HAL
open science

Conditional Non-parametric Bootstrap for Non-linear Mixed Effect Models

Emmanuelle Comets, Christelle Rodrigues, Vincent Jullien, Moreno Ursino

To cite this version:

Emmanuelle Comets, Christelle Rodrigues, Vincent Jullien, Moreno Ursino. Conditional Non-parametric Bootstrap for Non-linear Mixed Effect Models. Pharmaceutical Research, 2021, 38 (6), pp.1057-1066. 10.1007/s11095-021-03052-6 . hal-03280469

HAL Id: hal-03280469

https://hal.science/hal-03280469v1

Submitted on 19 Oct 2021

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

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

Pharmaceutical Research manuscript No.
(will be inserted by the editor)

Conditional non-parametric bootstrap for non-linear mixed effect models

Emmanuelle Comets · Christelle Rodrigues · Vincent Jullien · Moreno Ursino

Received: date / Accepted: date

Abstract ***Purpose:*** Non-linear mixed effect models are widely used and increasingly integrated into decision-making processes. Propagating uncertainty is an important element of this process, and while standard errors (SE) on parameters are most often computed using asymptotic approaches, alternative methods such as the bootstrap are also available. In this article, we propose a modified residual parametric bootstrap taking into account the different levels of variability involved in these models.

Methods: The proposed approach uses samples from the individual conditional distribution, and was implemented in R using the saemix algorithm. We performed a simulation study to assess its performance in different scenarios, comparing it to the asymptotic approximation and to standard bootstraps in terms of coverage, also looking at bias in the parameters and their SE.

Results: Simulations with an Emax model with different designs and sigmoidicity factors showed a similar coverage rate to the parametric bootstrap, while requiring less hypotheses. Bootstrap improved coverage in several scenarios compared to the asymptotic method especially for the variance param-

Emmanuelle Comets
Université de Paris, INSERM IAME; INSERM, CIC 1414; Rennes-1 University, France
16 rue Henri Huchard, 75018 Paris, France
Tel: +33-62582-4950
E-mail: emmanuelle.comets@inserm.fr

Christelle Rodrigues
INSERM U1129, Paris Descartes University, CEA, France

Vincent Jullien
UF Pharmacologie, GH Paris Seine Saint-Denis, Université Paris 13, France

Moreno Ursino
Centre de Recherche des Cordeliers, Inserm, Université de Paris, F-CRIN PARTNERS platform, AP-HP, France

eters. However, all bootstraps were sensitive to estimation bias in the original datasets.

Conclusions: The conditional bootstrap provided better coverage rate than the traditional residual bootstrap, while preserving the structure of the data generating process.

Keywords Bootstrap · Conditional distribution · Non-linear mixed effect models · Uncertainty of parameter estimates

1 Introduction

Nonlinear mixed effect models (NLMEM) are widely used namely in pharmacokinetics to characterise the evolution of drug concentrations after dose administration, and in pharmacodynamics where they are used to integrate disease models and pathophysiology to better understand the response of biological systems to treatment [1]. Modelling and simulation has become a key component in the analysis and design of clinical trials to establish the most promising therapeutic candidates, as well as the dosage regimens more likely to be effective and safe [2], and is also increasingly used to individualise diagnostics and treatments [3]. As the models become increasingly more comprehensive, and the extrapolations increasingly more sophisticated, it becomes also important to report the confidence we have in them to propagate this uncertainty into decision making processes and realistic evaluation. In the present article, we focus on the issue of standard errors of estimation (SE), which measure the precision of estimation for the parameters in the model. In maximum-likelihod estimation, they are often computed using an asymptotic approximation as the inverse of the Fisher information matrix [4], while Bayesian methods consider the entire posterior distributions and use it to assess and propagate uncertainty [5]. How well the asymptotic approximation holds in practice depends on the sample size, both in terms of number of subjects and in terms of number of samples per subject, and on the degree of non-linearity in the model [6]. Other approaches to derive the standard errors include log-likelihood profiling [7], Sampling Importance Resampling (SIR) [8] and bootstrap methods [9].

Bootstrap methods mimic the process of repeating a study several times to assess the distribution of an estimator, using only the data available in the original study through resampling. For independent and identically distributed observations, Efron [10] proposed resampling with replacement from the pool of available measurements, creating bootstrap datasets on which the estimator of interest is applied in order to approximate its distribution. In more complex settings, like linear regression or non-homoscedastic error models, issues such as within-subject correlation and heteroscedastic variances have to be considered. Several bootstrap approaches have been proposed in these situations, with the most intuitive being the case bootstrap, also called paired bootstrap [11], where the resampling is performed at the level of the subject.

