOQ). For MB-TOST, BLOQ concentrations were handled in the likelihood calculations using the method described in ref. 21 (i.e., the likelihood for a BLOQ observation is taken to be the likelihood that it is between 0 and the LOQ, given the current model parameter estimates). For NCA bootstrap, the BLOQ data were set to LOQ/2 as in Ref. 4.

### Evaluation

Both MB-TOST and NCA bootstrap were evaluated for BE on  $AUC_{0-t_{last}}$  and  $C_{max}$  (and not on  $AUC_{0-\infty}$ ) for the purpose of the comparison (see Supplementary Material S1 for the details of these metric calculations).

Despite the sparse sampling in our design, we used asymptotic SE for MB-TOST due to the large sampling size evaluated ( $N = 500$ ). Indeed, with  $\sqrt{N}$ -consistency for all PK parameters simulated and fitted with BSV (and WSV for crossover designs), we assumed asymptotic conditions were met.

For all approaches, the type I error was assessed under two different  $H_0$  hypotheses by setting  $\beta_{Cl/F}^T = \beta_{V/F}^T = \log(0.8)$  and  $\beta_{Cl/F}^T = \beta_{V/F}^T = \log(1.25)$ , corresponding to a simulated GMR for  $AUC_{0-t_{last}}$  and  $C_{max}$  of 1.25 and 0.8, respectively (see Supplementary Material S1).

We simulated under two different  $H_1$  hypotheses, first by setting  $\beta_{Cl/F}^T = \beta_{V/F}^T = \log(0.9)$  and second by setting  $\beta_{Cl/F}^T = \beta_{V/F}^T = \log(1) = 0$ , corresponding to a simulated GMR for  $AUC_{0-t_{last}}$  and  $C_{max}$  of 1.111 and 1, respectively. In all scenarios, period and sequence effects were assumed to be null (i.e.,  $\beta^P = \beta^S = 0$ ).

Therefore, we simulated 16 scenarios: two study designs (parallel or two-way crossover) times two sets of possible sampling times (5 or 10) and four hypotheses ( $\beta^T = \log(0.8), \log(0.9), \log(1), \log(1.25)$ ). Subsequently, our notation  $C10H_{0,0.8}$  refers to the scenario with a crossover design, a set of 10 sampling times to choose from and a simulated treatment effect  $\beta^T$  of  $\log(0.8)$ .

Estimation errors (EEs; the differences between the estimates and the simulated value) were calculated for model-based GMR and NCA bootstrap (geometric and arithmetic) mean ratio (MR) of AUC and  $C_{max}$ . Similarly for all scenarios, we calculated the empirical 5th and 95th percentiles of the GMR for MB-TOST (as the mean  $\pm u_{1-\alpha}$  × the standard deviation over all simulation estimates) and of the MR for the NCA bootstrap (geometric

and arithmetic, as the 5th and 95th percentiles over all simulation estimates) to be compared to the distribution of the estimated 5th and 95th percentiles obtained on each simulated data sets, for AUC and  $C_{\max}$ .

For the MB approach, relative EE (REE), relative bias (Rbias) and relative root mean square error (RRMSE) were calculated for fixed effects, random effects, and error model parameters.<sup>22</sup> Because the treatment, sequence, and period effects were simulated at zero, we did not calculate REE, Rbias, and RRMSE for the latter but absolute values of the corresponding metrics.

MB-TOST and NCA bootstrap were evaluated on 500 simulated data sets for each scenario. As NCA bootstrap runs faster, it was also evaluated on 10,000 simulated data sets for the challenging scenarios with 10 samples under one null and one alternative hypothesis:  $P10H_{0:1.25}$ ,  $P10H_{1:0.9}$ ,  $C10H_{0:1.25}$ , and  $C10H_{1:0.9}$ .

NCA bootstrap was evaluated with each of these scenarios, assuming there is no LOQ at 0.2 mg/L. Whereas MB-TOST was evaluated on scenarios  $P10H_{0:1.25}$  and  $C10H_{0:1.25}$  with no LOQ at 0.2 mg/L. In the absence of a defined LOQ, concentrations simulated below zero were set to 0.0001.

MB-TOST and NCA bootstrap were also evaluated on 500 simulated data sets for two parallel design scenarios with 10 possible sampling times and an increased residual variance at  $b = 30\%$  ( $P10H_{1:0.9}RUV_{0.3}$  and  $P10H_{0:1.25}RUV_{0.3}$ ) and two crossover design scenarios with 10 possible sampling times and increased WSV and residual variance at  $\gamma_{k_a} = \gamma_{V/F} = \gamma_{Cl/F} = 30\%$  and  $b = 30\%$  ( $C10H_{1:0.9}WSVandRUV_{0.3}$  and  $C10H_{0:1.25}WSVandRUV_{0.3}$ ).

## Implementation

For the NCA bootstrap method,  $B = 5000$  bootstrap samples were performed in the R statistical software version 4.0.2.

