Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Conditional bootstrap for non-linear mixed effects models

Sofia Kaisaridi
Infection, Anti-microbiens, Modélisation, Evolution
Moreno Ursino
Heka : Méthodes basées sur les données et modèles en santé pour l'acquisition de connaissances
Emmanuelle Comets
emmanuelle.comets@inserm.fr

Infection, Anti-microbiens, Modélisation, Evolution

Research Article

Keywords: Mixed model, SAEM algorithm, Conditional distribution, Uncertainty, saemix, Unbalanced design

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11398404/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Conditional bootstrap for non-linear mixed effects models

Sofia Kaisaridi (ORCID: 0000-0003-2110-6392)[1],
Moreno Ursino (ORCID: 0000-0002-5709-4322)[2],
Emmanuelle Comets (ORCID: 0000-0002-9150-9886)[1,3*]

[1*]Université Paris Cité & Université Sorbonne Paris Nord, INSERM, IAME, Faculté de Médecine Bichat-Claude Bernard, 16 rue Henri Huchard, Paris, 75018, France.
[2]Inserm, Centre de Recherche des Cordeliers, Sorbonne Université, Université Paris Cité, Paris, 75006, France.
[3]Université de Rennes, Inserm, EHESP, Irset - UMRS 1085, Rennes, 35000, France.

*Corresponding author(s). E-mail(s): emmanuelle.comets@inserm.fr;
Contributing authors: sofia.kaisaridi@icm-institute.org;
moreno.ursino@inserm.fr;

Abstract

Uncertainty in non-linear mixed effect models is often assessed using the Fisher information matrix to derive the standard errors of estimation. The bootstrap is an alternative to the asymptotic method, with different approaches to handle the different levels of individual and population variabilities.

We propose here a non-parametric bootstrap, cNP, resampling from the conditional distribution of the individual parameters, preserving the structure of the original data. cNP was implemented in the saemix package for R along with the case, parametric (Par), and non-parametric (NP) residual bootstraps. Coverage rates were compared in a simulation study using sigmoid Emax models, with rich, sparse and unbalanced designs, and 3 levels of residual variability.

The asymptotic method tended to produce lower than theoretical coverages for the variance terms. Bootstraps provided more adequate coverage, but none of the approaches maintained coverage when the residual error increased. Overall, the new cNP and the Case provided better coverage than the classical NP.

The new conditional non-parametric bootstrap could be used when it is important to preserve the structure of the original dataset, such as the number of observations or the repartition of covariates as it does not require stratification.

Keywords: Mixed model, SAEM algorithm, Conditional distribution, Uncertainty, saemix, Unbalanced design

Study Highlights

What is the current knowledge on this topic ?

Various bootstrap methods exist in non-linear mixed effect models. The simplest method is the Case bootstrap, resampling entire individuals. Stratifying is recommended to handle very unbalanced designs or when multiple covariates enter the model, but may be difficult in practice. Other bootstrap approaches resample at the level of the residuals, and may be more suitable in this case.

What question did this study address ?

This study proposes a non-parametric residual bootstrap which preserves the structure of the original data. By resampling from individual conditional distributions, we avoid the need to normalise the empirical residuals.

What does this study add to our knowledge ?

The case bootstrap was robust even in unbalanced designs. The Case and cNP bootstraps performed best, but the asymptotic method and all the bootstraps suffered strongly when the residual error increased.

How might this change drug discovery, development and/or therapeutics ?

The new conditional non-parametric bootstrap could be used when it is important to preserve the structure of the original dataset, such as the number of observations or the repartition of covariates as it does not require stratification.

1 Introduction

In pharmacological and medical studies multiple observations are generally collected from each individual followed during the trial, such as drug concentrations in pharmacokinetic (PK) studies, or biomarkers linked to disease progression or treatment effect. Hierarchical mixed effects models are used to simultaneously describe the evolution of the underlying physiological and pathological processes and handle different levels of variability, both within observations from the same individual (residual variability) and between individuals (between-subject variability). Drug development makes frequent use of non-linear mixed effects models (NLMEM) to analyse pharmacokinetic and pharmacodynamic data (Pillai et al. (2005)). The results from these models are then used to help design the next stages of clinical decisions and may even influence go/no-go decisions (Lalonde et al. (2007)). Informed decisions need to take into account the probability of success, and therefore incorporate uncertainty at different levels, including the uncertainty associated to parameter estimation. In NLMEM, the standard approach to compute estimation errors is an asymptotic approximation based on the Fisher Information Matrix, the inverse of which has been shown to be a lower bound on the standard errors of estimation (SE) of the model parameters (Davidian and Giltinan (1995)). The validity of this approximation requires that both the number of subjects and the number of observations in each subject tend to infinity (Vonesh and Carter (1992)), which in practice is seldom verified. Alternatively, bootstrap (Efron (1979)), log-likelihood profiling and Sampling Importance Resampling (SIR) (Dosne et al. (2017)) have been proposed.

The bootstrap was initially proposed as a widely applicable jackknife approach to estimate biases and SE (Efron (1979)) but then took on the more ambitious goal of building non-parametric confidence intervals (Efron (2000)). It consists in repeatedly resampling the observed data with replacement and applying the original analysis to create the distribution of an estimator or a statistic of interest. The simplest method

is the Case bootstrap where the entire vector of individuals is resampled. Given the hierarchical nature of the NLMEM, different approaches have been advocated to ensure a proper treatment of the inter-individual and within individual variabilities (Das and Krishen (1999)). Thai et al. (2013, 2014) investigated different bootstrap approaches including the parametric bootstrap (Par), resampling residuals from a theoretical distribution, and the non-parametric bootstrap (NP) resampling from the estimated residuals based on the Empirical Bayesian Estimates (EBE) of the parameters. They found that all the bootstraps performed better than the asymptotic method in NLMEM, none showing a clear advantage. One issue with the NP bootstrap however is shrinkage due to the regression to the mean of individual parameters estimates towards population parameters (Karlsson and Savic (2007)), which led to degraded performances especially in very non-linear models, despite correcting the variance of the residuals to account for shrinkage (Carpenter et al. (2003)). To overcome this, Comets et al. (2021) proposed a residual non-parametric bootstrap using samples from the conditional distribution of the individual parameters (Lavielle and Ribba (2016)) to resample the random effects. The objective of the present study is to extend this preliminary implementation to a full conditional non-parametric bootstrap (cNP) also resampling the residual errors, since these are affected by ϵ-shrinkage (Karlsson and Savic (2007)).

For comparability, we will continue to use the simulation setting of Comets et al. (2021), inspired by Plan et al. (2012). In a first scenario, we implement the full cNP and compare the performances with respect to the Case, NP and Par bootstraps using the previous simulated data in rich (4 doses) and sparse (2 doses) scenarios. In a second scenario, we challenged the bootstraps with unbalanced designs where the amount of information differs across subjects. Finally, we evaluate the impact of increasing the residual variability.

2 Materials and methods

2.1 Statistical model

Our framework is non-linear mixed effect models (Lavielle (2014)), where we denote y_{ij} $(i = 1, \ldots, N, j = 1, \ldots, n_i)$ j^{th} observation in the i^{th} subject, ξ_{ij} the corresponding design variables (eg measurement times and dose amounts in a PK study) and z_i the individual covariates. For models describing the evolution of continuous data, we generally assume that the observations y_{ij} follow a normal distribution with a mean defined by a structural model f representing the longitudinal evolution and depending on the individual parameters ψ_i. The standard deviation is defined through an error model g involving residual error coefficients denoted by σ:

$$y_{ij} = f(\xi_{ij}, \psi_i) + g(\xi_{ij}, \psi_i, \sigma) \cdot \epsilon_{ij} \tag{1}$$

The residual errors ϵ_{ij} are i.i.d normally distributed random variables. Individual parameters ψ_i are assumed to follow a multivariate distribution with known shape and unknown parameters, through a transformation h of a linear function of the fixed effects μ, possibly covariate effects β and random effects η_i, which are assumed to be normally distributed with mean 0 and variance-covariance matrix Ω:

$$\psi_i = h(\mu, \beta, z_i, \eta_i) \text{ where } \eta_i \sim N(0, \Omega) \tag{2}$$

Standard models for g include $g = \sigma$ (constant variance), $g(\xi_{ij}, \psi_i, \sigma) = \sigma\, f(\xi_{ij}, \psi_i)$ (proportional variance), or a combination of both.

In a frequentist framework, which will be our setting here, we estimate the population parameters $\theta = (\mu, \beta, \Omega, \sigma)$. We use the iterative SAEM algorithm (Kuhn and Lavielle (2005); Lavielle (2014)), where during the stochastic approximation step the unknown parameters η_i representing the random effects are sampled from their

conditional distributions to create a complete dataset from which the conditional log-likelihood is derived by stochastic approximation. At the end of the estimation process, individual parameters are estimated as the mode or the median from the conditional distribution obtained, refined at the end of the fit by repeated sampling through an Metropolis-Hastings procedure.

Parameter uncertainty is usually assessed in the software performing maximum likelihood estimation by evaluating the Fisher Information Matrix at the point estimates $\hat{\theta}$ and using the inverse as the variance-covariance matrix yielding the standard errors. Various approximations have been proposed to ease the computation (Retout et al. (2002); Nyberg et al. (2015)), the most commonly used being a first-order approximation yielding a block structure with one block for the fixed effects and one block for the parameters associated with the variabilities. A normal approximation then gives standard α-level confidence intervals (CI) for the p^{th} component of θ, denoted $\theta^{(p)}$, as:

$$\theta^{(p)} \in \left[\hat{\theta}^{(p)} - z_{\alpha/2}\hat{SE}(\hat{\theta}^{(p)}); \hat{\theta}^{(p)} + z_{1-\alpha/2}\hat{SE}(\hat{\theta}^{(p)});\right] \tag{3}$$

where z_q is the q-quantile of the normal distribution. In the following, we will denote this approach as the Asymptotic method and drop the superscript (p) for easier reading.

2.2 Bootstrap methods

Bootstrapping generates multiple pseudo-samples mimicking the original sample from which we obtain a distribution of the statistic of interest. Denoting B the number of bootstrap samples, a general bootstrap algorithm is:

1. Generate a bootstrap sample by resampling from the data and/or from the estimated model
2. Compute the estimates of the parameters of the model for the bootstrap sample

3. Repeat steps 1-2 B times to obtain the bootstrap distribution of the parameter estimates.

The mean and standard deviation of the distribution of the bootstrap estimates $\hat{\theta}_b^*$ $(b = 1...B)$ can be viewed as the bootstrap estimates of the parameters and their SE. These could be used to derive a CI through the normal approximation as in equation 3, however the bootstrap distribution may be asymmetric contrary to the asymptotic estimate and it is more appropriate to compute the non-parametric or percentile CI, defined as the bootstrap α-level CI from the bootstrap distribution (Davison and Hinkley (1997)):

$$\theta \in \left[\hat{\theta}^*_{\frac{\alpha}{2}}; \hat{\theta}^*_{(1-\frac{\alpha}{2})}\right] \tag{4}$$

where $\hat{\theta}^*_{(\frac{\alpha}{2}\cdot B)}$ is the $\frac{\alpha}{2}$-quantile of the bootstrap distribution. Other approaches to construct CI, including Student or normal-type approximations as well as bias-corrected CI, are discussed in particular in Davison and Hinkley (1997) and MacKinnon (2006).

Different approaches have been proposed to generate bootstrap samples in NLMEM to reflect the true data generating process with repeated measures within a subject and take into account the two levels of variability. The case bootstrap (Case) resamples entire observation vectors by resampling individuals. Parametric (Par) and non-parametric (NP) bootstraps on the other hand resample random effects and residual errors to reconstruct an observation vector, preserving the structure of the original dataset (detailed algorithms are given in online Appendix 5). Here, we extend the partial conditional non-parametric residual bootstrap using samples from the conditional distribution (Comets et al. (2021)) to also resample the residual errors, yielding the following algorithm (cNP):

1. fit the model to the data to obtain $\hat{\theta}$ and estimate the conditional distributions
2. obtain M samples for each vector η_i, η_i^m $(m = 1, ..., M)$ from the conditional distributions

3. center the η_i^m for each subject and draw a sample $\{\eta_i^*\}_{i=1,\dots N}$ with replacement
4. calculate the residuals $\epsilon_{ij}^m = \frac{y_{ij} - f(x_{ij},\hat{\mu},\eta_i^m)}{g(x_{ij},\hat{\mu},\eta_i^m,\sigma)}$ and center them
5. draw a sample $\{\epsilon_{ij}^*\}$ with replacement globally from $\{\epsilon_{ij}^{*m}\}$
6. generate the bootstrap responses $y_{ij}^* = f(x_{ij},\hat{\mu},\eta_i^*) + g(x_{ij},\hat{\mu},\eta_i^*,\sigma)\ \epsilon_{ij}^*$

In the present study we performed the resampling of residuals in steps 3. and 5. separately for each subject within their own distribution, but in homogeneous populations it would also be possible to sample globally over the residuals obtained for the entire population.

2.3 Simulation study

We performed three sets of simulations to evaluate our new bootstrap approach. First, we applied the cNP to the previous simulations used in the assessment of the partial cNP for random effects in Comets et al. (2021) for consistency. Second, we investigated the influence of the amount of information on the uncertainty estimates by simulating unbalanced designs. Third, we increased the residual variability to study the balance between interindividual and residual variability.