Alternatively, the residual bootstrap resamples residuals in the empirical distribution (non-parametric residual bootstrap) or in an external distribution with suitable characteristics (parametric boostrap) [12]. Bootstraps were extended to NLMEM by considering the different levels of variability involved at the individual and population level. Das and Krishen [13] proposed a residual bootstrap involving two levels of resampling. In small sample studies, however, the distribution of the residuals does not reflect the variability in the population because of a phenomenon called shrinkage, causing a regression to the mean due to the limited amount of information to jointly estimate a large number of parameters and/or the residual variability. Carpenter et al. [14] proposed to correct the distribution of residuals before resampling, using a ratio between the empirical variance-covariance matrix and the estimated variance matrix of the random effects. Different bootstrap methods were summarised and evaluated both in linear and non-linear mixed effect models by Thai et al. [9, 15]. For linear mixed effect models with homoscedastic variance, the case, parametric and non-parametric bootstraps exhibited good performances in terms of absence of bias and coverage rates, while the asymptotic method proved reliable for the estimation of the parameters and their SE [9]. These bootstraps were then evaluated in NLMEM, where the results were more contrasted. In this more challenging settings, all bootstraps and the asymptotic method exhibited some measure of bias, especially on a model involving non-linear pharmacokinetics where some parameters have strong non-linearity [15]).

The objective of the present manuscript was to develop a new bootstrap approach and evaluate it by comparison to usual bootstraps in NLMEM. Our proposed approach, the conditional non-parametric bootstrap, is a residual bootstrap using samples from the the conditional distribution of the individual parameters, which have been shown to correct the shrinkage in the empirical Bayes estimates in non-linear models [16]. We use simulations inspired by the study performed by Plan et al. [17], using an Emax and a Hill model to investigate the influence of non-linearity with both sparse and rich designs.

2 Methods

2.1 Statistical models

Let the random variable y_{ij} denote the observation of the longitudinal data, which we will assume in this work to be continuous, at time t_{ij} for subject $i = 1, ..., N$ $(j = 1, ..., n_i)$. Following [18], the NLMEM can be formalised as a hierarchical model involving a parametric structural model f defining the longitudinal evolution, depending on fixed parameters μ and subject-specific random effects η_i with a parametric distribution, coupled with a probability distribution on the measurements, here a Gaussian with a standard deviation defined by a parametric function g:

$$\begin{cases} y_{ij} = f(x_{ij}, \theta_i) + g(x_{ij}, \theta_i, \sigma)\ \epsilon_{ij} \\ \theta_i = h(\mu, \eta_i) \\ \eta_i \sim N(0, \Omega) \\ \epsilon_{ij} \sim N(0, 1) \end{cases} \tag{1}$$

In these equations, θ_i denotes the individual parameters, which we assume are related through a function h to the p-dimensional vector containing the fixed effects μ, and the q-dimensional vector containing the random effects η_i. ϵ_{ij} is a random variable assumed to be normally distributed. σ denotes variance parameters entering the function g, which expresses the standard deviation of the measurement error and is generally either constant (homoscedastic variance) or a function of f. The random effects η_i and the residual errors ϵ_{ij} are assumed to be independent for different subjects and to be independent of each other for the same subject. We will denote by $\xi_i = \{x_{i1}, ..., x_{in_i}\}$ the design variables, for example the n_i sampling times for subject i.

The parameters of the model $\Psi = (\mu, \Omega, \sigma)$ can be estimated by maximum likelihood. In NLMEM, the likelihood associated with equation (1) is intractable as individual likelihoods need to integrate out the unknown parameters θ_i over their distribution $\mathcal{D}_\theta$. In this study, we used the SAEM algorithm to obtain parameter estimates [19, 18]. The maximum likelihood estimate (MLE) $\hat{\theta}$ of θ is asymptotically normally distributed with mean θ and asymptotic covariance matrix given by the inverse of the Fisher information matrix (FIM), denoted M_F, usually obtained through a first-order approximation of the model [4]. During the estimation process, the individual parameters θ_i act as nuisance parameters, but their conditional distribution is used during the stochastic approximation phase to sample the unknown parameters η_i and obtain a complete dataset from which the conditional log-likelihood is derived [19]. The estimated individual parameters, called empirical Bayes estimates (EBE), can be defined as the mode or the median of the conditional distribution obtained at the end of the estimation process.

2.2 Conditional bootstrap

Bootstrap consists in repeatedly generating pseudo-samples distributed according to the same distribution as that the original sample. In NLMEM, bootstrap methods include the case bootstrap (Case), resampling the entire vector of individual observations, the parametric bootstrap resampling residuals from a theoretical distribution (Par) and the non-parametric bootstrap resampling from estimated residuals (NP) [15]. We propose a new bootstrap method, which we call conditional non-parametric residual bootstrap (cNP), using samples from the conditional distributions instead of shrinkage-corrected empirical Bayes estimates.