For the model-based approach, the Monolix 2018R2 software<sup>23</sup> was used which implements the SAEM algorithm for parameters.<sup>24</sup> Population parameters were estimated with 300 exploratory and 100 smoothing iterations, ensuring an appropriate convergence of SAEM visually assessed on a couple of data sets. Further, because there was only one observation per subject per occasion, 10 Markov chains were used to improve the precision of the estimation. This choice was motivated by a previous simulation study which already recommended to use 10 Markov chains with three sampling times per subject.<sup>13</sup> Initial values were set at the true value (i.e., the simulation value [see section simulation settings and Supplementary Material S2]) for all parameters except the treatment, period, and sequence effects all set to zero. Standard

errors were computed using the stochastic approximation method.<sup>25</sup>

BE tests were performed on GMR for MB-TOST and MR for NCA bootstrap (geometric and arithmetic), of  $AUC_{0-t_{\text{last}}}$  and  $C_{\max}$  in the R statistical software version 4.0.2.

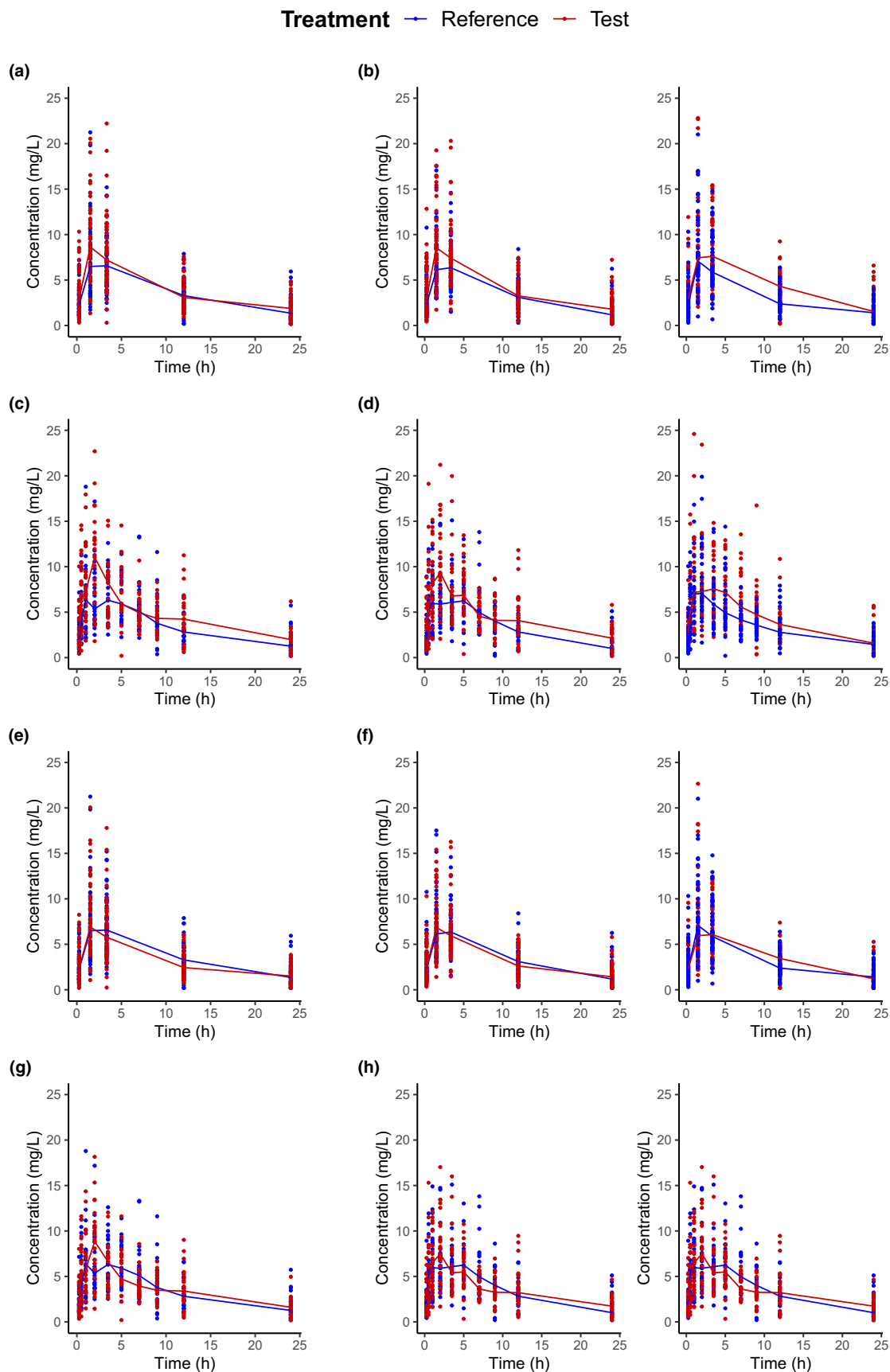
## RESULTS

Figure 1 represents a typical data set under the hypotheses of no BE ( $\beta_{Cl/F}^T = \beta_{V/F}^T = \log(0.8)$  top two lines) and equivalence ( $\beta_{Cl/F}^T = \beta_{V/F}^T = \log(1)$  bottom two lines) for scenarios with parallel (left column) and crossover (middle and right columns) designs and choosing the individual sample from a set of five or 10 possible times. As simulating  $\beta_{Cl/F}^T = \beta_{V/F}^T = \log(0.8)$  corresponds to  $\beta_{AUC}^T = \beta_{C_{\max}}^T = \log(1.25)$  (see Supplementary Material S1), the test treatment curve appears above the reference treatment curve on plots A-D of Figure 1. Consequently, simulating  $\beta_{Cl/F}^T = \beta_{V/F}^T = \log(1)$  leads to the reference and test treatment curves being overlaid on plots E-H of Figure 1. There were very few BLOQ data in our simulations, with 0.2% to 1.6% of simulated concentrations below 0.2 mg/L at 12 and 24 h post-dose respectively, for scenario  $P10H_{0:1.25}$  which simulates the lowest concentrations in the treatment group.

