Estimating the uncertainty of parameters in non-linear mixed-effects models by bootstrap: simulation studies to compare different bootstrap methods

Emmanuelle Comets[1,4] & Hoai-Thu Thai[1,2] & Christine Veyrat-Follet[2] & Nicholas HG Holford[3] & France Mentré[1]

[1] *IAME, UMR 1137, INSERM, 75018 Paris, France; Univ Paris Diderot, Sorbonne Paris Cité, 75018 Paris, France*

[2] *Drug Disposition Department, Sanofi, Paris, France*

[3] *Department of Pharmacology and Clinical Pharmacology, University of Auckland, Auckland, New Zealand*

[4] *CIC 0203, U. Rennes-I, CHU Pontchaillou, 35000 Rennes, France*

Acknowledgments: this work was performed by Hoai-Thu Thai during her PhD.

Résumé. Les modèles non-linéaires à effets mixtes sont utilisés abondamment en pharmacométrie pour décrire l'évolution dans le temps des concentrations de médicaments et de la réponse clinique, et sont des outils adaptés aux données longitudinales recueillies dans les essais cliniques. Le développement récent de nouvelles méthodes et algorithmes d'estimation, couplé à la puissance grandissante des capacités de calcul, permettent aujourd'hui de construire des modèles de plus en plus complexes. Ces méthodes permettent d'obtenir des estimations des paramètres ainsi que de l'incertitude associée, sous forme d'erreurs d'estimation. Dans les modèles complexes, ces erreurs d'estimation sont parfois difficiles à obtenir ou biaisées. Dans le présent travail, nous avons exploré des méthodes bootstrap comme alternatives à l'approximation asymptotique généralement utilisée. Nous avons notamment comparé des méthodes ré-échantillonnant les deux niveaux de variabilités présents dans les modèles mixtes, la variabilité inter et intraindividuelle, par des études de simulation dans les modèles linéaires et non-linéaires. Nos simulations montrent que le bootstrap peut améliorer les estimations asymptotiques pour certains paramètres, en particulier dans les modèles fortement non-linéaires. Cependant, certaines situations, fréquemment rencontrées en pratique, rendent le bootstrap difficile à utiliser, notamment quand le protocole expérimental est très déséquilibré ou quand certains paramètres sont mal estimés dans le modèle.

Mots-clés. Bootstrap, modèles non-linéaires à effets mixtes, pharmacocinétique, pharmacométrie, erreurs d'estimation sur les paramètres, données longitudinales.

Abstract.

Nonlinear mixed effect models are intensively used in pharmacometry to describe the evolution of drug concentrations and clinical responses with time, and are adapted to

handle longitudinal data collected in clinical trials. Over the past decades, major improvements in estimation methods and computer algorithms, along with increases in computing capacities, have expanded the range and complexity of models that can be explored. Estimation methods provide not only estimates of the parameters, but assess also the uncertainty in these parameter estimates. However, the estimates of standard errors of estimation in complex models may be biased or sometimes cannot be obtained. In the present work, we investigated by simulation the uncertainty obtained by different bootstrap methods. We explored bootstrap methods resampling two levels of variability (between subject and residual) in nonlinear mixed-effects models (NLMEM), comparing them to the uncertainty obtained by an asymptotic approach. We show that the bootstraps only provide better estimates of uncertainty in NLMEM with high nonlinearity compared to the asymptotic method. The case bootstrap, which resamples at the level of the individual, performs as well as the nonparametric bootstrap of both random effects and residuals. However, the improvement gained using bootstrap depends on the setting and the parameters considered, and bootstrap can be difficult to perform for instance in unbalanced designs, where the stratification approaches we considerd were not sufficient.

Keywords. Bootstrap, non-linear mixed-effects models, pharmacokinetics, pharmacometry, estimation error on parameters, longitudinal data.

Introduction

Pharmacometry, the science of quantitative pharmacology [1], makes heavy use of nonlinear mixed effect models to describe the factors influencing the variability in drug response and drug kinetics. The data collected in these studies are longitudinal data, characterised by repeated measurements in a number of individual units (such as subjects). Nonlinear mixed effect models describe the typical evolution of a process through a common mathematical function, and handle the variability through subject-specific random effects, which are characterised by a statistical distribution. These hierarchical models have been widely used to investigate the pharmacokinetics (PK) and pharmacodynamics (PD) of drugs in clinical trials during drug development, as well as to help provide adequate dosage regimens through therapeutic drug monitoring. This field has seen recently a surge of methodological developments, with new and powerful estimation methods now implemented in dedicated as well as general statistical software [2, 3, 4].