The algorithm is as follows: after fitting the model to the data, we obtain K samples for each vector η_i, η_{ik} $(k = 1, ..., K)$ from the conditional distri-

butions, center them and resample vectors of individual parameters η_i^* with replacement. In this work we implement this step by sampling from all conditional samples over $i = 1...N, k = 1...K$ but we could also resample for each individual within his or her conditional samples. Error residuals are sampled from the centered and shrinkage-corrected residuals as in NP bootstrap. More details on bootstrap algorithms and implementation are given in supplementary material.

2.3 Simulation study

Models: The model used for this simulation is a Hill (or sigmoid E_{max}) model, a standard model in dose-response studies where the effect of a drug in response to a dose d, $E(d)$ follows:

$$E(d) = E_0 + E_{max} \frac{d_i^\gamma}{d_i^\gamma + ED_{50}^\gamma} \tag{2}$$

It involves 4 parameters, the initial effect E_0, the maximum effect E_{max}, the concentration at which half the maximum effect is achieved EC_{50} and the sigmoidicity factor γ which controls the non-linearity of the model through the curvature. Interindividual variability was modelled through a log-normal distribution for all parameters, except for γ which was assumed to be the same for all subjects (no IIV). A correlation was simulated between E_{max} and EC_{50}, and we used a proportional error model with coefficient of variation σ. The parameters used in the simulation were the same as those used by Plan et al. [17], and are given in Table 1.

Table 1 Parameters used in the simulation for the E_{max} model (see [17]).

Parameter	*Value*	*Parameter*	*Value*
E_0 (-)	5	$\omega^2_{E_0}$	0.09
E_{max} (-)	30	$\omega^2_{E_{max}}$	0.49
ED_{50} (mg)	500	$\omega^2_{EC_{50}}$	0.49
γ (-)	1 or 3	$\mathrm{cov}(E_0, E_{max})$	0.245
σ (-)	0.1		

Scenarios: We investigated different scenarios to assess the influence of non-linearity, with $\gamma = 1$ and 3, and design (Table 2). For each value of γ we considered two designs, a rich design with N=100 subjects given 4 doses each (0, 100, 300, 1000), and a sparse design with N=200 subjects divided in 4 groups of 50 subjects given 2 doses each among the following 4 combinations (0,1000),(100,1000),(0,300) and (100,300). To evaluate the performance of bootstraps with a smaller sample size, the two scenarios with a Hill model were also evaluated with N=20 subjects (rich design) and 20 subjects per group (sparse design).

Table 2 Simulation scenarios.

Simulation scenarios	*Value of γ*	*Design*	*Number of subjects*
$S_{Emax,R}$	1	rich	100
$S_{Emax,S}$	1	sparse	200
$S_{Hill,R}$	3	rich	100
$S_{Hill,S}$	3	sparse	200
$S_{Hill,smallR}$	3	rich	20
$S_{Hill,smallS}$	3	sparse	80

For each scenario, we simulated K=200 datasets. Bootstrap distributions were obtained with B=100 bootstrap samples for each bootstrap. Additional simulations with up to B=500 bootstrap samples were used to assess the number of bootstrap samples necessary.

Bootstrap evaluation: The bootstraps were evaluated in terms of coverage rate, and we also computed the bootstrap mean $\hat{\theta}_{B,k}$, the bootstrap standard error (SE) and the bootstrap confidence interval (CI) for each estimated parameter from the bootstrap distributions. The bootstrap parameters estimates and their SE were defined as unbiased with relative error within $\pm 5\%$, moderately biased (from $\pm 5\%$ to $\pm 10\%$) and strongly biased ($> \pm 10\%$). The coverage rate of the 95% CI was considered to be good (from 90% to 100%), low (from 80% to 90%) and poor ($< 80\%$). A good bootstrap was defined as a method without bias for parameter estimates and their corresponding SE, and ensuring a good coverage rate of the 95% CI.

All computations and simulations were performed in R [20] using the development version of the saemix library [21]. Details of the computations are given in Supplementary Material.

3 Results

3.1 Bootstrap distributions

As an illustration, Figure 1 shows the distribution of the parameters obtained for one randomly selected simulated dataset obtained with the different bootstraps in scenario $S_{Hill,R}$ (Appendix B shows example of simulated data). For each parameter, we plot in red the true value (used in the simulation) and in blue the value estimated by saemix (for E_0 and ω_{Emax} the two values are almost identical so the lines are overlayed). In this example the distributions are for the most part similar, but we notice some discrepancy for ω_{Emax} where the NP bootstrap is shifted to the left, and in some cases the distribution appears more peaked, usually with Par or with Case.