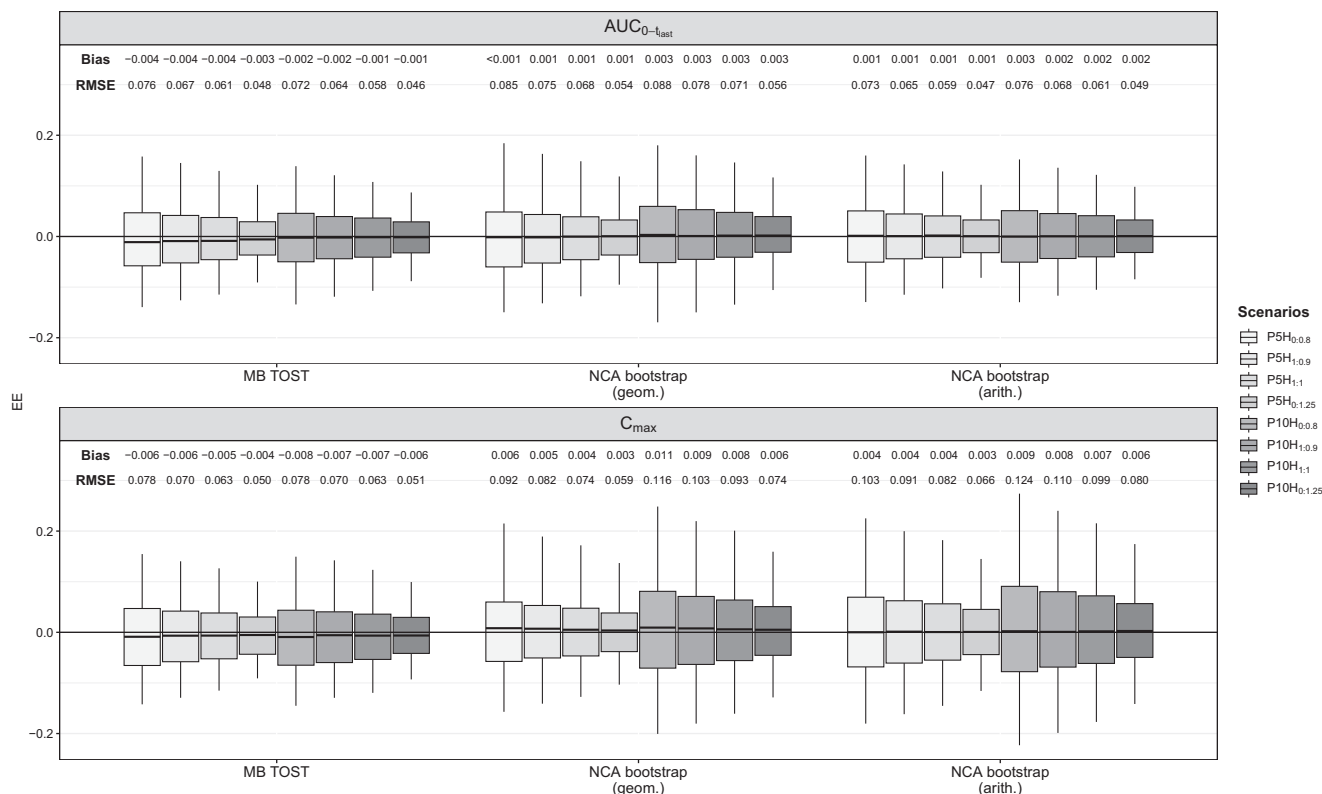
Model-based GMR and NCA bootstrap MR (geometric and arithmetic) for  $AUC_{0-t_{\text{last}}}$  and  $C_{\max}$  were unbiased and precise (see Figures 2 and 3). These results validated the parameter estimation from both the model-based and NCA approaches. The evaluation of the model parameter estimation is detailed in Supplementary Material S2.

Compared to parallel studies, crossover studies resulted in smaller 90% CIs (see Figures S5 and S6) and consequently higher power estimates (see Table 1) close to 100%, as expected. The number of samples to choose from impacted only the percentiles of the NCA bootstrap MR for  $C_{\max}$ . Because the AUC is an average over the concentrations, it is expected to be less sensitive to the design than  $C_{\max}$  which relies mostly on one measurement. Indeed, for designs with 10 rather than five samples to choose from, and consequently half of the sample size at each sampling time, the 90% CIs were significantly larger. This phenomenon led to 50% drops in power for parallel designs. For crossover designs, the drop in power was modest under the scenario with  $\beta^T = \log(0.9)$ . Indeed, in crossover designs, we simulated a rather low WSV which resulted in small 90% CIs and consequently a power close to one. Of note, 90% CIs derived from the empirical percentiles were even smaller.