Simulation model:

We used the model initially proposed by Plan et al. (2012) in a comparison of several algorithms, where different scenarios allowed to investigate the influence of non-linearity and design on the performance of estimation methods. The data follows a dose-response model:

$$f(x_{ij},\psi_i) = \mathrm{E}_{0_i} + \mathrm{E}_{max_i} \cdot \frac{x_{ij}^{\gamma}}{x_{ij}^{\gamma} + \mathrm{ED}_{50_i}^{\gamma}} \tag{5}$$

This model involves 4 parameters, the initial effect E_0, the maximum effect $\mathrm{E}_{\max}$, the concentration at which half the maximum effect is achieved ED_{50} and the sigmoidicity factor γ which controls the non-linearity of the model through the

curvature. The parameters were assumed to follow a log-normal distribution, except for γ, which was assumed to be the same for all subjects (no variability), so that $\psi_i = \{\mathrm{E}_{0,i}, \mathrm{E}_{max,i}, \mathrm{ED}_{50,i}, \gamma\}$. A covariance was simulated between $\mathrm{E}_{\max}$ and ED_{50}, and a proportional error model with coefficient σ was used. Two values of γ were used, $\gamma = 1$ ($\mathrm{E}_{\max}$ model) and $\gamma = 3$ (Hill model) to evaluate how model non-linearity affected the results. The values of the parameters used in the simulations are presented in Table 1.

Simulation scenarios:

In the first scenario, we used the datasets already simulated for the paper by Comets et al. (2017) (available online in a Zenodo repository), which included a rich (4 doses) and sparse (2 doses) design for each of the two models ($\mathrm{E}_{\max}$ with $\gamma = 1$ and Hill with $\gamma = 3$). In the second scenario, we considered unbalanced designs representing situations where some subjects receive a full set of doses while others receive only low, only high, or a mix of doses, as well as a mix of sparse (2 doses) and rich (4 doses) design, as illustrated in Figure 1. The third scenario used the same settings as the first scenario but increasing σ to 0.3 then 0.5. Detailed description of the designs are given in online Appendix 5. For each dataset and method, $B = 200$ bootstrap samples were used. This number could be increased to better approach extreme quantiles (Austin and Leckie (2020)).

Bootstrap evaluation:

For each simulated dataset $k = 1, 2, ..., K$ we computed the following statistics of interest obtained from B bootstrap samples by the four methods, for each estimated parameter: the bootstrap mean $\hat{\theta}_{B,k}$, the bootstrap standard error $\widehat{SE}(\theta)_{B,k}$, obtained as the standard deviation of the bootstrap distribution $\widehat{SE}(\theta)_{B,k} = SD(\hat{\theta}_{B,k})$, and the bootstrap CI obtained using the quantiles of the bootstrap distribution.

Our main evaluation metric was the coverage rate for the 90% CI obtained by the four bootstraps and the Asymptotic method, calculated using the SE estimated by saemix. The coverage rate of a CI of parameter is the proportion of times that it contains the true value of the parameter, reflecting the accuracy of each approach to estimate uncertainty. The 90% CI ($\alpha = 0.1$) was chosen to allow visualising both undercoverage and overcoverage. We also report the results for the more standard 95% CI ($\alpha = 0.05$) in online Appendix 5. Fluctuations due to the simulations was assessed by computing the SE related to the Monte Carlo procedure as suggested by Morris et al. (2019), and used to choose the number of simulations (K).

$$MC.SE = \sqrt{\frac{E(coverage)(1 - E(coverage))}{K}} \tag{6}$$

Here we used $K = 200$ for an expected MC.SE of approximately 2%.

2.4 Implementation

Simulations were performed in R (R Development Core Team (2006)) using R version 4.0.4(2021-02-15). The parameters of the models were estimated using the saemix library (Comets et al. (2017)) (version 3.2). The default options were used to run the SAEM algorithm; the initial parameters for the fixed and random effects were set to the true value in all runs as the objective was not to evaluate the estimation method but the bootstrap performances. The results were analysed in R. The runs were set up using a laptop and the simulation studies were run on the *"Centre de Biomodélisation IAME"* (CATIBioMed cluster, Bichat hospital, Paris). For the evaluation of the designs we used the R package PFIM 4.0 (Dumont et al. (2018)) available on the PFIM website. where we evaluated the scenarios presented in this study, minus the covariance term as the software does not allow a full covariance

matrix. The simulated datasets used in this study are available on Zenodo (repositories: 10.5281/zenodo.4059718 and 10.5281/zenodo.10779192).

3 Results

3.1 Original designs

In Figure 2 we plot the coverage rates of 90% CI as points, along with their MC uncertainty (as errorbars) for the four designs in the original simulation scenario. Non-parametric CI are presented with a different colour for each method. We observe similar patterns as in the previous analysis for the 95% CI (Comets et al. (2021)), with the asymptotic method showing significant undercoverage for several parameters. The sparse design with a Hill model ($S_{Hill,S}$) was the most challenging case, with all bootstraps failing to get good coverages for at least one parameter. In this scenario the case bootstrap showed the most overcoverage, which wasn't apparent previously with a 95% CI.

Our proposed cNP performs better than the classical NP bootstrap across the scenarios and as well as the Case and the Parametric bootstrap, although the CI for the fixed effects are larger in $S_{\text{Hill,S}}$ scenario. The NP underestimates the coverage for some parameters in every design, especially in $S_{Hill,R}$. Supplementary C shows plots of the bias with respect to the true values for the different approaches (left column), and we can see that for the two rich designs the estimation by saemix is unbiased for all parameters, suggesting that the poor performance with NP is not due to biased estimates of the population parameters. The plots in the right hand column in Figure 6, which compare the SD of the bootstrap distributions to the empirical SE, show that the SE tend to be estimated with a bias around ±10% (depending on the parameter) which could indicate that the NP has more difficulties in narrowing on the extreme quantiles, while the cNP may be able to correct for this using samples accepted by the Metropolis-Hastings algorithm. The same plots provide a possible explanation for the

undercoverage observed for the Asymptotic method, as they show that this method consistently underestimates the SE for most designs.

3.2 Unbalanced designs

Figure 3 shows the same coverage rate with the Hill model for the four unbalanced designs (see Supplementary D for the E_{max} scenarios, and the 95% coverage rates under both models).

Across all unbalanced designs and as previously, the asymptotic method showed some undercoverage, but less important probably due to the designs being more informative (see Appendix D). The bootstraps generally had adequate coverage, except for the NP for the design $S_{Hill,end}$. The Case bootstrap in this case was not stratified as the results were already good without stratification. To investigate further we plotted the relative bias on the parameters and their SE compared to the true values (figure 11, Appendix D) and found that for this design, both the Par and NP bootstrap showed significant bias for the random effects both for the parameter estimate and for the SE, while the Case and the cNP bootstraps were within ±10%. This trend was apparent across all designs, with Par and NP presenting more bias in the unbalanced designs compared to the original designs.

3.3 Increased error

Figure 4 contrasts the coverage rates for the 90% CI in the rich and sparse Hill designs (see Appendix E for additional results) from the original scenario ($\sigma = 0.1$, left) with the scenarios with increased error coefficient $\sigma = 0.3$ and $\sigma = 0.5$ (middle and right columns). The undercoverage already observed in figure 2 for the coverage rate of fixed effects by the Asymptotic method becomes more pronounced as σ increases, affecting all parameters except E_0. Figure 17 shows that this may again be linked to underestimation of the SE, and is also logically higher in the sparse design. The

behaviour of the NP boostrap is somewhat inconsistent, sometimes correcting the poor coverages observed with $\sigma = 0.1$ and sometimes inflating them.

For the variance parameters, in the rich design the asymptotic method again underestimates the SE leading to undercoverage, but this time we see a marked degradation of both the NP and more surprisingly of the Par bootstrap as σ increases. The Case and the cNP perform generally better but show undercoverage for ω_{E_0} and σ. At the highest level of residual error tested here all bootstraps provided very poor coverage for σ, with a largely underestimated uncertainty for this parameter. The same pattern can be seen for the sparse design, but there the asymptotic method only shows a slight overestimation for the random effects with less impact on the SE, and as a consequence the coverage rate are adequate. The plots of bias compared to the true parameters in Appendix E in both the rich and sparse designs show that both the NP and the Par bootstraps tend to amplify the bias from the original estimates, which are used to calibrate the theoretical distributions for Par and as priors for NP, and can explain why these bootstraps behave poorly, while both the Case and cNP rely more strongly on the data and can partly correct for the bias.

4 Discussion

The aim of this study was to extend the conditional non-parametric bootstrap approach for NLMEM proposed in Comets et al. (2021) by also resampling the residual error, and evaluate this full cNP by comparison to three bootstrap methods previously used in NLMEM (Thai et al. (2014)), case (also called pairs bootstrap), non-parametric and parametric bootstraps. We also investigated the influence of an imbalance in the number of samples across subjects and of the magnitude of the residual error. Coverage rates were our main evaluation metric since the bootstrap is a non-parametric approach, and the bootstrap distributions may deviate from the Gaussian distribution obtained through a normal approximation. Additional results are given in Appendices

for the bias on parameters using the mean value of the bootstrap distribution as the bootstrap estimates, and the comparison between the empirical SE and the standard deviation of the bootstrap distribution which is often used as an approximation to the SE.

The cNP uses the conditional distributions, obtained as a by-product of the estimation process in the SAEM algorithms (Lavielle and Ribba (2016)). Samples from the conditional distribution are used to produce informative diagnostics by avoiding shrinkage when the number of individual samples is small. Here the comparison between cNP and NP shows that they allow to generate more representative bootstrap datasets than using empirical individual residuals with a population-level correction (Carpenter et al. (2003); Morris (2002)). This could mean that the exchangeability assumption, a core bootstrap principle, is better respected (Davison and Hinkley (1997)), as the cNP resamples residuals within their own conditional distributions which have equal sampling probability.

A major motivating factor behind the development of the cNP was its ability to generate bootstrap samples with the same design as the original dataset, since resampling at the level of the residuals preserves the original design and covariate structure without needing to stratify over variables of interest. Case bootstrap, widely used for its simplicity and general applicability, requires adjusting test statistics and does not preserve the design (MacKinnon (2006)). Here, when we simulated unbalanced designs with a varying amount of information per subject, both cNP and Case estimated the parameters with low bias, except in $S_{Hill,low}$ for the covariance term, where it was limited enough to avoid affecting the coverage. In general, all bootstraps performed well in the unbalanced designs, and the Case bootstrap did not need a stratification to maintain appropriate coverages. The predicted RSE (online Appendix 5.1), which we used during the design of the simulation study to ensure parameters were identifiable, suggest that the unbalanced designs were probably in fact not challenging enough,

especially when we compare to the predicted RSE from the original design. Stratification could be explored in scenarios involving covariates in the model, but our initial results suggest that the Case bootstrap remains robust.

All bootstraps however failed to some degrees in the final simulation setting with high σ. The plots of bias shown in the online Appendices show an increased bias for the variance parameters, most affected by the undercoverage, indicating that the bootstrap distributions move away from the true value. More generally, bootstrap methods find it difficult to recover from poor estimates on the original datasets, and we even found that Par and NP amplify the bias in the asymptotic method (see for example fig 16 and 15 in online Appendix 5.1). This corroborates other works showing that bootstrap approaches don't perform well in small samples, where there is not enough information to estimate variabilities: in Broeker and Wicha (2020), the Case bootstrap was implemented and performed very poorly in small samples, especially compared to an approach combining univariate log-likelihood profiling and Sampling Importance Resampling (Dosne et al. (2017)) or a full Bayesian approach. Alternatively, a hybrid approach using Bayesian sampling from a frequentist prior using a Metropolis-Hasting algorithm coupled with the SAEM approach was implemented recently in saemix and shown to perform well with small sample sizes Guhl et al. (2024).

In the present paper with larger sample sizes, the Case and cNP generally performed similarly and were least sensitive in this scenario, so that it would be interesting to compare the cNP in the setting of Broeker and Wicha (2020). The poor performance of Par was somewhat unexpected, as it samples in a distribution with the true shape, but it indicates a high sensitivity to biased estimates from the estimation on the original data, even amplifying it as the bootstrap estimates drift further away from the asymptotic estimates. Because the cNP uses an acceptation-rejection algorithm to sample from the conditional distribution, it may be able to correct better for the drift. One final note is that we could not examine each bootstrap distribution produced

in this simulation study, but in practice, we would recommend to plot the bootstrap distributions as their shape and support space can be an indication of the stability from one bootstrap run to another. Patently aberrant estimates can occur especially in sparse designs with high variance, and could be excluded before building the confidence intervals. Another necessary check is model adequacy which can be examined by diagnostics (Nguyen et al. (2017)) as well as run assessment (Lavielle (2014)) to check the sensitivity of the model estimates to different initial starting values and random seeds.

5 Conclusion

We proposed a new bootstrap approach for non-linear mixed effect models using samples from the conditional distribution instead of estimated residuals. We implemented it in version 3.2 of the saemix package for R, and compared its performances to other widely used bootstrap methods, for a model with continuous outcomes. The cNP bootstrap improved the coverage rate in the different scenarios tested in this study over the traditional non parametric bootstrap, and was as efficient as the more simple case bootstrap. Although it is more computationally intensive than Case bootstrap, it has the advantage of preserving the structure of the original dataset, including the covariate distribution and design, which make it conceptually simpler to implement in complex models where Case bootstrap may need stratification to maintain identifiability. Bootstrap approaches however are not a magic wand and cannot compensate for poor estimates, such as variance terms in small sample size or poorly identifiable parameters. In particular we recommend performing thorough diagnostics of model adequacy and stability before applying a bootstrap approach.

Acknowledgments

The authors would like to thank Hervé Le Nagard, Lionel de la Tribouille and Rémy Bertino for the use of the CATIBioMed computer facility.