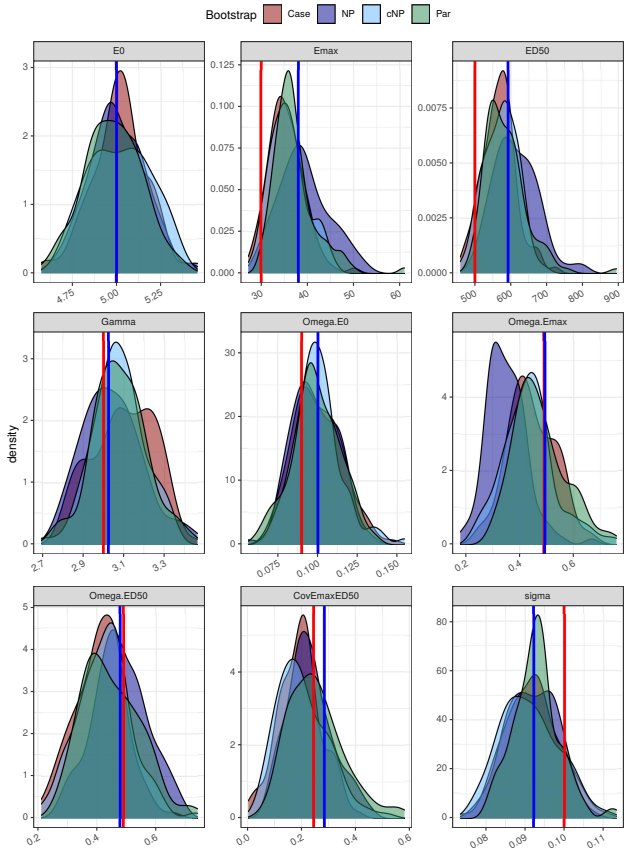

Fig. 1 Bootstrap distributions obtained for one randomly sampled distribution in scenario $S_{\mathrm{Hill,R}}$. The red line in each plot represents the true value of the parameter and the blue line the value estimated by saemix.

3.2 Bootstrap performances

Figure 2 shows the total relative bias on parameter estimates and their SE for the 6 scenarios simulated in this study. Dashed lines delineate absolute relative biases within 5% and dotted lines absolue relative biases within 10%. To make the graph easier to read, each line of the plot, representing one scenario, has its own scale on the Y-axis because of the difference in information between the different designs. The results from the estimation by saemix on each dataset are shown in grey, and the SE in this case are obtained using the asymptotic FIM, and the bootstrap distributions in different colours. The plots in the left column, representing the relative bias on the fixed and random parameters in the model, show a similar pattern across the different bootstraps: all the fixed parameters are estimated without bias across all scenarios except $S_{Hill,smallS}$ which had the least information. Concerning random effects, the covariance term seems most sensitive to a sparse design. For the rich designs (scenarios $S_{Emax,R}$ and $S_{Hill,R}$), the asymptotic method underestimated the SE for the random effects in $S_{Emax,R}$ and for nearly all the parameters in $S_{Hill,R}$ with biases as high as -30%. With the rich design however, the bootstraps manage to compensate some of the bias and reduce it to the 10% range. Additional results suggest that biased bootstrap estimates are associated with biased estimates on the original dataset (Appendix C).

With sparse designs ($S_{Emax,S}$ and $S_{Hill,S}$), although the bootstraps remain biased for the estimation of the SE, they correct in part the high bias of the asymptotic method. Finally, the last two lines show the scenarios with a small number of subjects, where all methods fail to estimate properly the parameters and their SE, although there again most of the bias on the parameters is in fact due to poor estimates by saemix.

Comparing the conditional non-parametric bootstrap (cNP) with the non-parametric bootstrap (NP), we note that cNP behaves generally similarly to NP in the first three scenarios in terms of parameter estimates, while it performs slightly better in $S_{Hill,S}$ (sparse design, higher model non-linearity) and in the scenarios with less subjects except for the estimation of the covariance term in scenario $S_{Hill,smallR}$. The estimates of the SEs are also on the whole slightly more accurate with cNP.