For AUC, both MB-TOST and NCA bootstrap (geometric and arithmetic) obtained controlled type I error rates



**FIGURE 1** Concentrations versus time following the reference and test treatments for the first simulated clinical study for different simulation scenarios with parallel and two-way crossover designs (period 1 on the left and period 2 on the right). The lines are connecting the median in each group successive timepoints. (a) P5H<sub>0:0.8</sub>; (b) C5H<sub>0:0.8</sub>; (c) P10H<sub>0:0.8</sub>; (d) C10H<sub>0:0.8</sub>; (e) P5H<sub>1:1</sub>; (f) C5H<sub>1:1</sub>; (g) P10H<sub>1:1</sub>; and (h) C10H<sub>1:1</sub>.



**FIGURE 2** Boxplot (2.5th, 25th, 50th, 75th, and 97.5th percentiles) of estimation errors (EEs) of the geometric mean ratio (MB-TOST) and the mean ratio (NCA bootstrap geometric and arithmetic) for  $AUC_{0-t_{last}}$  and  $C_{max}$ , for 500 simulated data sets with parallel designs.  $AUC_{0-t_{last}}$ , area under the time-concentration curve from time zero until the last sampling time;  $C_{max}$ , maximum concentration; EEs, estimation errors; MB-TOST, model-based two one-sided test; NCA, noncompartmental analysis.

on 500 simulated data sets (see Table 2) with the exception of a slightly inflated estimate in scenario  $P10H_{0,0.8}$  for NCA bootstrap geometric (0.076).

In terms of power, MB-TOST and NCA bootstrap arithmetic obtained similar mean estimates, whereas NCA bootstrap geometric power estimates were consistently lower in parallel designs.

For  $C_{max}$ , the 90% CIs of all approaches were larger than the 90% CIs derived from the empirical percentiles. Further, the 90% CIs of the NCA bootstrap (geometric and arithmetic) MR was wider than that for the model-based GMR. Consequently, type I error rates were significantly conservative, for NCA bootstrap in general with the exception of scenario  $P5H_{0,0.8}$  when using arithmetic means and for scenarios  $C5H_{0,0.8}$  and  $C10H_{0,0.8}$  for MB-TOST.

In terms of power, MB-TOST had a similar power as NCA bootstrap in crossover designs, but higher power than NCA bootstrap in parallel designs. Of note for  $C_{max}$ , conversely to AUC, NCA bootstrap geometric had slightly higher power than NCA bootstrap arithmetic in parallel studies, but the two were similar in crossover designs.

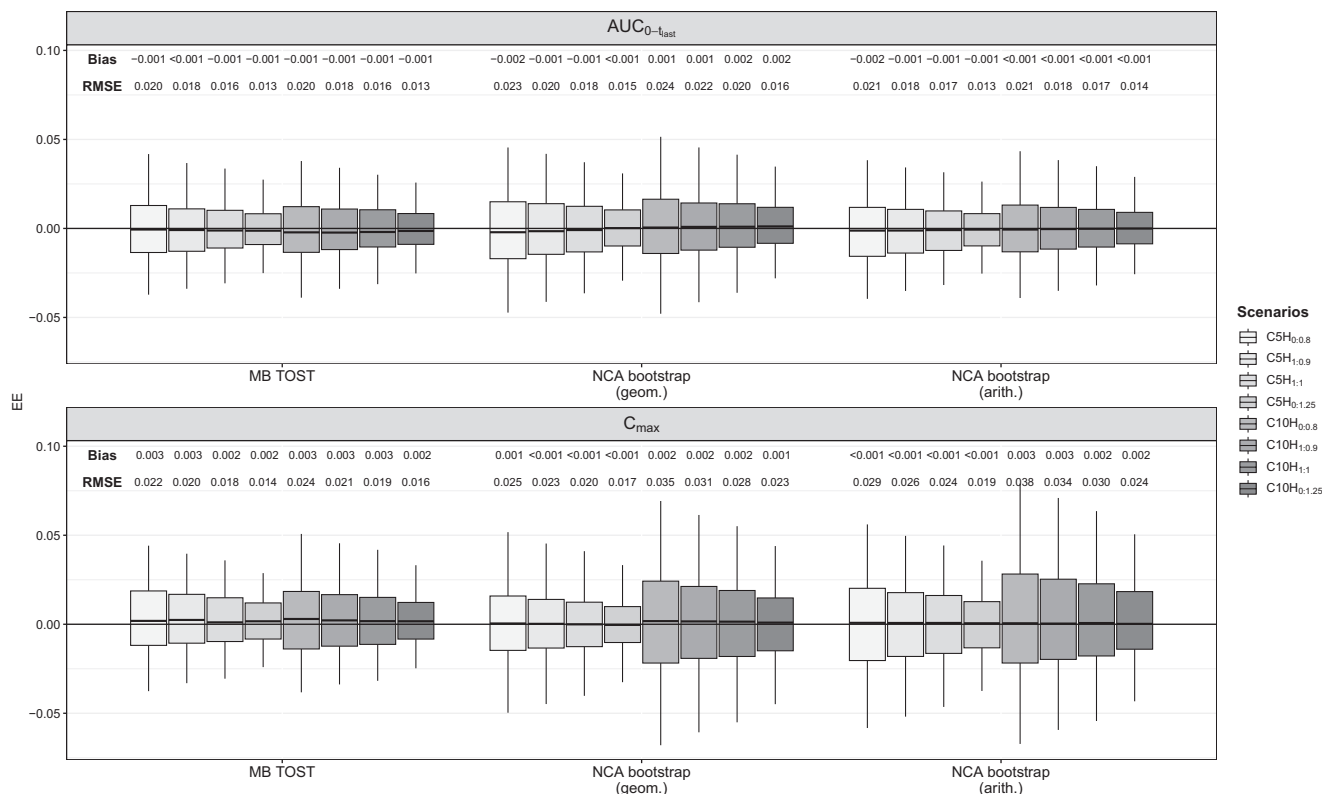
On the scenarios where we investigated NCA bootstrap on 10,000 simulated data sets, the type I error rates converged closer to the 5% target value and with narrower CIs