Disclosure Statement

Author contributions: E.C. and S.K. designed the research. S.K. performed the research, analysed the data and wrote the first draft of the manuscript. M.U. and E.C. discussed and interpreted the results. E.C. finalised the manuscript with M.U. and S.K. providing input and discussion elements.

Funding: No funding was received for this work. E.C. has received consulting fees from Sanofi, bearing no relationship with the present manuscript.

References

Austin PC, Leckie G (2020) Bootstrapped inference for variance parameters, measures of heterogeneity and random effects in multilevel logistic regression models. J Stat Comput Simul 90:3175–99. URL https://doi.org/10.1080/00949655.2020.1797738

Broeker A, Wicha SG (2020) Assessing parameter uncertainty in small-n pharmacometric analyses: value of the log-likelihood profiling-based sampling importance resampling (llp-sir) technique. J Pharmacokinet Pharmacodyn 47:219–28

Carpenter JR, Goldstein H, Rasbash J (2003) A novel bootstrap procedure for assessing the relationship between class size and achievement. J R Stat Soc Ser C Appl Stat 52(4):431–443

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

Comets E, Rodrigues C, Jullien V, et al (2021) Conditional non-parametric bootstrap for non-linear mixed effect models. Pharm Res 38:1057–66

Das S, Krishen A (1999) Some bootstrap methods in non-linear mixed-effects models. J Stat Plan Inference 75:237–245

Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data. CRC, London

Davison AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge

Dosne A, Bergstrand M, Harling K, et al (2017) An automated sampling importance resampling procedure for estimating parameter uncertainty. J Pharmacokinet Pharmacodyn 44:509—-520

Dumont C, Lestini G, Nagard H, et al (2018) Pfim 4.0, an extended r program for design evaluation and optimization in nonlinear mixed-effect models. Comput Methods Programs Biomed 156

Efron B (1979) Bootstrap methods: another look at the jacknife. Ann Stat 7:1–26

Efron B (2000) The bootstrap and modern statistics. J Am Stat Soc 95:1293–6

Guhl M, Bertrand J, Fayette L, et al (2024) Uncertainty computation at finite distance in nonlinear mixed effects models - a new method based on metropolis-hastings algorithm. The AAPS Journal 26(3):53

Karlsson M, Savic R (2007) Diagnosing model diagnostics. Clin Pharmacol Ther 82:17–20

Kuhn E, Lavielle M (2005) Maximum likelihood estimation in nonlinear mixed effects models. Comput Stat Data Anal 49:1020–1038

Lalonde RL, Kowalski KG, Hutmacher MM, et al (2007) Model-based drug development. Clin Pharmacol Ther 82:21–32

Lavielle M (2014) Mixed effects models for the population approach: models, tasks, methods and tools. Chapman & Hall CRC Biostatistics Series, Boca Raton, FL

Lavielle M, Ribba B (2016) Enhanced method for diagnosing pharmacometric models: random sampling from conditional distributions. Pharm Res 33:2979–88

MacKinnon J (2006) Bootstrap methods in econometrics. Econ Rec 82:S2–S18

Morris J (2002) The BLUPs are not "best" when it comes to bootstrapping. Stat Probab Lett 56:425–30

Morris TP, White IR, Crowther MJ (2019) Using simulation studies to evaluate statistical methods. Stat Med 38(11):2074–2102

Nguyen THT, Mouksassi M, Holford N, et al (2017) Model evaluation of continuous data pharmacometric models: Metrics and graphics. CPT Pharmacometrics Syst Pharmacol 6(2):87–109. https://doi.org/10.1002/psp4.12161

Nyberg J, Bazzoli C, Ogungbenro K, et al (2015) Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics studies. Br J Clin Pharmacol 79(1):6–17. https://doi.org/10.1111/bcp.12352

Pillai GC, Mentré F, Steimer JL (2005) Non-linear mixed effects modeling – from methodology and software development to driving implementation in drug development science. J Pharmacokinet Pharmacodyn 32:161—-83

Plan E, Maloney A, Mentré F, et al (2012) Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for dose-response models. AAPS J 14:420–432

R Development Core Team (2006) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org, ISBN 3-900051-07-0

Retout S, Mentré F, Bruno R (2002) Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. Stat Med 21(18):2623–2639

Thai H, Mentré F, Holford NH, et al (2013) A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. Pharm Stat 12:129–40

Thai H, Mentré F, Holford NH, et al (2014) Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics. J Pharmacokinet Pharmacodyn 41:15–33

Vonesh EF, Carter RL (1992) Mixed-effects nonlinear regression for unbalanced repeated measures. Biometrics 48:1–17

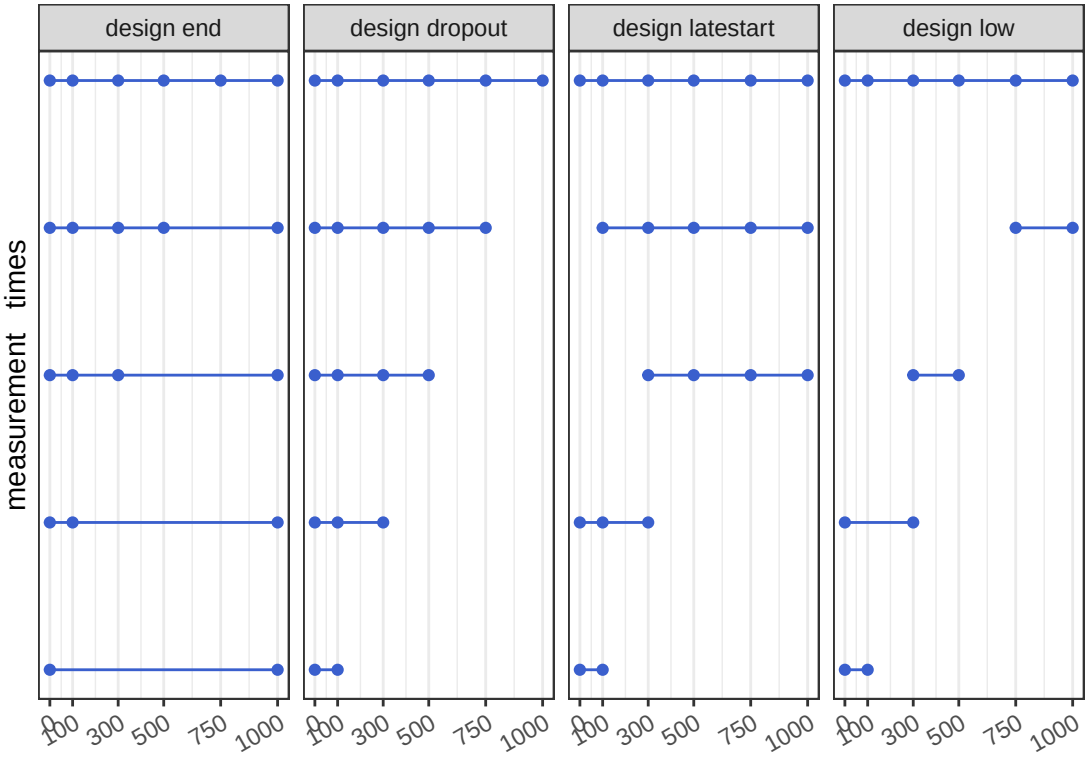


Fig. 1 : Measurement times in the four unbalanced designs. The first 3 designs involved 100 subjects divided in 5 groups of 20 with 2 to 6 doses, and the last group included 80 subjects with 2 doses and 20 subjects with 6 doses.

Tables

Table 1 Parameters used in simulations

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

Figures

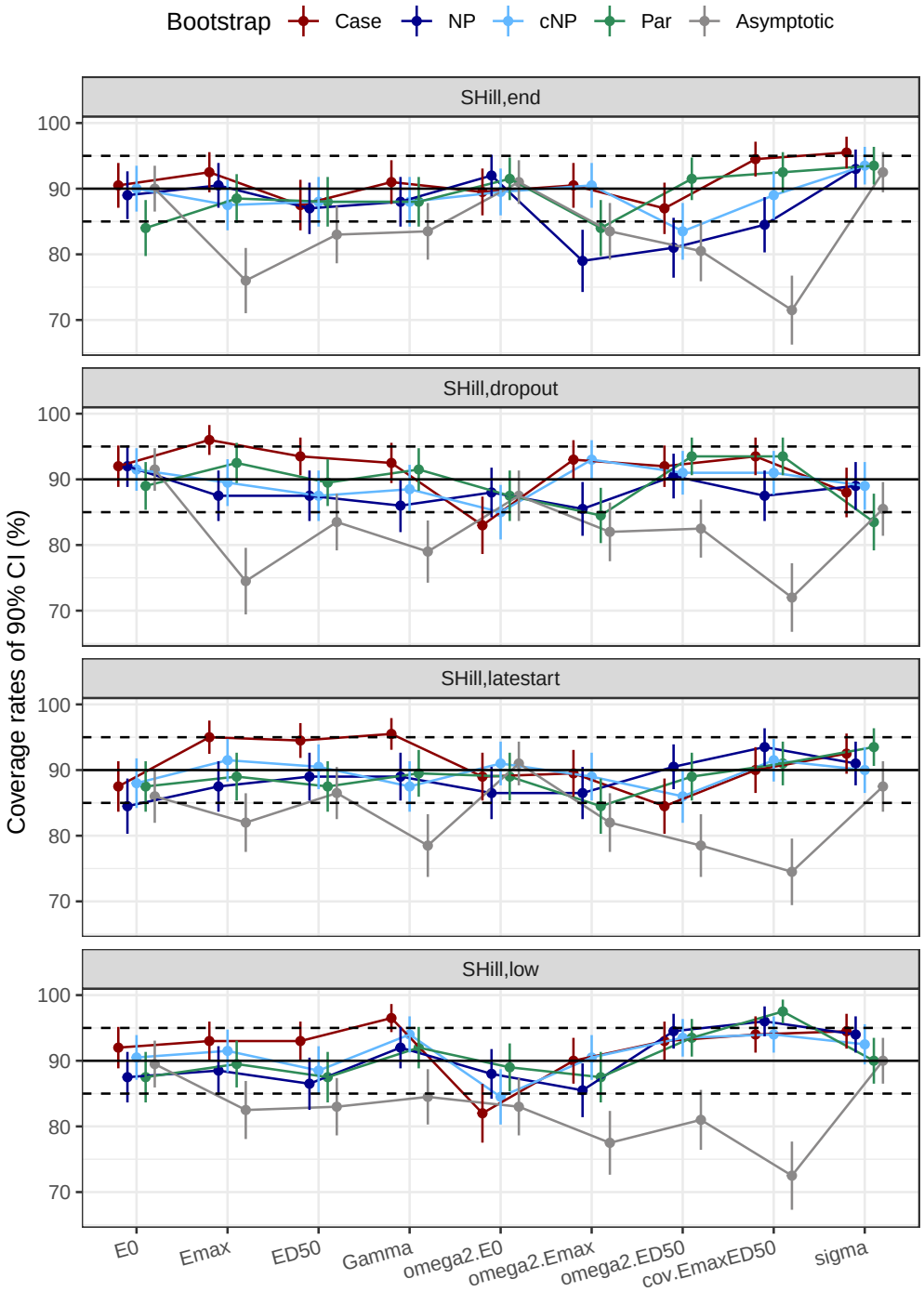


Fig. 3 Coverage rates of the 90% CI and errorbars of their MC uncertainty for the unbalanced designs for the S_{Hill}. Dashed lines indicate a coverage of 85 and 95%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches.

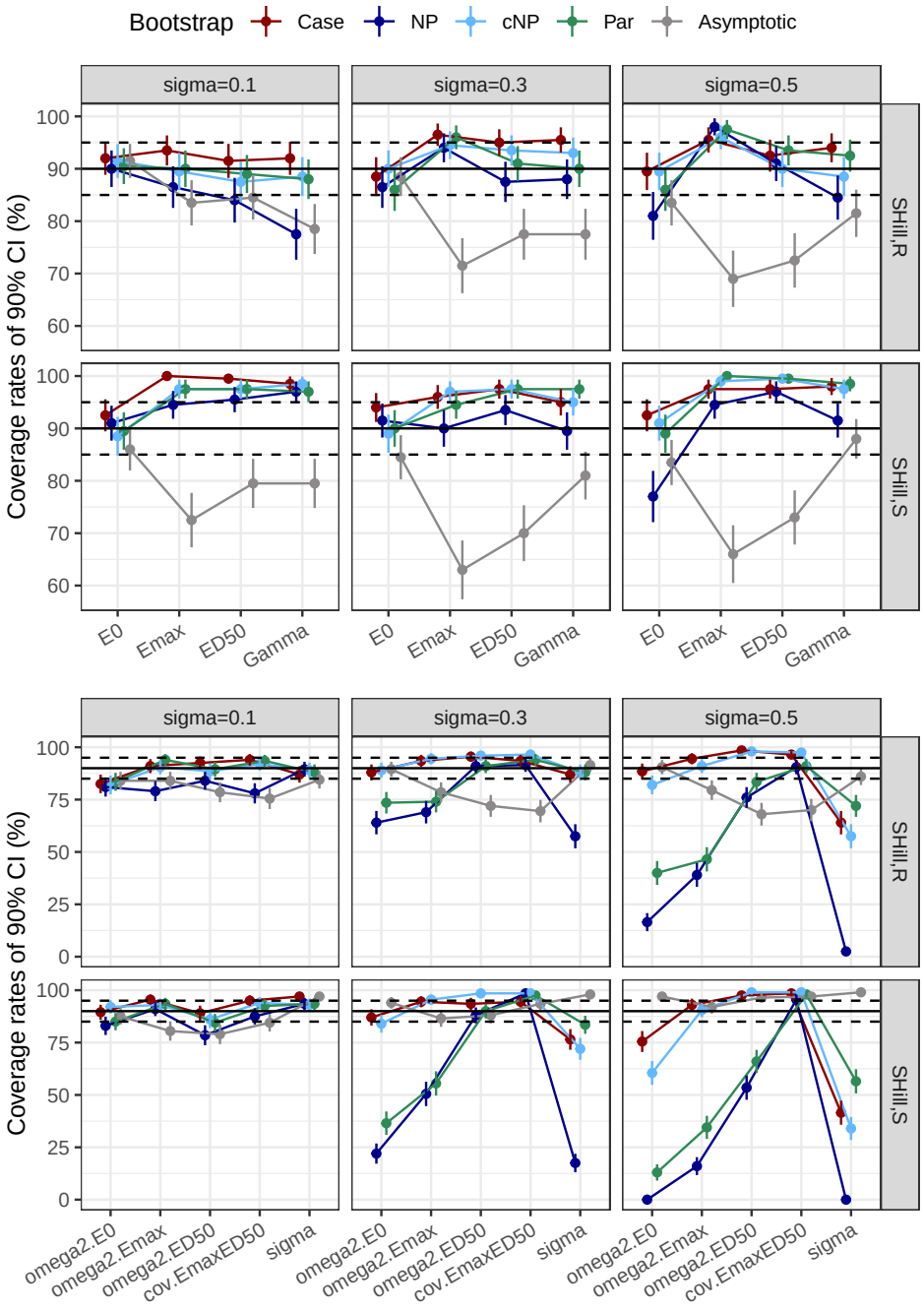


Fig. 4 Coverage rates of the 90% CI and errorbars of their MC uncertainty for the original design and the two designs with increased error coefficient $\sigma = 0.3$ and $\sigma = 0.5$ under the S_{Hill}. Dashed lines indicate a coverage of 85 and 95%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches. At the first pair of lines we show the fixed effects and at the second the random effects. Each pair of lines has the same scale on the Y-axis but we separated the fixed effects (top two lines) and the variance parameters (bottom two lines) because of the different scales.