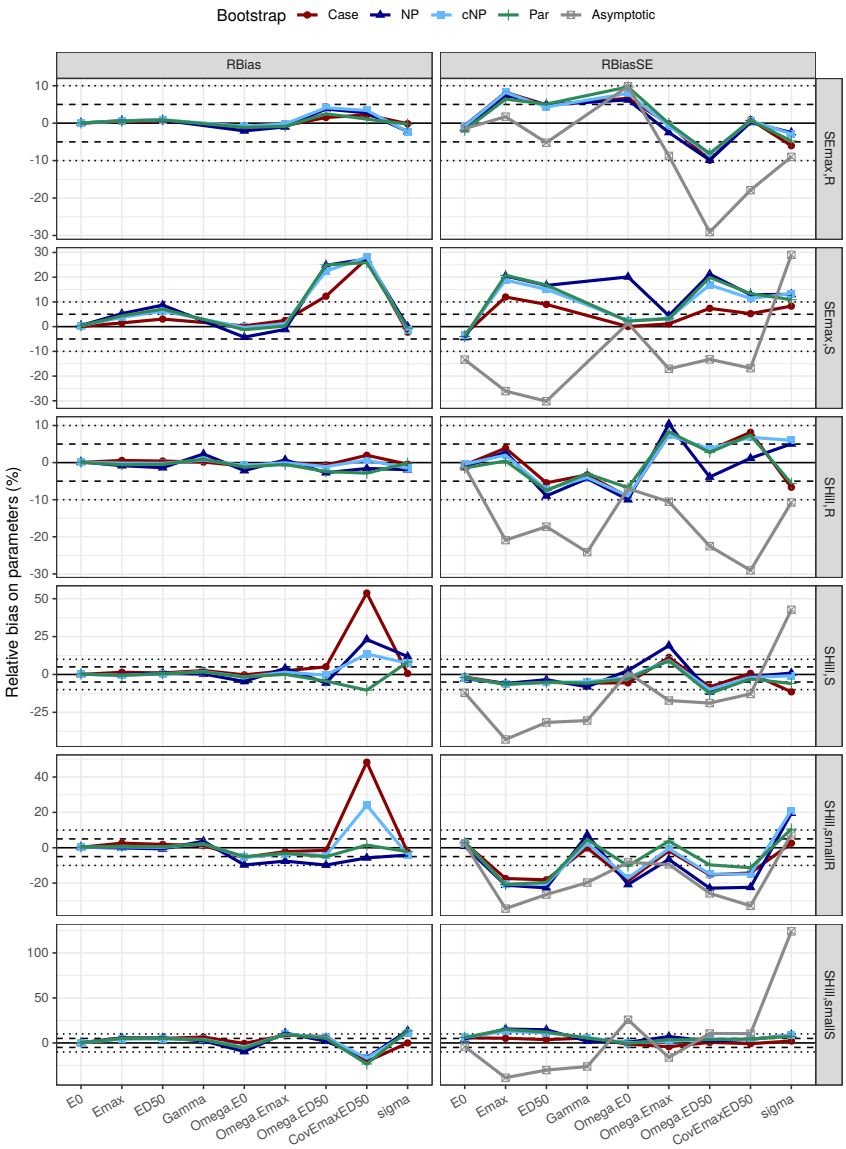

Fig. 2 Relative bias on parameters (left) and SE (right) for the 6 scenarios simulated in this study. Dotted lines delineate absolute relative biases within 5% and dashed lines absolue relative biases within 10%. Each line of plot has its own scale on the Y-axis.

In Figure 2 we used the standard deviation of the bootstrap distribution to estimate the SE, however the bootstrap is a non-parametric approach designed to produce the distribution of the estimator, and this distribution may not be symmetrically distributed. Therefore in Figure 3 we plot the coverage for each parameter in the 6 scenarios (rich designs in the left column and sparse designs in the right column), using the 4 bootstrap distributions and the asymptotic confidence intervals.