compared with 500 simulated data sets (Supplementary Material S5). Other trends were similar to those observed in Tables 1 and 2. When we analyzed the data without a LOQ at 0.2 mg/L, the results were generally consistent (Supplementary Material S6) for AUC.

We observed no impact of inflating the simulated RUV on the type I error of MB-TOST and NCA bootstrap for AUC or  $C_{max}$  (Supplementary Material S4), in general. Inflating the simulated WSV only resulted in a slight inflation of the type I error of NCA bootstrap arithmetic for AUC in crossover designs. For a parallel design, the power estimate for AUC decreased by 28%, 21%, and 19% with MB-TOST, NCA bootstrap geometric mean, and NCA bootstrap arithmetic, respectively, whereas on the crossover design they decreased by 8%, 17%, and 7%, respectively. Likewise, the power estimate for  $C_{max}$  decreased by 26% and 21% with MB-TOST using a parallel and crossover design, respectively, and decreased by more than 60% for both NCA bootstrap methods and designs.

Overall, MB-TOST and NCA bootstrap performed similarly for AUC. For  $C_{max}$ , NCA bootstrap tended to be conservative and less powerful for parallel designs compared to MB-TOST, likely due to the more precise estimates of MB-TOST when the PK model is correctly specified.





**FIGURE 3** Boxplot (2.5th, 25th, 50th, 75th, and 97.5th percentiles) of EEs of the geometric mean ratio (MB-TOST) and the mean ratio (NCA bootstrap geometric and arithmetic) for  $AUC_{0-t_{last}}$  and  $C_{max}$  for 500 simulated data sets with two-way crossover designs.  $AUC_{0-t_{last}}$ , area under the time-concentration curve from time zero until the last sampling time;  $C_{max}$ , maximum concentration; EEs, estimation errors; MB-TOST, model-based two one-sided test; NCA, noncompartmental analysis.

## DISCUSSION

In this study, we used NLME tools to simulate PK BE studies for a generic ophthalmic drug product with one sampling timepoint per subject and per occasion. Consistent with our previous research, we estimated treatment effects on all model parameters<sup>12,13,15,26</sup> and evaluated various scenarios, including the study design (parallel or crossover), set of sampling times, and GMR. The aim was to evaluate and compare the performances of MB-TOST<sup>15</sup> and the NCA bootstrap<sup>4</sup> using concentration geometric or arithmetic means.

All model parameters were well-estimated except for the additive term of the combined error model, possibly due to the limitations of using a single sample per patient (see Tables S1–S8 in Supplementary Material S2). Whether a proportional error model or fixing the additive term would have solved that issue requires further evaluation.<sup>27</sup> It had, however, no impact on the estimation of the treatment, sequence, or period effects parameters as well as on the GMRs of AUC and  $C_{max}$ .

Both MB-TOST and the NCA bootstrap (geometric and arithmetic) performed similarly in terms of type I error rate for AUC. For  $C_{max}$ , the NCA bootstrap approach

almost always provided conservative type I error rates with values below 0.05. For parallel designs, NCA bootstrap, which does not need to specify the underlying PK model, led to noticeable lower power for  $C_{max}$  as compared to MB-TOST when the underlying PK model is correctly specified and the test drug has the same PK structural model as the reference drug. We used optimal design to choose the sampling times of the sparse set which did not include the  $t_{max}$  according to the fixed effects. This may have favored MB-TOST as NCA relies on observed sampling times to estimate  $C_{max}$ . However, this advantage is difficult to appreciate as due to BSV all subjects had different  $t_{max}$ . Power was also particularly impacted by the number of sampling timepoints. Indeed, with 10 sampling times to choose from, fewer subjects were allocated to informative timepoints, which reduced the informativeness of the design<sup>28</sup> and led to larger 90% CIs of the GMR and MR for  $C_{max}$ . The simulation showed that estimates for  $C_{max}$  were not as efficient as for AUC because  $C_{max}$  is only based on one value whereas AUC is based on an average over the concentrations.

This work presents a notable limitation in the use of the simulated PK model structure, as well as the simulated values as initial guess for the parameters (except for

**TABLE 1** Study power [95% CI] using parallel and two-way crossover study designs with 500 simulated data sets (estimates in brackets are obtained with 10,000 simulated data sets).