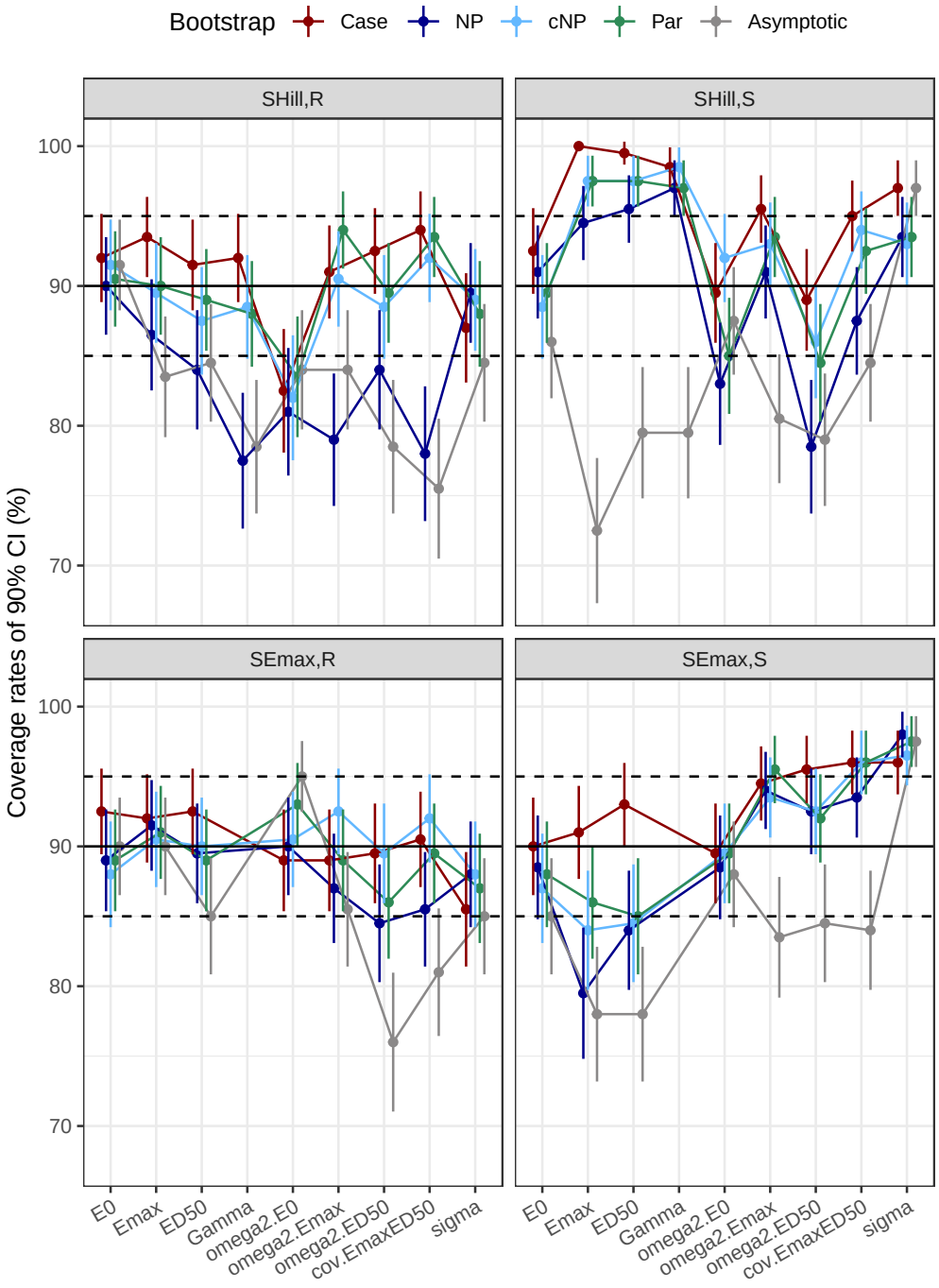


Fig. 2 Coverage rates of the 90% CI along with the errorbars of their MC uncertainty for the original designs. Dashed lines indicate a coverage of 85 and 95%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches.

Supplementary information. Appendix A details the bootstrap algorithms used in this study. Appendices B to E include additional results for the different scenarios.

Declarations

Funding

None

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Statements of ethical approval

Not applicable

Data availability

The simulated datasets used in this study are available on Zenodo (repositories: 10.5281/zenodo.4059718 and 10.5281/zenodo.10779192).

Code availability

The saemix package is available on CRAN (current version: 3.5), and successive versions are archived on the IAME research center github (https://github.com/iame-researchCenter/saemix), where the user guide for saemix can be downloaded (https://github.com/iame-researchCenter/saemix/blob/main/docsaem.pdf). The development version of the package is available on the development github repository (https://github.com/saemixdevelopment/saemixextension).

Author Contribution information

Sofia Kaisaridi: Writing – review & editing, Software, Methodology. Moreno Ursino: Writing - review. Emmanuelle Comets: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis.

Supplementary A - Bootstrap algorithms

Methods

Bootstrapping generates multiple pseudo-samples mimicking the distribution of the original sample. Denoting B the number of bootstrap samples, a general bootstrap algorithm is:

1. Generate a bootstrap sample by resampling from the data and/or from the estimated model
2. Compute the estimates of the parameters of the model for the bootstrap sample
3. Repeat steps 1-2 B times to obtain the bootstrap distribution of the parameter estimates.

Case bootstrap (Case):

This method consists of resampling with replacement the entire subjects (ξ_i,$\mathbf{y}_i$ where $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{in_i})'$) from the original data before modelling. It is also called the *paired bootstrap*. It is the most straightforward way to do bootstrapping and is mostly assumption-free. Stratification may be applied when the data is non-homogeneous to preserve the distribution of a covariate or of a sampling schedule.

Parametric residual bootstrap (Par):

The parametric bootstrap requires the strongest assumptions as it depends both on the model and the distributions of parameters and errors. This method resamples the residuals by simulating from the theoretical distributions obtained after fitting the model, here for the random effects from the normal distributions adjusted the estimated mean and variance of the initial fit. In non-linear mixed effect models, two levels of variability must be considered, within-subject and between different subjects. The bootstrap sample is obtained as follows:

1. fit the model to the data to obtain $\hat{\theta}$

2. draw a sample $\{\eta_i^*\}_{i=1,\ldots,N}$ from a normal distribution with mean zero and covariance matrix $\hat{\Omega}$
3. draw a sample $\{\epsilon_{ij}^*\}_{j=1,\ldots n_i}$ from a normal distribution with mean zero and covariance matrix $\mathcal{I}_{n_i}$ where $\mathcal{I}$ denotes the identity matrix
4. generate the bootstrap responses $y_{ij}^* = f(x_{ij}, \hat{\mu}, \eta_i^*) + g(x_{ij}, \hat{\mu}, \eta_i^*, \sigma)\ \epsilon_{ij}^*$ (where we explicitly replace ψ_i by $\hat{\mu}$ and η_i^* to highlight the dependency on the resampled random effects)

Non-parametric residual bootstrap (NP):

This method resamples with replacement from the residuals obtained after model fitting for all subjects. The bootstrap sample is obtained as follows:

1. fit the model to the data to obtain $\hat{\theta}$
2. estimate the individual parameters $\hat{\eta}_i$ as the EBE, then center and normalise them as detailed below
3. draw a sample η_i^* with replacement for $i = 1, \ldots N$ from the resulting set
4. calculate the residuals $\hat{\epsilon}_{ij} = y_{ij} - f(x_{ij}, \hat{\mu}, \hat{\eta}_i)$, then center and normalise them
5. draw a sample $\{\epsilon_{ij}^*\}_{j=1,\ldots n_i}$ with replacement globally from the resulting set
6. generate the bootstrap responses $y_{ij}^* = f(x_{ij}, \hat{\mu}, \eta_i^*) + g(x_{ij}, \hat{\mu}, \eta_i^*, \sigma)\ \epsilon_{ij}^*$

The normalisation in step (b) and (d) is necessary because the EBE generally suffer from regression to the mean and their variance may be considerably smaller than the true variability in the population (Karlsson and Savic (2007)). To correct for this, Carpenter et al. (2003) suggested centering the residuals, to ensure their distribution has mean 0, and inflating the variance by using the ratio between the estimated and empirical variance-covariance matrices. The transformation of random effects was carried out using the eigenvalue decomposition (EVD) proposed by Thai et al. (2014) to limit numerical difficulties due to almost singular matrices. This procedure reads as follows:

1. center the raw estimated random effects: $\tilde{\eta_i} = \hat{\eta_i} - \bar{\eta_i}$
2. obtain the EVD of the estimated variance-covariance matrix: $\hat{\Omega} = V_\Omega \; D_\Omega V_\Omega^T$ where D_Ω is the diagonal matrix containing the eigenvalues of $\hat{\Omega}$
3. obtain the EVD of S, the variance-covariance matrix of the centered random effects: $S = V_S \; D_S V_S^T$ where D_S is the diagonal matrix containing the eigenvalues of S
4. calculate the correction matrix A_η using these two decompositions as $A_\eta = V_S \; D_S^{-1/2} \; V_\Omega \; D_\Omega^{-1/2}$
5. transform the centered random effects using the ratio A_η: $\hat{\eta}_i^{'} = \tilde{\eta_i} \times A_\eta$

Similarly, the residuals are transformed using the empirical variance:

1. center the raw estimated residuals: $\tilde{\epsilon}_{ij} = \hat{\epsilon}_{ij} - \bar{\epsilon}_{ij}$
2. calculate the correction factor $A_\sigma = 1/\sigma_{emp}$ where σ_{emp} is the empirical standard deviation of the centered residuals
3. transform the centered residuals using the ratio A_σ: $\hat{\epsilon}_{ij}^{'} = \tilde{\epsilon}_{ij} \times A_\sigma$

Implementation and runtimes

The bootstrap methods were implemented for the saemix package in version 3.2 (Comets et al. (2017)) and is available on CRAN. Development versions are downloadable on github https://github.com/saemixdevelopment/saemixextension.

In terms of runtimes, the cNP was slower than the other bootstraps by 30 to 100% in the various scenarios, with the limiting step being the computation of the conditional distribution. However, although this made our simulations comparatively slower for this bootstrap than for the other residual-based bootstraps, in practice for a real data analysis the conditional distribution is a recommended step in the estimation as it makes the diagnostic graphs more informative, therefore the additional computational cost incurred by using cNP is mostly taken care of during the estimation itself and can be leveraged for better model diagnostics and a more precise estimation of the log-likelihood using importance sampling. The computation time increases with the

number of bootstraps. Here we used 200 bootstrap samples for each approach to keep the simulations manageable and because a calibration in Comets et al. (2021) showed the coverages to be stable, but the number of samples should be increased when considering more extreme quantiles.

Supplementary B - Simulation scenarios

In the first scenario, a rich and a sparse design were simulated for each model (E_{max} with $\gamma = 1$ and Hill with $\gamma = 3$). The rich design included N=100 subjects given 4 doses each (0, 100, 300, 1000), and the sparse design 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).

In the second scenario, we considered unbalanced designs with non-homogenous amount of information across subjects. We evaluated using the Fisher information matrix (FIM) several combinations from the dosing set (0, 100, 300, 500, 750, 1000), with 100 subjects divided in 5 groups of 20 with 2 to 6 doses to select the 3 designs shown in the first 3 facets of Figure 1, representing situations where some subjects receive a full set of doses while others receive only low, only high, or a mix of doses. We simulated a last scenario (last facet of the figure) where 80 subjects received only 2 doses while 20 subjects received the full set of doses, to mimic a mix of sparse and rich designs. The expected RSE according to PFIM was reasonable (see online Appendix 5).