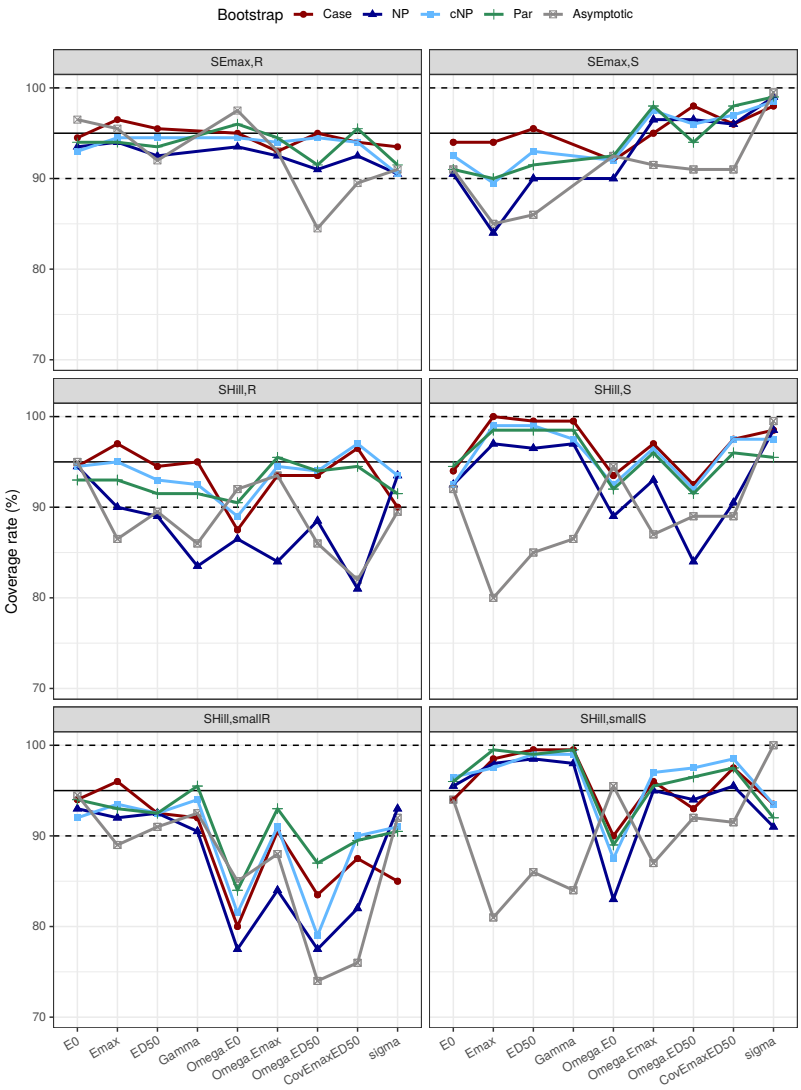

Fig. 3 Coverage rates for the 6 scenarios simulated in this study. Dotted lines indicate a coverage of 90 and 100%.

This plot shows good coverage for the case, parametric, and conditional non-parametric bootstraps in the first four scenarios, while the asymptotic method struggles with the non-linearity in the Hill model (see also Appendix D and E). cNP performs better than NP in all scenarios investigated, with coverage rates closer to the nominal value (a comparison between the coverage rates for these two approaches is given in Figure S10, Appendix I). Similar bias and coverage rate were found for σ in both bootstrap approaches, which is consistent with using the same correction in the cNP as in the NP bootstrap. To complement the coverage itself, Appendix F shows a figure showing the percentage of cases in which the true parameter is either lower or higher than the confidence interval. This figure indicates that apart from very few cases, lower coverage is due to the confidence interval being too low with respect to the true value of the parameter. We also note in $S_{Hill,S}$ and $S_{Hill,smallS}$ that except for E_0, the bootstraps for fixed effects tend to yield coverages close to 100%, suggesting overly wide confidence intervals. To investigate this further we also computed 50% CI (see Appendix G) and found that indeed, coverages could go as high as 70% in particular in $S_{Hill,S}$ and $S_{Hill,smallS}$ for these parameters, with none of the bootstraps appearing to have a clear advantage.

Concerning runtimes, to obtain 100 bootstrap samples for one dataset from the SHill,R scenario, parametric and standard residual bootstraps took 5% longer than the case bootstrap, while cNP took 40% longer. Most of the additional burden for the cNP bootstrap is however due to the step involving the estimation of the conditional distribution in the current, non optimised, version of the code, and the conditional distribution is normally estimated to perform diagnostic graphs after an saemix run. Finally, additional results concerning the influence of the number of bootstrap samples on the performances are given in Appendix H.

4 Discussion

The aim of the present article was to propose a new bootstrap approach for non-linear mixed effect models, and evaluate it by comparison to three bootstrap methods previously recommended [15]. We use samples from the conditional distributions proposed by Lavielle and Ribba [16] to bypass the issue of shrinkage for non-parametric bootstrap.

In the present article, we use a simple simulation example based on a sigmoid E_{max} model as a proof of concept to evaluate this new bootstrap, on the assumption that bootstrap methods should provide good results in simple cases first if they are to be trusted with more complex situations. In our simulations, the conditional non-parametric bootstrap showed good coverage in different scenarios, and performed better than the standard non-parametric bootstrap. Its performance was comparable to the case and parametric bootstraps, which showed good performance in a simulation setting with a homogenous and informative experimental design, in line with previous studies [9, 15]. A question arising at the time was whether it would be possible to improve the correction