	$C_{max}$					
	$AUC_{0-t_{last}}$			$C_{max}$		
	NCA bootstrap		MB-TOST	NCA bootstrap		Arithmetic
Geometric	Arithmetic	Geometric		Arithmetic		
P5H <sub>1,0.9</sub>	0.692 [0.649; 0.732]	0.600 [0.556; 0.643]	0.696 [0.654; 0.736]	0.598 [0.554; 0.641]	0.388 [0.345; 0.432]	0.378 [0.335; 0.422]
P10H <sub>1,0.9</sub>	0.696 [0.654; 0.736]	0.554 [0.509; 0.598] (0.565 [0.555; 0.575])	0.676 [0.633; 0.717] (0.681 [0.672; 0.690])	0.588 [0.543; 0.632]	0.194 [0.160; 0.231] (0.196 [0.188; 0.204])	0.170 [0.138; 0.206] (0.166 [0.159; 0.173])
P5H <sub>1,1</sub>	0.950 [0.927; 0.967]	0.916 [0.888; 0.939]	0.968 [0.949; 0.982]	0.926 [0.899; 0.947]	0.760 [0.720; 0.797]	0.664 [0.621; 0.705]
P10H <sub>1,1</sub>	0.968 [0.949; 0.982]	0.884 [0.853; 0.911]	0.970 [0.951; 0.983]	0.896 [0.866; 0.921]	0.370 [0.328; 0.414]	0.294 [0.254; 0.336]
C5H <sub>1,0.9</sub>	0.998 [0.989; 1.000]	1 [0.993; 1.000]	1 [0.993; 1.000]	0.998 [0.989; 1.000]	1 [0.993; 1.000]	0.996 [0.986; 1.000]
C10H <sub>1,0.9</sub>	0.998 [0.989; 1.000]	1 [0.993; 1.000] (1 [1.000; 1.000])	1 [0.993; 1.000]	0.998 [0.989; 1.000]	0.972 [0.953; 0.985] (0.973 [0.970; 0.976])	0.932 [0.906; 0.952] (0.949 [0.944; 0.953])
C5H <sub>1,1</sub>	1 [0.993; 1.000]	1 [0.993; 1.000]	1 [0.993; 1.000]	0.998 [0.989; 1.000]	1 [0.993; 1.000]	1 [0.993; 1.000]
C10H <sub>1,1</sub>	1 [0.993; 1.000]	1 [0.993; 1.000]	1 [0.993; 1.000]	0.998 [0.989; 1.000]	1 [0.993; 1.000]	1 [0.993; 1.000]

Note: The power estimates presented here are not corrected for the departure from the nominal level of the test.

Abbreviations:  $AUC_{0-t_{last}}$ , area under the time-concentration curve from time zero until the last sampling time; CI, confidence interval;  $C_{max}$ , maximum concentration; MB-TOST, model-based two one-sided test; NCA, noncompartmental analysis.

**TABLE 2** Type I error rates [95% CI] using parallel and two-way crossover study designs with 500 simulated data sets (estimates in brackets are obtained with 10,000 simulated data sets).

	$C_{max}$					
	$AUC_{0-t_{last}}$			$C_{max}$		
	NCA bootstrap		MB-TOST	NCA bootstrap		Arithmetic
Geometric	Arithmetic	Geometric		Arithmetic		
P5H <sub>0,0.8</sub>	0.068 [0.048; 0.094]	0.058 [0.039; 0.082]	0.064 [0.044; 0.089]	0.054 [0.036; 0.078]	0.026 [0.014; 0.044]	0.034 [0.020; 0.054]
P10H <sub>0,0.8</sub>	0.068 [0.048; 0.094]	0.076 [0.054; 0.103]	0.066 [0.046; 0.091]	0.060 [0.041; 0.085]	0.024 [0.012; 0.042]	0.026 [0.014; 0.044]
P5H <sub>0,1.25</sub>	0.038 [0.023; 0.059]	0.058 [0.039; 0.082]	0.052 [0.034; 0.075]	0.038 [0.023; 0.059]	0.028 [0.015; 0.047]	0.040 [0.025; 0.061]
P10H <sub>0,1.25</sub>	0.052 [0.034; 0.075]	0.064 [0.044; 0.089] (0.056 [0.052; 0.061])	0.066 [0.046; 0.091] (0.057 [0.052; 0.061])	0.032 [0.018; 0.051]	0.014 [0.006; 0.029] (0.020 [0.018; 0.023])	0.012 [0.004; 0.026] (0.020 [0.017; 0.023])
C5H <sub>0,0.8</sub>	0.044 [0.028; 0.066]	0.066 [0.046; 0.091]	0.062 [0.043; 0.087]	0.028 [0.015; 0.047]	0.016 [0.007; 0.031]	0.024 [0.012; 0.042]
C10H <sub>0,0.8</sub>	0.048 [0.031; 0.071]	0.052 [0.034; 0.075]	0.056 [0.038; 0.080]	0.026 [0.014; 0.044]	0.014 [0.006; 0.029]	0.014 [0.006; 0.029]
C5H <sub>0,1.25</sub>	0.060 [0.041; 0.085]	0.050 [0.033; 0.073]	0.050 [0.033; 0.073]	0.062 [0.043; 0.087]	0.024 [0.012; 0.042]	0.024 [0.012; 0.042]
C10H <sub>0,1.25</sub>	0.052 [0.034; 0.075]	0.068 [0.048; 0.094] (0.062 [0.057; 0.067])	0.070 [0.049; 0.096] (0.056 [0.052; 0.061])	0.060 [0.041; 0.085]	0.010 [0.003; 0.023] (0.013 [0.011; 0.015])	0.014 [0.006; 0.029] (0.015 [0.013; 0.018])