In the third scenario, we performed two additional simulations following the original designs (rich with N=100, n_i=4, and sparse with N=100, n_i=2), where we increased the error coefficient σ to 0.3 and 0.5, leaving the other parameters unchanged. In Table 2 we present the details for each simulation scenario in this study.

Table 2 Simulation scenarios

Simulation	Design	Subjects	Doses
Original	Rich	100	4
	Sparse	200 (groups of 50)	2
Unbalanced		100 (groups of 20)	2,3,4,5,6
Increased error	Rich	100	4
(σ=0.3 or 0.5)	Sparse	200 (groups of 50)	2

Supplementary C - Original designs

Coverage rates

In figure 5 we plot the coverage rates for the 95% confidence intervals and their MC uncertainty for the original simulation scenarios.

Bias

Figure 6 shows the bias on the parameters and their S.E. obtained by all the bootstraps as well as the asymptotic method for the original designs. We calculate the bootstrap relative bias (RB) for the parameters with respect to the true values θ_0:

$$RBbias(\hat{\theta})_{true} = \frac{1}{K} \cdot \sum_{k=1}^{K} \frac{\hat{\theta}_{B,k} - \theta_0}{\theta_0} \times 100 \tag{7}$$

We compared the bootstrap SE with the empirical SE obtained by K simulated datasets giving the "true" value observed across the simulations:

$$\mathrm{SE}_{\mathrm{empirical}}(\hat{\theta}) = \sqrt{\frac{1}{K-1} \cdot \sum_{k=1}^{K} (\hat{\theta}_k - \theta_0)^2} \tag{8}$$

where θ_k is the value of the parameter estimated in the k^{th} dataset and θ_0 the true value. The relative bias on SE by each bootstrap method is then computed as the percentage difference between the average bootstrap SE and the empirical SE:

$$RBbias(\mathrm{SE}(\hat{\theta})) = \frac{1}{K} \cdot \sum_{k=1}^{K} \frac{\widehat{\mathrm{SE}}_{B,k} - \mathrm{SE}_{\mathrm{empirical}}(\hat{\theta})}{\mathrm{SE}_{\mathrm{empirical}}(\hat{\theta})} \times 100 \tag{9}$$

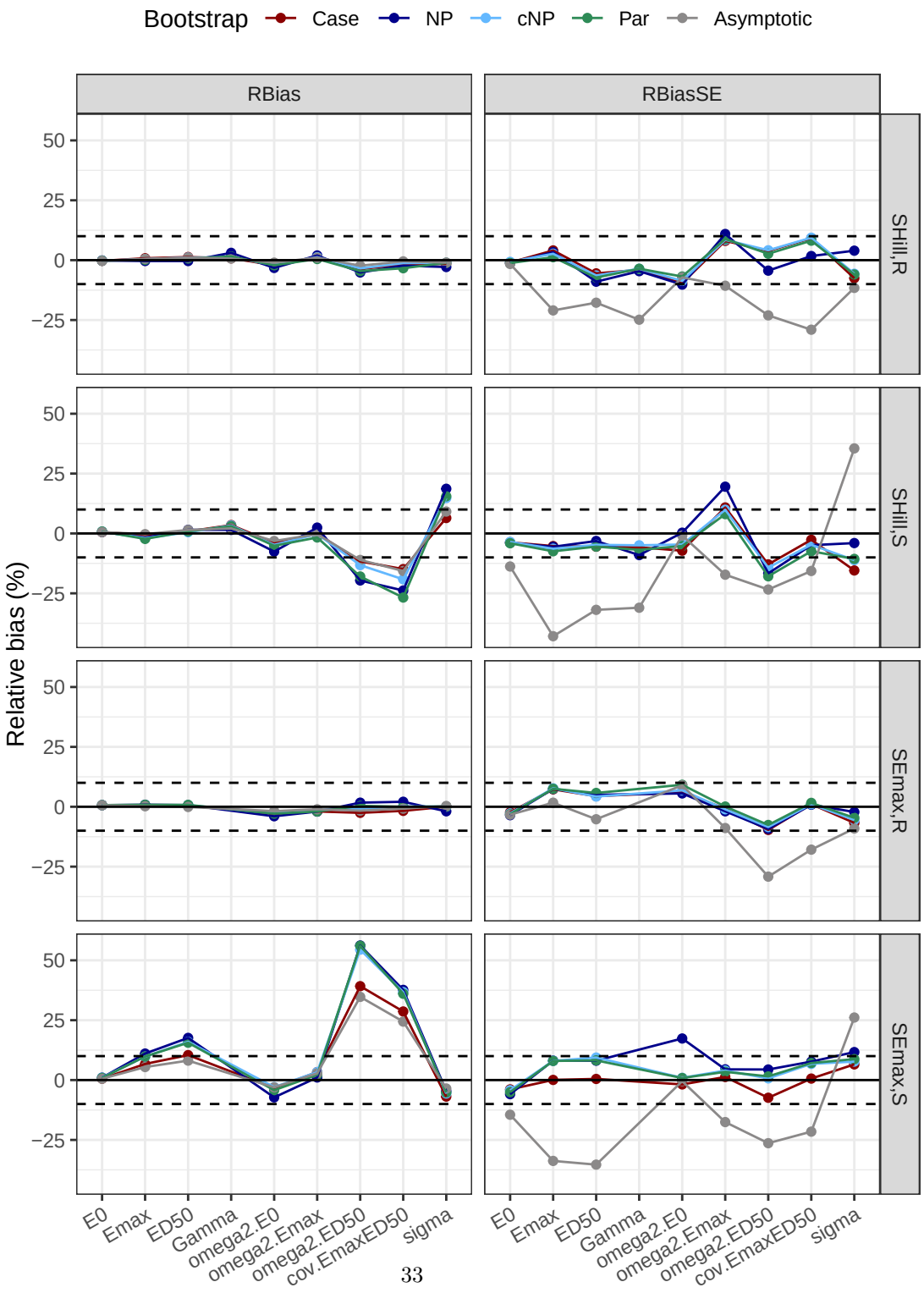


Fig. 6 Relative bias on parameters (left) and SE (right) for the original designs. Dashed lines delineate absolute relative biases within 10%. The same scale on the Y-axis is used

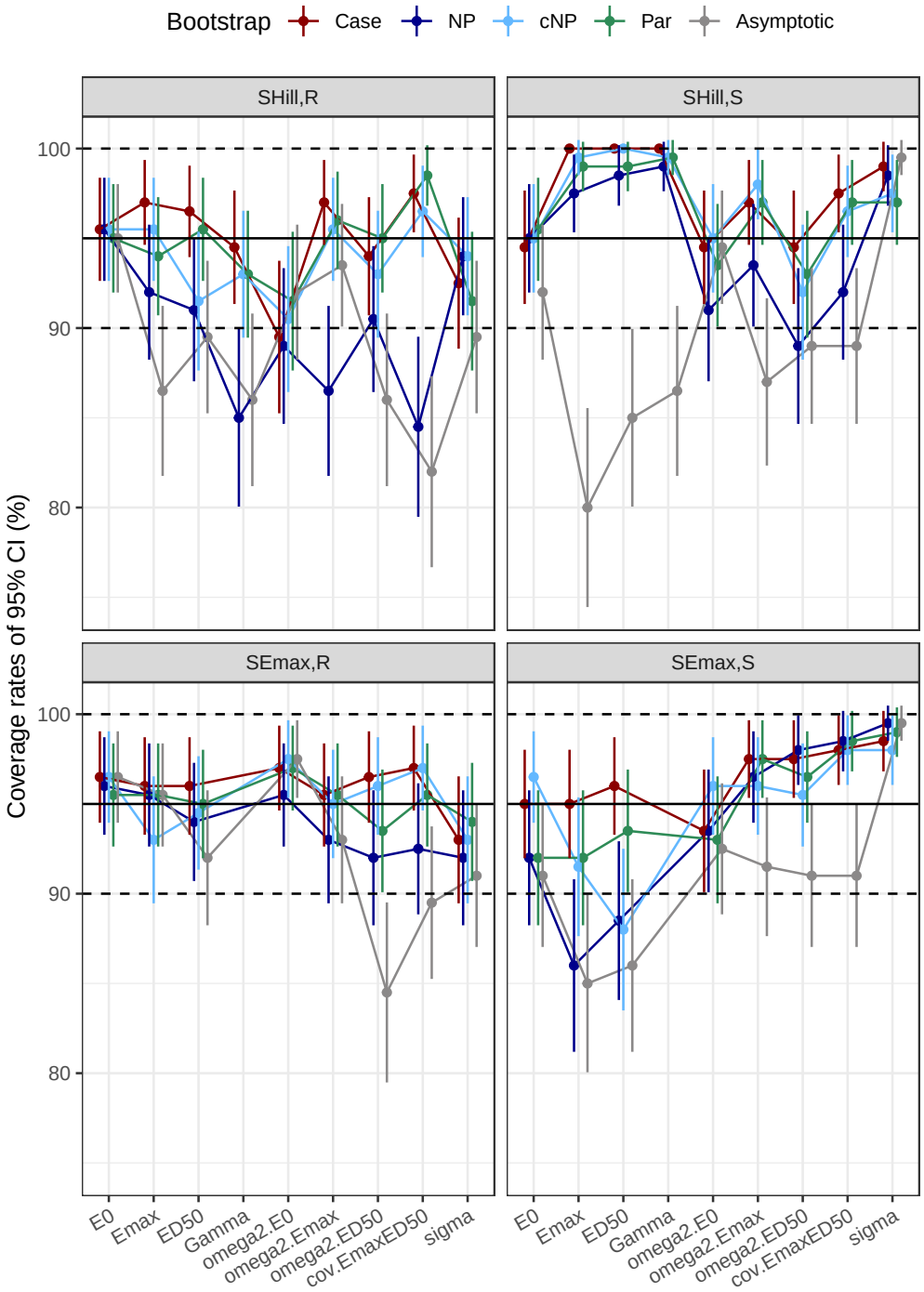


Fig. 5 Coverage rates of 95% CI and errorbars of their MC uncertainty for the original designs. Dashed lines indicate a coverage of 90 and 100%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches.

Supplementary D - Unbalanced designs

Design evaluation

In Figure 7 we present the expected RSE % calculated by the R package PFIM 4.0 for each parameter. The results for each unbalanced design is presented in a different colour and compared to the ones from the original rich designs (with red colour).

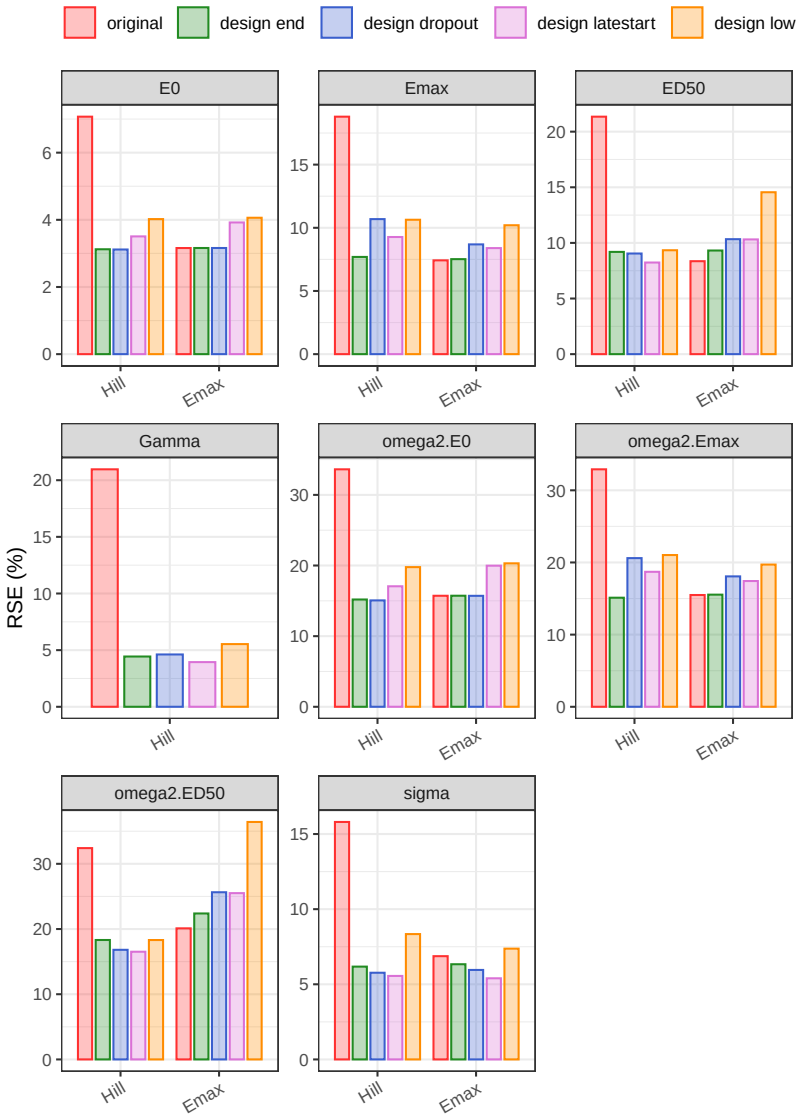


Fig. 7 Barplots of predicted RSE by PFIM 4.0, across the unbalanced designs compared to the original rich designs for each parameter

5.1 Coverage rates

Figure 8 shows the 90% coverage rates obtained by all the bootstraps as well as the asymptotic method for the E_{max} scenarios in the four unbalanced designs, with similar findings as in Figure 3. In figures 9 and 10 we plot the coverage rates for the 95% confidence intervals in the Hill and the E_{max} scenarios respectively. The results with the E_{max} were qualitatively similar to those obtained with a Hill model, except for the $S_{E_{max},dropout}$ scenario where the cNP had wider CI than expected for E_{max} and ED_{50} while NP and Par had narrower CI than expected for $\omega_{ED_{50}}$.