of the residuals before resampling, to benefit from the non-parametric bootstrap's ability to maintain the structure of the original dataset, while being less dependent on model assumptions than the parametric bootstrap. That question echoed similar interrogations around the usefulness of graphical and statistical diagnostics for NLMEM [22, 23] in designs with low information where the empirical Bayes estimates exhibit high shrinkage. Recently, Lavielle and Ribba [16] proposed new diagnostic tools which use the conditional distribution of individual parameters, obtained as a product of stochastic estimation methods that have been gaining increasing popularity in the field of pharmacometrics and beyond. These conditional distributions allow to bypass the issue of shrinkage, and it was therefore a natural step to consider using these distributions to sample from. The individual conditional distributions are a by-product of the estimation in the SAEM algorithm, and the mode or mean of which are used as estimates of the individual parameters. With linearisation-based algorithms, these individual parameters are obtained by minimising a Bayesian criterion, but additional computations could be used to derive the same distributions. Sampling individual parameters from the individual conditional distributions $p(\psi_i|y_i)$ generates a sample of parameters representative of the population distribution $p(\psi_i)$ [16]. Diagnostic graphs using these samples, for example of parameters versus covariates, do not suffer from shrinkage. When we use these resampled parameters to build the bootstrapped datasets in the cNP bootstrap, we therefore expect to generate datasets representative of the true interindividual variability. In unbalanced designs, it might also be worth considering resampling the random effects within each subjects conditional distributions; indeed, the posterior distributions capture the uncertainty on individual parameter estimates, reflecting the informativeness of the individual sampling design which differs according to the number and position of the samples. As such, our cNP bootstrap retains the attractive feature of the NP bootstrap of preserving the structure of the original dataset with an unchanged number of samples and subjects, and combines it with a more appropriate distribution for the random effects. An extension for future work will therefore be to investigate unbalanced designs, where some subjects may have more information than others, such as the design mixing rich and sparse data in [15], where the non-parametric bootstrap greatly overestimated the SE for some variance parameters but a stratified version led to a much lower coverage. By using conditional distributions instead of theoretical distributions as in the parametric bootstrap, we could also expect the cNP to be more robust to misspecifications in the random effects. This will however need to be tested in a further study, as we show here that bootstraps cannot recover from very poor asymptotic estimates. Also, conditional distributions involve model assumptions on the distribution of the random effects, and their properties have not been studied fully yet, especially in small samples.

For the residual errors ϵ_{ij}, we used the same correction in cNP as in the NP bootstrap, inflating the residuals so that their overall variability matches the population estimate of the residual error. However, this is not entirely satisfactory as the $\hat{\epsilon}_{ij}$ are also affected by shrinkage, called ϵ-shrinkage by

Karlsson and Savic [22], and a global level correction incurs similar penalties as for η-shrinkage. This could be the reason why cNP and the NP bootstraps gave similar results for σ. In the present study, the designs were homogenous in terms of informativeness, but it would be interesting to evaluate the difference between cNP and NP, in unbalanced designs. A further extension to the cNP will be to compute the residual errors for the conditional samples drawn from the individual conditional distribution of the parameters, and resampling from these to reflect the individual residual variability.

Although the bootstrap is often branded as an approach to gain a good understanding of the uncertainty in small samples, Broeker and Wicha have shown that this is mostly wishful thinking as the performance in small datasets is not good [24]. We confirm this here in the two scenarios $S_{\mathrm{Hill,smallR}}$ and $S_{\mathrm{Hill,smallS}}$. With even lower number of subjects, especially in the sparse design, the bias on the parameter estimates for the asymptotic method was even more apparent (data not shown) and the bootstrap approaches could not recover from these poor estimates. With these scenarios the coverage for all bootstraps was lower than 90%, and there was no clear trend in whether the bootstraps under- or over-estimated the parameters and their SE, which is a cautionary message for the usage of bootstrap in practice as from present and previous work [15, 24] the conditions in which bootstrap performances are adequate are not obvious. In our simulations, we found that the performance of the bootstraps degrade as the informativeness of the design decrease. Tools like PFIM [25] can be used to evaluate the design before running a bootstrap approach, to assess whether a bootstrap procedure is appropriate.

A final point is the high computational effort incurred. Even the simplest version, the case bootstrap, requires a large number of runs; its theoretical shortcomings, such as the difference in information between bootstrapped datasets in unbalanced designs, or the difficulty in handling covariate distributions when they enter a model, have caused other even more computationally intensive bootstrap algorithms to be developed [9]. These bootstraps require repeated model evaluations on top of the sampling step, making them somewhat slower. However, these approaches only need to be run for key models such as the final model in an analysis. On top of this, an additional source of uncertainty comes from the model itself, as a model is rarely an isolated object but an element in model building; methods like model averaging or sensitivity analyses should also be used to gain a full understanding of uncertainty [26].

Conclusion

Our new conditional residual bootstrap for NLMEM improved the coverage rate compared to the classical non-parametric residual bootstrap, proving nearly as efficient as the parametric bootstrap, which in these simulations where we used the true distributions for the parametric bootstrap can be considered as a gold standard. The case bootstrap also performed well, and all three bootstraps provided more appropriate estimates of the SE than the

asymptotic method in sparse designs. The bootstrap estimates of the parameters and their SE showed bias in the presence of non-linearity and sparse information, highlighting the importance of considering the entire bootstrap distribution for uncertainty instead of using a normal approximation with mean and standard deviation derived from the distribution.

Funding

This work was performed during the authors' research time and no other funding was used.

Conflict of interest

None.