Abbreviations:  $AUC_{0-t_{last}}$ , area under the time-concentration curve from time zero until the last sampling time; CI, confidence interval;  $C_{max}$ , maximum concentration; MB-TOST, model-based two one-sided test; NCA, noncompartmental analysis.

the treatment effects which were fixed to 0) in the MB approach, whereas the model structure and parameter values are unknown in reality. In this paper, we did not test the performance of MB-TOST if the underlying PK model is not correctly specified, which is likely with a limited number of sampling points. This is mitigated, however, by the fact that BE studies occur downstream of the drug development process when a non-negligible knowledge has been accumulated on the PKs of the drug at least in the reference treatment arm.<sup>6</sup> Moreover, the model would be subjected to a thorough evaluation prior to its use for the BE analysis.<sup>7</sup> Nonetheless, using the true model and parameter estimates provided a favorable setup for MB-TOST and challenging its robustness to departure from the simulated PK model and parameter values needs to be explored in the future.

Another limitation is only 500 datasets were simulated under each of the 16 scenarios due to the limitation of computation power, which can impact the precision of the evaluation of the performance (e.g., bias, type 1 error rate, and power). Computing times ranged in average from 15 min per data set for parallel designs to 45 min per data set for crossover designs, on a similar computing setup. Notably, it is the default number for simulation-based goodness of fit plots in PsN, a recognized tool for NLMEM.<sup>29</sup> Additionally, the number of bootstraps for the NCA approaches could also have been increased from 5000<sup>4</sup> to 10,000.

In addition, the lowest concentration level in real studies is zero, but, in simulation, negative concentration levels can happen, which is a limitation of using simulation to generate concentration data. For those concentrations simulated below zero, they are imputed as 0.0001 when no LOQ is defined. When a LOQ is defined, the likelihood is taken to be between zero and the LOQ for all simulated BLOQ concentrations including those with 0 to LOQ (e.g., 0.2) values as well as negative (<0) concentration values for MB-TOST. Likewise, for NCA bootstrap, LOQ/2 is imputed for both 0–0.2 mg/L values and negative (<0) concentration values. These are typically used methods to handle BLOQ values, however, for simulated concentration values below zero, these imputations may impact bias, type 1 error rate, and power. This limitation also contributes to the observed inflated type 1 error rate above 0.05 in certain scenarios, besides the limited simulation data sets (500).

In addition, we considered the same magnitude of BSV and WSV for  $k_a$ ,  $Cl/F$ , and  $V/F$ . Further, the initial 16 scenarios considered a residual variance at  $b = 10\%$  and a WSV at 15%, which are rather low. We explored the impact of increasing both variances at 30% in four additional scenarios and found no impact on MB-TOST and NCA bootstrap type I error rate and, as expected, a decrease

in power although to a remarkable extent for  $C_{max}$  with NCA bootstrap approaches on both parallel and crossover designs.

In conclusion, our simulations show that MB-TOST demonstrates higher power and controlled type I error rate when compared to NCA-TOST in single sample PK studies, provided that the underlying PK model is correctly specified, and the test drug has the same PK structural model as the reference drug. As such, MB-TOST may potentially serve as an alternative approach to evaluate BE for generic ophthalmic drug products.

## AUTHOR CONTRIBUTIONS

All authors wrote the manuscript. K.F., W.S., G.S., S.G., L.Z., L.F., F.M., and J.B. designed the research. C.T., F.L., and J.B. performed the research. C.T. analyzed the data.

## ACKNOWLEDGMENTS

The authors would like to acknowledge all their collaborators in the project “Evaluation of model-based Bioequivalence (MBBE) statistical approaches for sparse design PK studies” under the FDA contract 75F40119C10111.

## FUNDING INFORMATION

This work was supported by the U.S. Food and Drug Administration, Silver Spring, MD 20993, USA, under FDA contract 75F40119C10111.

## CONFLICT OF INTEREST STATEMENT

M.D., K.F., W.S., G.S., S.G., L.Z., and L.F. are employed by the U.S. Food and Drug Administration. As Editor-in-Chief of *CPT:Pharmacometrics and Systems Pharmacology*, F.M. was not involved in the review or decision process for this paper. The opinions expressed in this manuscript are those of the authors and should not be interpreted as the position of the U.S. Food and Drug Administration. All other authors declared no competing interests for this work.