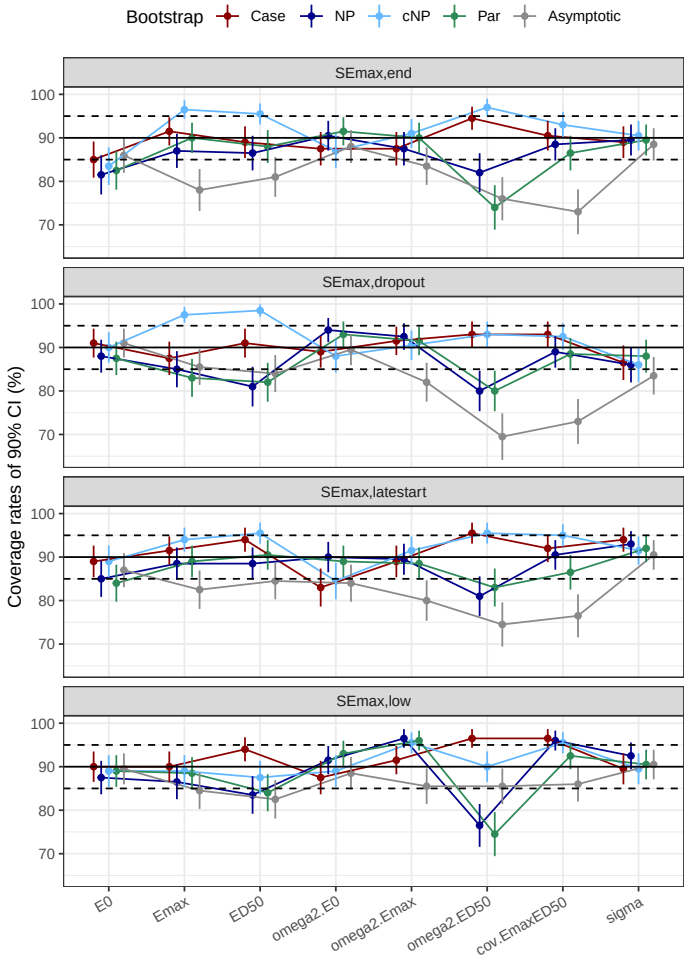


Fig. 8 Coverage rates of 90% CI and errorbars of their MC uncertainty for the unbalanced designs for the $S_{E_{max}}$. Dashed lines indicate a coverage of 85 and 95%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches.

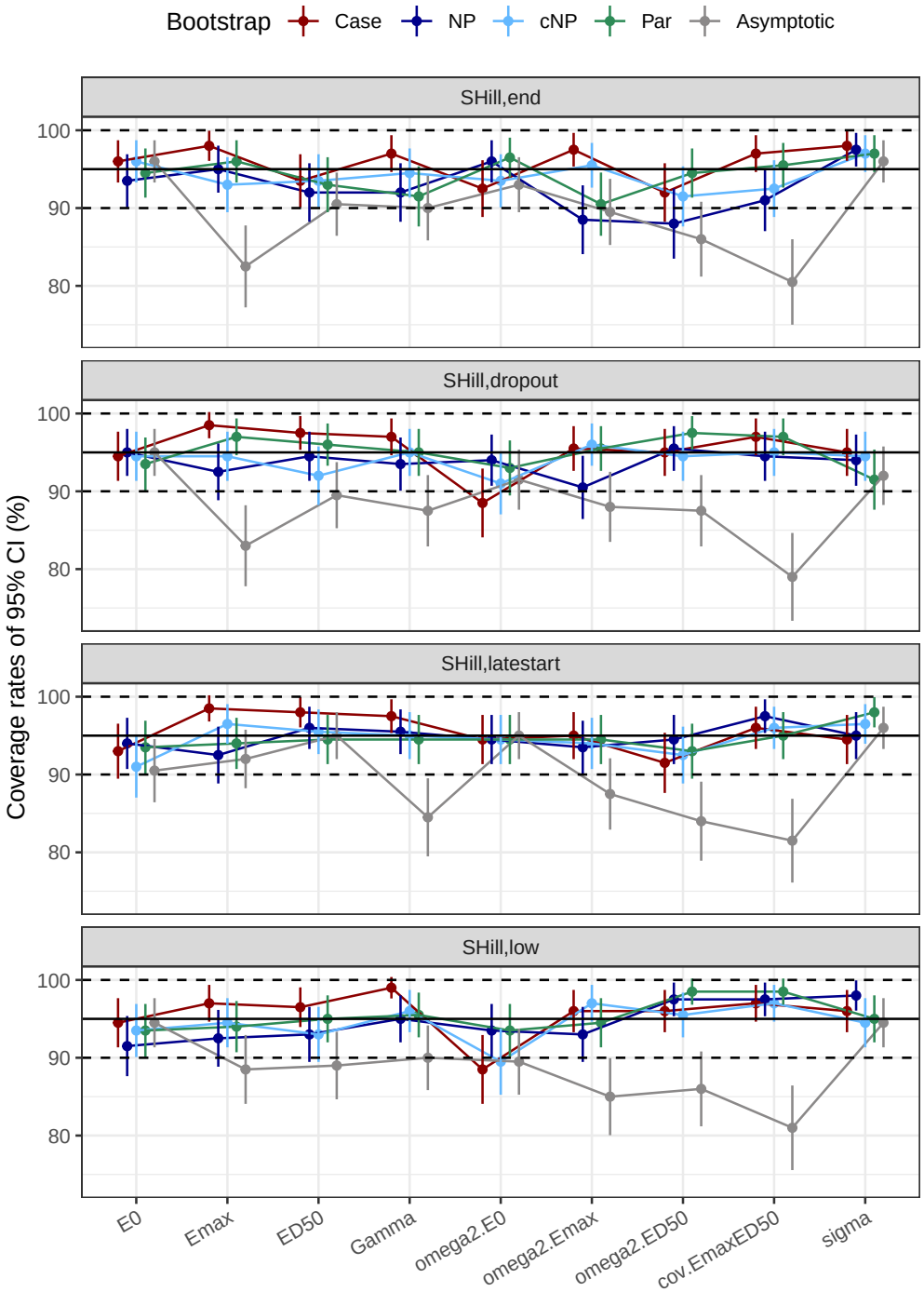


Fig. 9 Coverage rates of 95% CI and errorbars of their MC uncertainty for the unbalanced designs for the S_{Hill}. Dashed lines indicate a coverage of 90 and 100%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches.

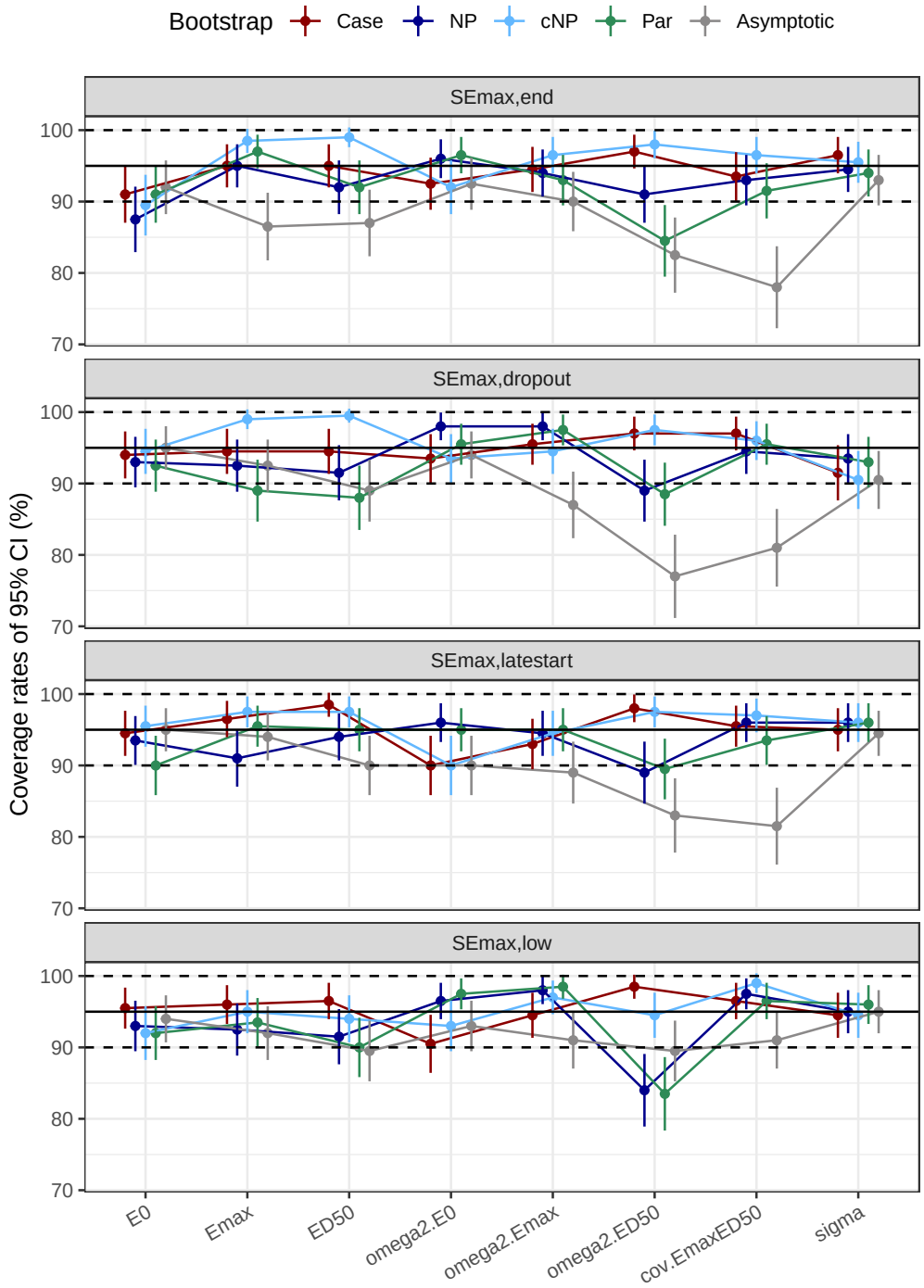


Fig. 10 Coverage rates of 95% CI and errorbars of their MC uncertainty for the unbalanced designs for the $S_{\mathrm{E}_{max}}$. Dashed lines indicate a coverage of 90 and 100%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches.

Bias

Figure 11 shows the bias on the parameters and their S.E. obtained by all the bootstraps as well as the asymptotic method for the Hill scenarios in the four unbalanced designs. In Figure 12 we plot the respective results for the E_{max} scenarios.

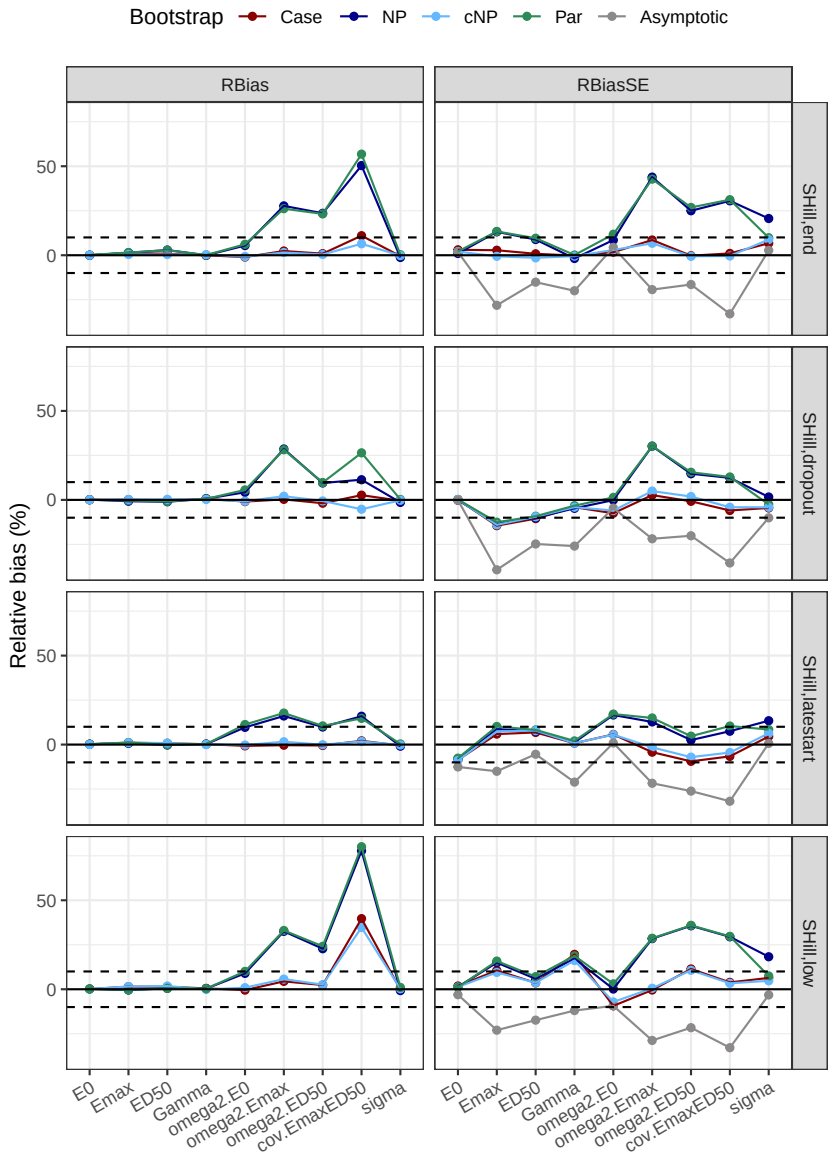


Fig. 11 Relative bias on parameters (left) and SE (right) for the original dosing scheme and the four unbalanced designs under the S_{Hill}. Dashed lines delineate absolute relative biases within 10%. The same scale on the Y-axis is used

Supplementary E - Increased error

Empirical SE

In Figure 13 we present the empirical SE for the scenarios with increased coefficient compared to the original designs with $\sigma = 0.1$. Each graph window corresponds to a different parameter and the empirical SE for each scenario is presented with a bar of a different colour across the four designs tested.