Author Contributions

Conceptualisation, E.C.; formal analysis and simulations, E.C. and C.R.; methodology, M.U.. and E.C.; software: E.C. and C.R.; visualisation, E.C. and M.U.; writing, original draft, C.R. and E.C.; writing, review and editing, E.C., M.U. C.R. and V.J.

Data Availability

All data generated or analysed during this study are included in this published article and its supplementary information files (see Appendix A) and are available on Zenodo (`https://doi.org/10.5281/zenodo.4059718`).

References

1. Lalonde RL, Kowalski KG, Hutmacher MM, Ewy W, Nichols DJ, Milligan PA, et al. Model-based drug development. Clin Pharmacol Ther. 2007;82:21–32.
2. Wang Y, Zhu H, Madabushi R, Liu Q, Huang SM, Zineh I. Model-informed drug development: current US regulatory practice and future considerations. Clinical Pharmacology & Therapeutics. 2019;105(4):899–911.
3. Desmée S, Mentré F, Veyrat-Follet C, Sébastien B, Guedj J. Nonlinear joint models for individual dynamic prediction of risk of death using Hamiltonian Monte Carlo: application to metastatic prostate cancer. BMC Med Res Methodol. 2017;17:105.

4. Retout S, Mentré F, Bruno R. Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. Statistics in Medicine. 2002;21(18):2623–2639.
5. Gelman A, Carlin J, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. London: Chapman & Hall /CRC Press; 2013.
6. Dartois C, Lemenuel-Diot A, Laveille C, Tranchand B, Tod M, Girard P. Evaluation of uncertainty parameters estimated by different population PK software and methods. J Pharmacokinet Pharmacodyn. 2007;34:289–31.
7. Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. Biometrics. 1990;46:673–87.
8. Dosne AG, Niebecker R, Karlsson MO. dOFV distributions: a new diagnostic for the adequacy of parameter uncertainty in nonlinear mixed-effects models applied to the bootstrap. Journal of Pharmacokinetics and Pharmacodynamics. 2016;43:597–608.
9. Thai H, Mentré F, Holford NH, Veyrat-Follet C, Comets E. A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. Pharmaceutical Statistics. 2013;12:129–40.
10. Efron B. Bootstrap methods: another look at the jacknife. Annals of Statistics. 1979;7:1–26.
11. Kümmel A, Peter L Bonate JD, Krause A. Confidence and Prediction Intervals forPharmacometric Models. CPT Pharmacometrics Syst Pharmacol. 2018;7:360–73.
12. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1994.
13. Das S, Krishen A. Some bootstrap methods in non-linear mixed-effects models. Journal of Statistical Planning and Inference. 1999;75:237–45.
14. Carpenter JR, Goldstein H, Rasbash J. A novel bootstrap procedure for assessing the relationship between class size and achievement. Applied Statistics. 2003;52:431–43.
15. Thai H, Mentré F, Holford NH, Veyrat-Follet C, Comets E. Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics. Journal of Pharmacokinetics and Pharmacodynamics. 2014;41:15–33.
16. Lavielle M, Ribba B. Enhanced method for diagnosing pharmacometric models: random sampling from conditional distributions. Pharmaceutical Research. 2016;33:2979–88.
17. Plan E, Maloney A, Mentré F, Karlsson M, Bertrand J. Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for dose-response models. The AAPS Journal. 2012;14:420–32.
18. Lavielle M. Mixed effects models for the population approach: models, tasks, methods and tools. Boca Raton, FL: Chapman & Hall CRC Biostatistics Series; 2014.
19. Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics and Data Analysis. 2005;49:1020–

38.
20. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2006. ISBN 3-900051-07-0. Available from: `http://www.R-project.org`.
21. Comets E, Lavenu A, Lavielle M. Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. Journal of Statistical Software. 2017;80:1–41.
22. Karlsson M, Savic R. Diagnosing model diagnostics. Clin Pharmacol Ther. 2007;82:17–20.
23. Combes FP, Retout S, Frey N, Mentré F. Powers of the likelihood ratio test and the correlation test using empirical bayes estimates for various shrinkages in population pharmacokinetics. CPT: Pharmacometrics and Systems Pharmacology. 2014;3:e109.
24. Broeker A, Wicha SG. Assessing parameter uncertainty in small-n pharmacometric analyses: value of the log-likelihood profiling-based sampling importance resampling (LLP-SIR) technique. Journal of Pharmacokinetics and Pharmacodynamics. 2020;47:219–28.
25. Dumont C, Lestini G, Le Nagard H, Mentré F, Comets E, Nguyen T, et al. PFIM 4.0, an extended R program for design evaluation and optimization in nonlinear mixed-effect models. Computer Methods and Programs Biomedicine. 2018;156:217–29.
26. Buatois S, Ueckert S, Frey N, Retout S, Mentré F. Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. AAPS J. 2018;20:56.