## ORCID

Coralie Tardivon  <https://orcid.org/0000-0002-2304-9910>

Florence Loingeville  <https://orcid.org/0000-0002-4278-2974>

Mark Donnelly  <https://orcid.org/0000-0001-7537-9615>

Liang Zhao  <https://orcid.org/0000-0002-0257-9082>

France Mentré  <https://orcid.org/0000-0002-7045-1275>

Julie Bertrand  <https://orcid.org/0000-0002-6568-1041>

## REFERENCES

1. U.S. Food and Drug Administration. Bioequivalence studies with pharmacokinetic endpoints for drugs submitted under an ANDA guidance for industry. 2021. Accessed March 27, 2023. <https://www.fda.gov/media/87219/download>

2. Choi S, Lionberger RA. Clinical, pharmacokinetic, and in vitro studies to support bioequivalence of ophthalmic drug products. *AAPS J*. 2016;18:1032-1038.
3. FDA Office of Generic Drugs. Draft guidance on loteprednol etabonate. 2019. Accessed March 27, 2023. [https://www.accessdata.fda.gov/drugsatfda\\_docs/psg/PSG\\_210565.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/psg/PSG_210565.pdf)
4. Shen M, Machado SG. Bioequivalence evaluation of sparse sampling pharmacokinetics data using bootstrap resampling method. *J Biopharm Stat*. 2017;27(2):257-264.
5. FDA Office of Generic Drugs. Draft guidance on loteprednol etabonate; tobramycin. 2020. Accessed March 27, 2023. [https://www.accessdata.fda.gov/drugsatfda\\_docs/psg/PSG\\_050804.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/psg/PSG_050804.pdf)
6. Mould DR, Upton RN. Basic concepts in population modeling, simulation, and model-based drug development. *CPT Pharmacometrics Syst Pharmacol*. 2012;1:e6.
7. Seng Yue C, Ozdin D, Selber-Hnatiw S, Ducharme MP. Opportunities and challenges related to the implementation of model-based bioequivalence criteria. *Clin Pharmacol Ther*. 2019;105(2):350-362.
8. Panhard X, Mentré F. Evaluation by simulation of tests based on non-linear mixed-effects models in pharmacokinetic interaction and bioequivalence cross-over trials. *Stat Med*. 2005;24(10):1509-1524.
9. Lindstrom ML, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990;46(3):673-687.
10. Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. *Ann Stat*. 1999;27:94-128.
11. Dubois A, Gsteiger S, Pigeolet E, Mentré F. Bioequivalence tests based on individual estimates using non-compartmental or model-based analyses: evaluation of estimates of sample means and type I error for different designs. *Pharm Res*. 2010;27(1):92-104.
12. Dubois A, Lavielle M, Gsteiger S, Pigeolet E, Mentré F. Model-based analyses of bioequivalence crossover trials using the stochastic approximation expectation maximisation algorithm. *Stat Med*. 2011;30(21):2582-2600.
13. Loingeville F, Bertrand J, Nguyen TT, et al. New model-based bioequivalence statistical approaches for pharmacokinetic studies with sparse sampling. *AAPS J*. 2020;22(6):1-8.
14. Nie L. Convergence rate of MLE in generalized linear and non-linear mixed-effects models: theory and applications. *J Stat Plan Infer*. 2007;137:1787-1804.
15. Möllenhoff K, Loingeville F, Bertrand J, et al. Efficient model-based bioequivalence testing. *Biostatistics*. 2020;23:314-327. doi:10.1093/biostatistics/kxaa026
16. FDA Office of Generic Drugs. Draft Guidance on Progesterone. 2011. Accessed March 27, 2023. [https://www.accessdata.fda.gov/drugsatfda\\_docs/psg/Progesterone\\_caps\\_19781\\_RC02-11.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/psg/Progesterone_caps_19781_RC02-11.pdf)
17. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokin Biopharm*. 1987;15(6):657-680.
18. EMA. Guideline on the Investigation of Bioequivalence. 2014. Accessed March 27, 2023. [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf)
19. Lavielle M. *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools*. Chapman and Hall/CRC; 2014.
20. Oehlert GW. A note on the Delta method. *Am Stat*. 1992;46(1):27-29.
21. Panhard X, Samson A, Mentré F. Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: application to HIV dynamics model. *Comput Stat Data Anal*. 2006;51(3):1562-1574.
22. Savic R, Radojka M, Mentré F, Lavielle M. Implementation and evaluation of the SAEM algorithm for longitudinal ordered categorical data with an illustration in pharmacokinetics-pharmacodynamics. *AAPS J*. 2011;13(1):44-53.
23. Team Monolix, Lixoft. The Monolix Software 2018 - R2. 2018.
24. Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Anal*. 2005;49(4):1020-1038.
25. Louis TA. Finding the observed information matrix when using EM algorithm. *J R Stat Soc B*. 1982;44:226-233.
26. Guhl M, Mercier F, Hofmann C, et al. Impact of model misspecification on model-based tests in PK studies with parallel design: real case and simulation studies. *J Pharmacokin Pharmacodyn*. 2022;49(5):557-577.
27. Plan EL, Maloney A, Mentré F, Karlsson MO, Bertrand J. Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for dose-response models. *AAPS J*. 2012;14(3):420-432.
28. Nyberg J, Bazzoli C, Ogungbenro K, et al. Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics studies. *Br J Clin Pharmacol*. 2015;79(1):6-17.
29. Keizer RJ, Karlsson MO, Hooker A. Modeling and simulation workbench for NONMEM: tutorial on Pirana, PsN, and Xpose. *CPT Pharmacometrics Syst Pharmacol*. 2013;2:e50.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Tardivon C, Loingeville F, Donnelly M, et al. Evaluation of model-based bioequivalence approach for single sample pharmacokinetic studies. *CPT Pharmacometrics Syst Pharmacol*. 2023;12:904-915. doi:[10.1002/psp4.12960](https://doi.org/10.1002/psp4.12960)