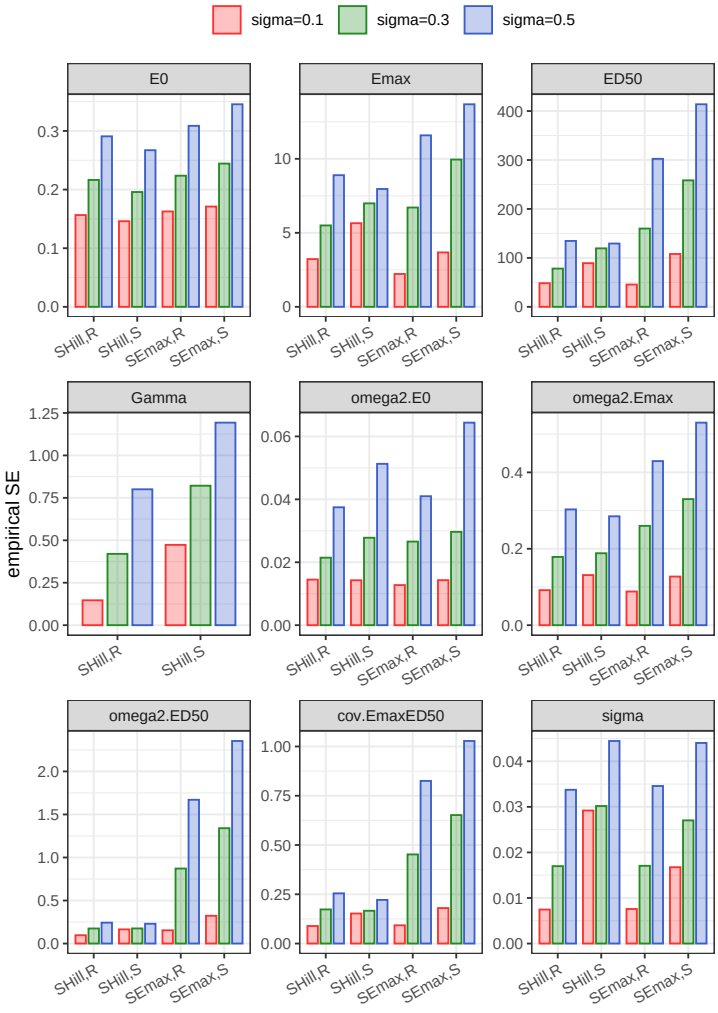


Fig. 13 Barplots of emprical SE for the scenarios with increased coefficient compared across the four designs tested

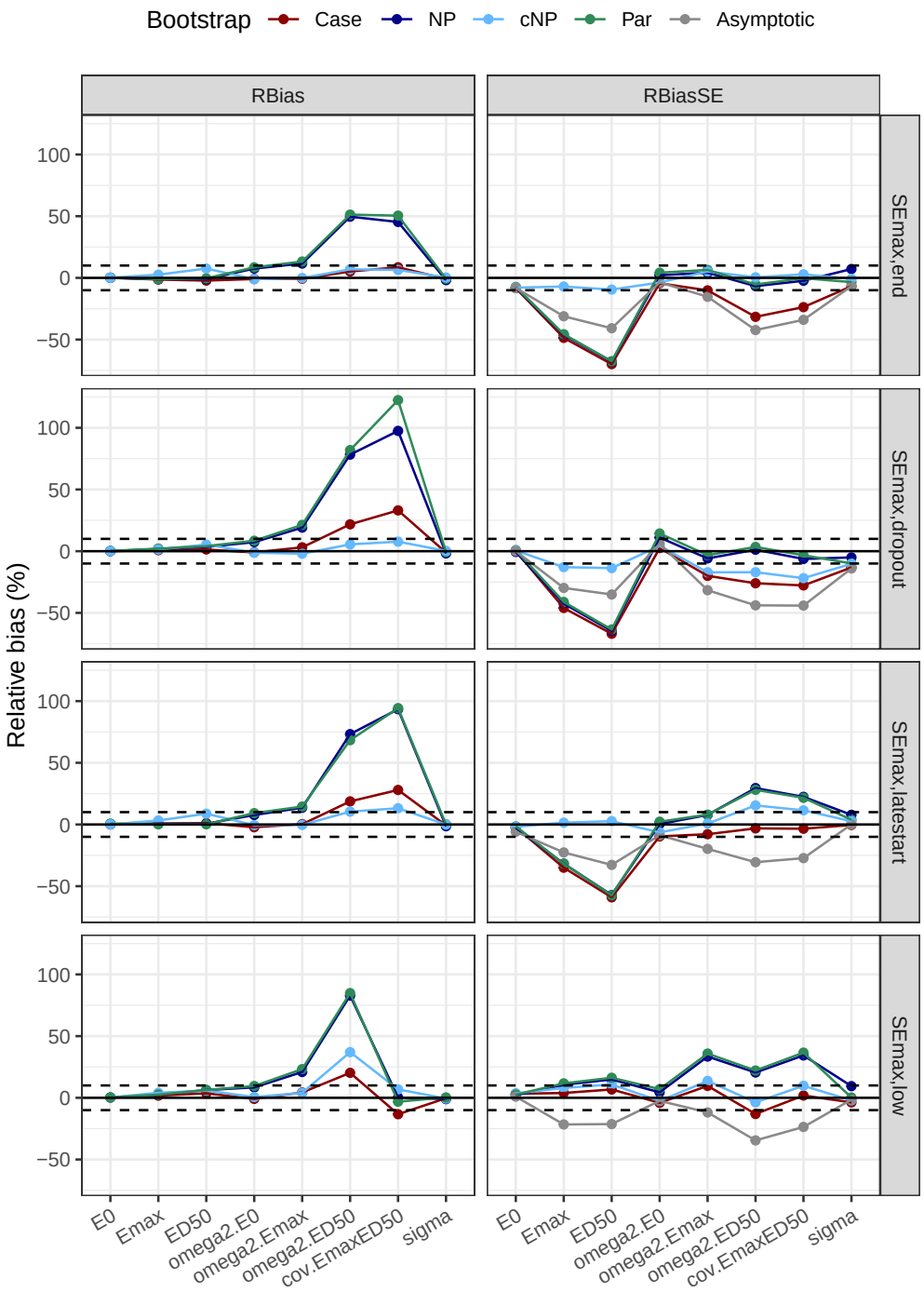


Fig. 12 Relative bias on parameters (left) and SE (right) for the original dosing scheme and the four unbalanced designs under the S_{Emax}. Dashed lines delineate absolute relative biases within 10%. The same scale on the Y-axis is used

Coverage rates

Figure 14 shows the coverage rates for the 90% CI obtained by all the bootstraps as well as the asymptotic method for the fixed effects (first pair of lines) and the random effects (second pair of lines) for the E_{max} scenarios in the original design and in the designs with increased error coefficient σ at 0.3 and 0.5. In Figures 15 and 16 we present the 95% coverage rates. All bootstraps show some bias as σ increases, which could suggest that with high residual variability random and residual variability may become difficult to separate. However, there was no correlation between the estimates of variance terms and the estimates of σ (data not shown) in our simulations to indicate this.

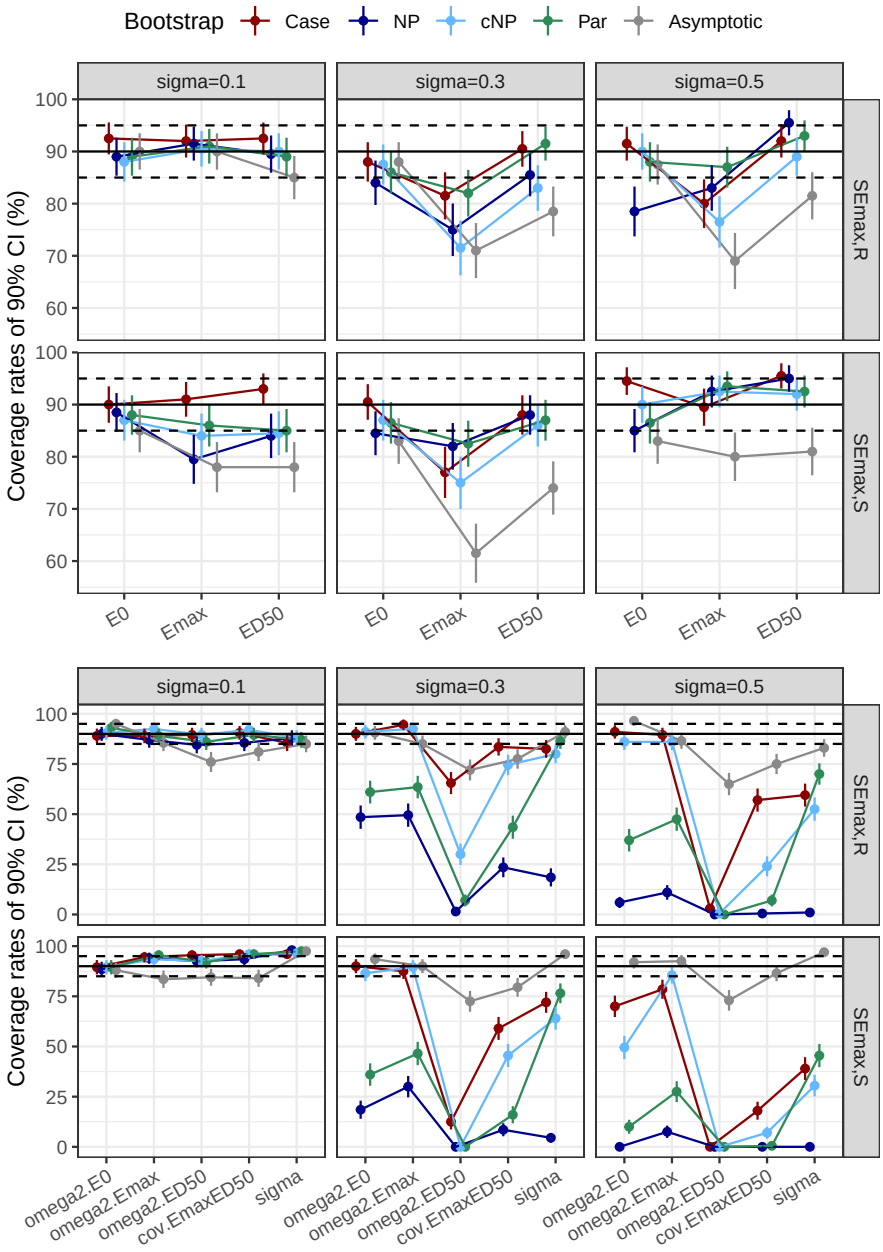


Fig. 14 Coverage rates of 90% CI and errorbars of their MC uncertainty for the original design and the two designs with increased error coefficient $\sigma = 0.3$ and $\sigma = 0.5$ under the $S_{\mathrm{E}_{max}}$. Dashed lines indicate a coverage of 85 and 95%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches. At the first pair of lines we show the fixed effects and at the second the random effects. Each pair of lines has the same scale on the Y-axis

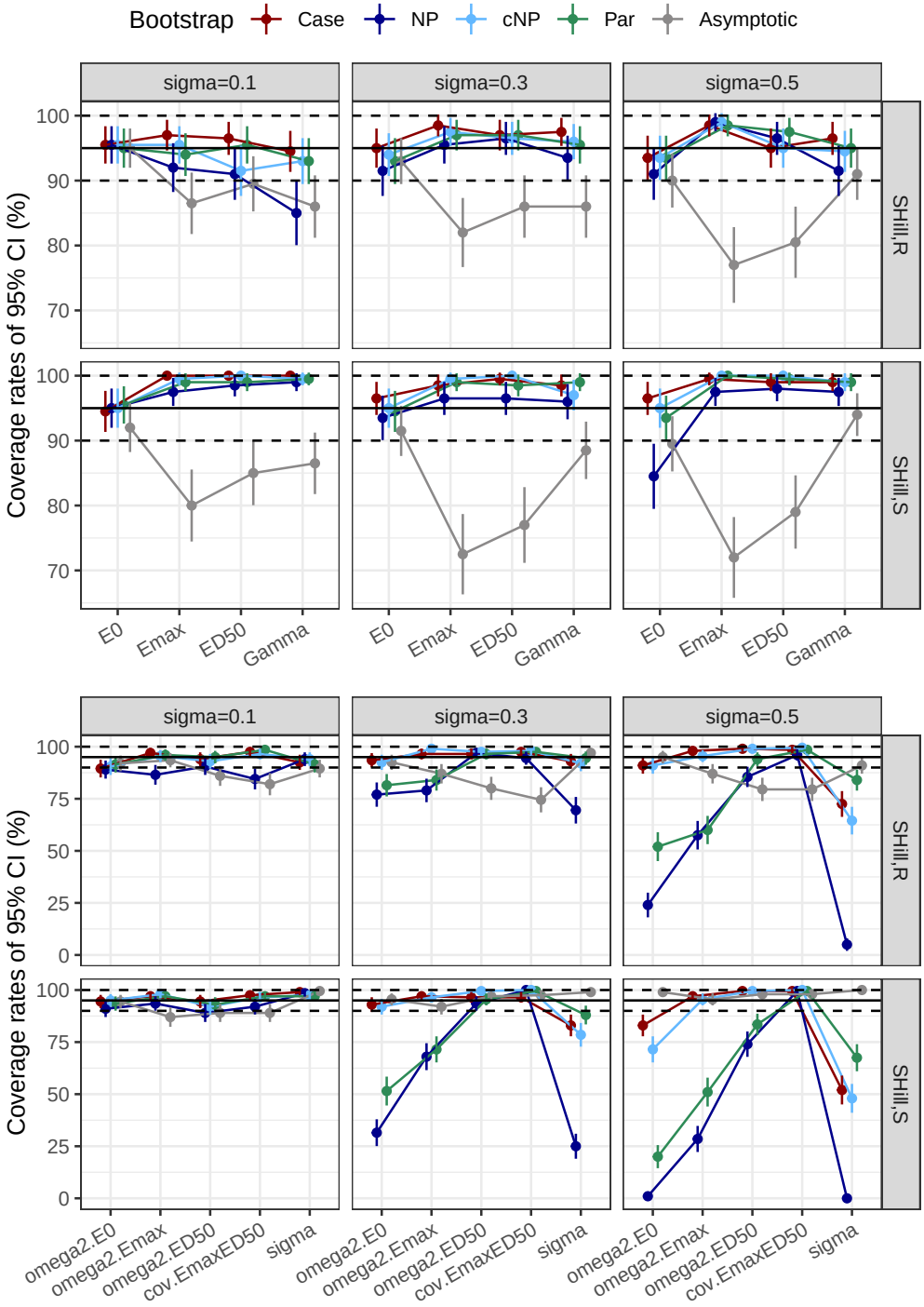


Fig. 15 Coverage rates of 95% CI and errorbars of their MC uncertainty for the original design and the two designs with increased error coefficient $\sigma = 0.3$ and $\sigma = 0.5$ under the S_{Hill}. Dashed lines indicate a coverage of 90 and 100%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches. At the first pair of lines we show the fixed effects and at the second the random effects. Each pair of lines has the same scale on the Y-axis

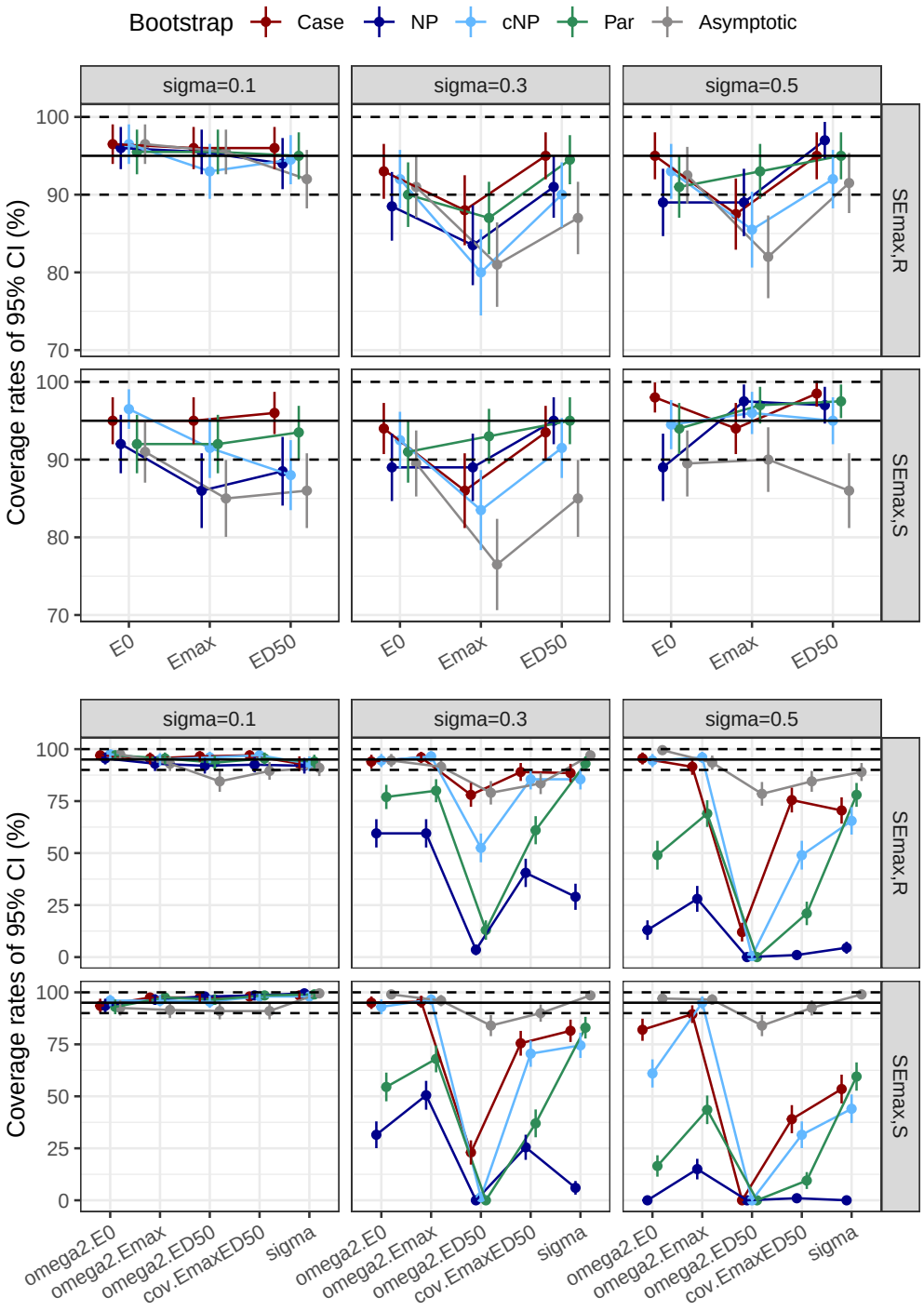


Fig. 16 Coverage rates of 95% CI and errorbars of their MC uncertainty for the unbalanced designs. Dashed lines indicate a coverage of 90 and 100%. The same scale is used on the Y-axis for the 4 graphs. The X-axis is jittered to avoid superimposing the different approaches.

Bias

Figure 17 shows the bias on the parameters and their S.E. obtained by all the bootstraps as well as the asymptotic method for the Hill scenarios in the original design and in the designs with increased error coefficient σ at 0.3 and 0.5. In Figure 18 we plot the respective results for the E_{max} scenarios.

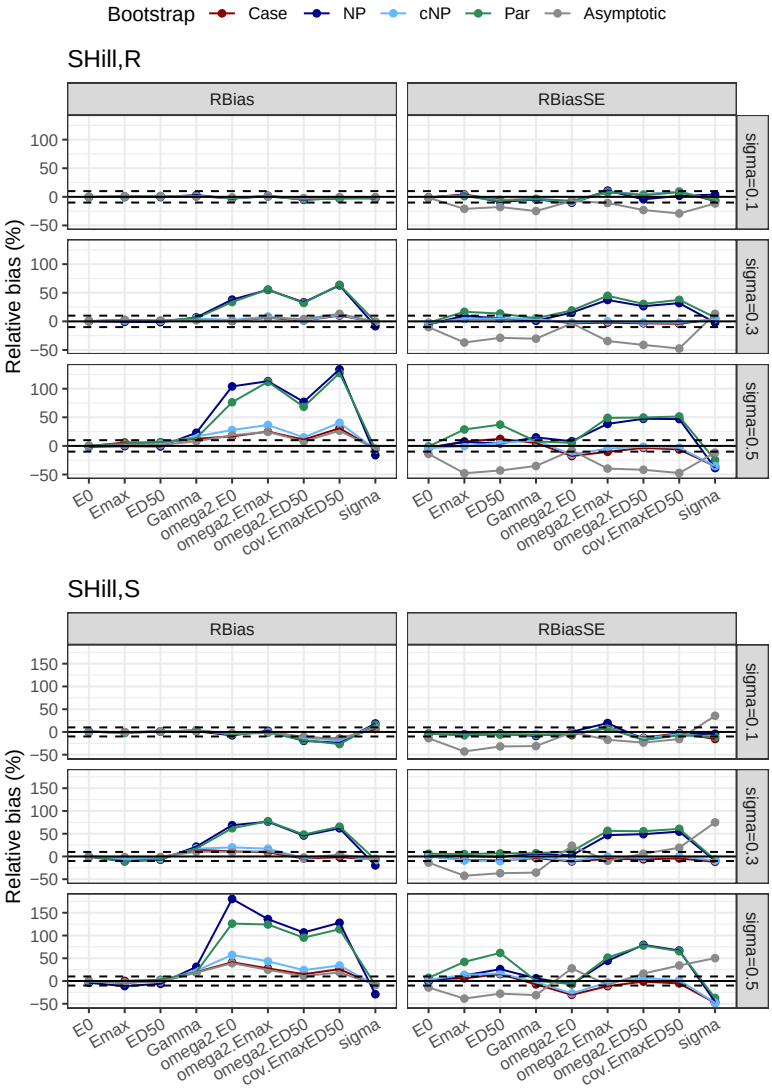


Fig. 17 Relative bias on parameters (up) and SE (down) for the original dosing scheme and the two scenarios with increased error coefficient $\sigma = 0.3$ and $\sigma = 0.5$ under the S_{Hill}. Dashed lines delineate absolute relative biases within 10%. A fixed scale is used on the Y-axis

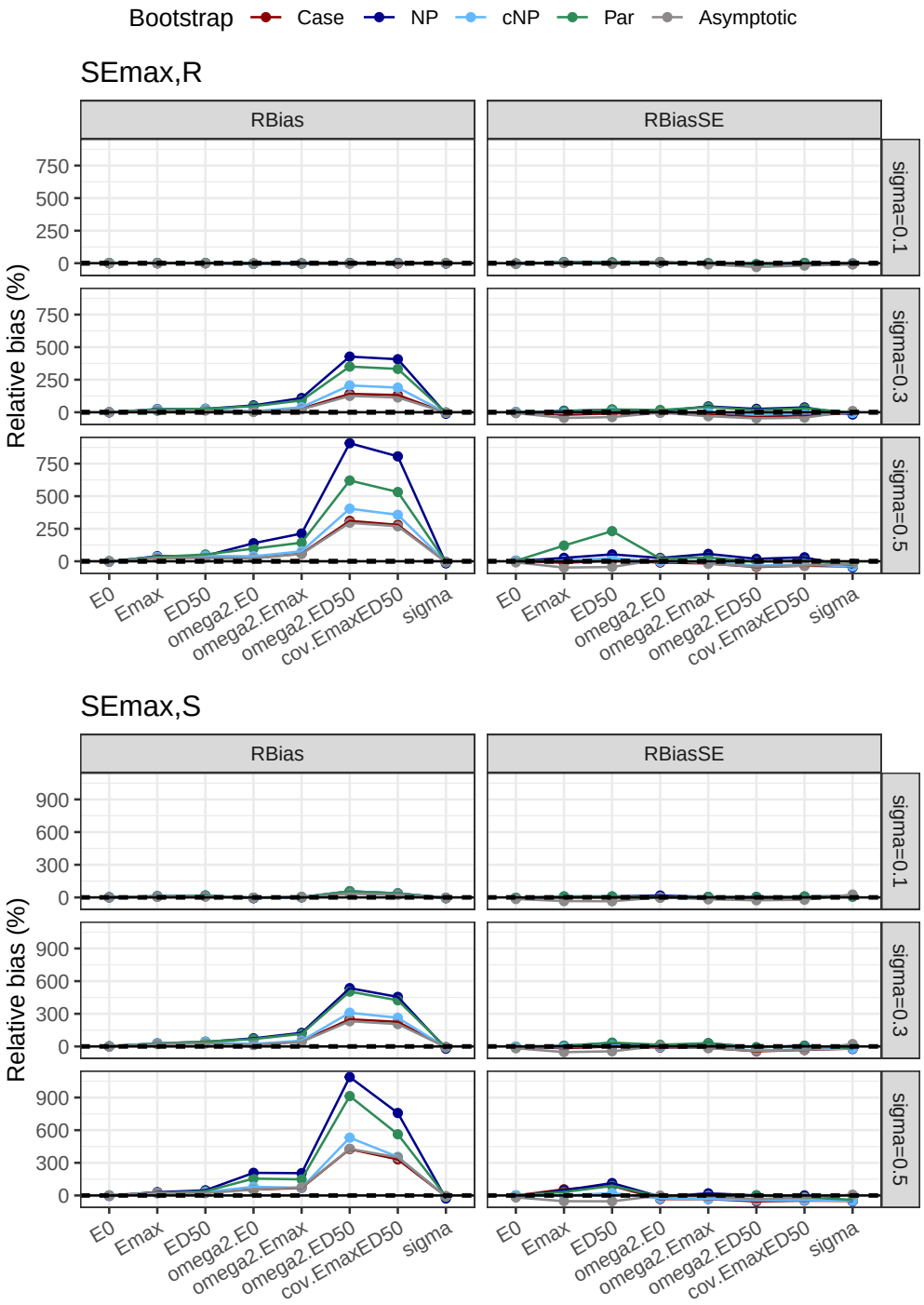


Fig. 18 Relative bias on parameters (up) and SE (down) for the original dosing scheme and the two scenarios with increased error coefficient $\sigma = 0.3$ and $\sigma = 0.5$ under the S_{Emax}. Dashed lines delineate absolute relative biases within 10%. A fixed scale is used on the Y-axis