Standard errors (SE) of parameter estimates in population PKPD analyses are generally computed by estimation software using an asymptotic approximation involving the Fisher information matrix. Bootstrap approaches, which compute a distribution of the estimate through resampling, provide an alternative estimate of SE and confidence intervals [5]. They have been extensively studied in statistics to provide the distribution of an estimator or a test statistic [6], and applied to parameter estimates in linear and nonlinear regression [7]. They consist in repeatedly resampling from an initial set of values to obtain the distribution of an estimator. The simplest approach is the so-called paired or

case bootstrap. In regression analysis, the case bootstrap consists in resampling pairs of dependent and independent variables with replacement from the pool of such pairs present in the original dataset. With longitudinal data, case bootstrap consists in resampling at the individual level the entire vector of observations. This type of resampling is considered to be non-parametric since no assumption is made on the distribution of the observations. Several alternative bootstrap approaches have been proposed in the context of regression analysis: residual bootstrap, consisting in resampling empirical residuals from the fit of the model to the original dataset, parametric bootstrap, which resamples residuals from the estimated distribution, or external bootstrap, which uses a distribution with adequate properties. In mixed-effect models, bootstrapping has not been well studied. To date, most PK/PD applications employed the non-parametric case bootstrap, because of its simplicity [8]. Das [9] summarises the extensions of the different types of bootstrap to mixed-effect models. The idea behind the bootstrap is that the resampling procedure should respect the true data generating process, with repeated measures within a subject. In mixed-models, it should therefore handle the two levels of variability in these models, between-subject (BSV) and residual (RUV), and residual-based bootstrap methods which resample both random effects and residuals may be a good way to approach the data generating process [10].

The aims of this work were a) to present and extend different bootstrap methods which can be applied for mixed-effect models, and b) to evaluate the performance of these methods to estimate uncertainty of parameters using simulation studies. As bootstrap is a very time-consuming approach, we first evaluated different methods in linear mixed effect models to eliminate the methods with poor performance, and then performed a comparison of the best methods in a non-linear mixed effect model. The proposed methods were applied to the real datasets in both studies.

Methods

In the first part of this work [11], we reviewed bootstrap methods and applied them to linear mixed-effects models. Different bootstrap approaches were investigated, including parametric and non-parametric bootstraps of both BSV and RUV, residual-alone bootstraps and the non-parametric case bootstrap. We investigated corrections of random effects and residuals using the ratio between empirical and estimated variance-covariance matrices [12] to account for the underestimation of variances.

The motivating example for the simulation study was a subset of a clinical study describing the natural evolution of Parkinson's disease over a 2-year drug-free period, where the clinical response was modelled by a linear change in a composite score measuring the degradation of cognitive function [13]. Different designs were simulated: a rich design (N=100 subjects and n=7 samples), a sparse design (N=30/n=3) in which the variance parameters are less well estimated, and a large error design (N=100/n=7).

The bootstrap methods were implemented in R [14], using the lme function 2.14.1 from

the nlme package to fit the data. We compared two estimation methods, restricted maximum likelihood (REML) predominantly used in linear mixed-effect models, and Maximum likelihood (ML), generally used in non-linear mixed-effect models. For each design, 1000 replicates were simulated, with 1000 bootstrap samples per replicate for each bootstrap method. The bootstrap approaches were compared with respect to bias of parameters, standard errors (SE) and coverage rate of the 95% confidence intervals.

We then investigated the performance of the three best bootstrap approaches in a more complex setting involving model non-linearity as well as heteroscedasticity in the residual error [15]. Real data from a PK study on aflibercept (Zaltrap^{TM}) [16], a novel anti-VEGF drug, was collected from two clinical trials in cancer patients, and used to design the simulation study in non-linear mixed-effect models. In the simulations, we used a simplified PK model describing only free aflibercept PK using a two-compartment infusion model with first order elimination for the sake of computational time.

The SAEM algorithm implemented in MONOLIX 4.1.2 was used to obtain the ML estimates of the model parameters. Again we considered different designs: a balanced setting, evaluating both a rich (N=30/n=9) and a sparse design (N=70/n=4), and a an unbalanced setting reflecting the structure of the real dataset, with a mix of rich and sparse sampling. An additional simulation was performed with a Michaelis-Menten elimination instead of first-order elimination in the rich balanced design to increase model non-linearity. We investigated stratification approaches for the unbalanced design. The performance of each bootstrap method was assessed as in the linear case.

Results

With linear mixed-effect models, our simulation studies evidenced a good performance of the case bootstrap and the non-parametric/parametric bootstraps of both random effects and residuals, across the different designs we investigated. On the other hand, the bootstrap methods which resampled only the residuals and the bootstraps combining case and residuals performed poorly. The asymptotic method showed a good performance, as expected for the linear mixed-effect model. REML and ML provided similar bootstrap estimates of uncertainty, but there was slightly more bias and poorer coverage rate for variance parameters with ML in the sparse design.

The different methods were then applied to the real dataset. The different bootstrap distributions yielded similar values for the mean estimates, and these estimates were close to those provided by the asymptotic method. However, we found differences in the estimates of SE for some parameters, which was in part due to the unbalanced design.

The case bootstrap and the bootstraps resampling residuals and random effects either parametrically or non-parametrically were then applied to the non-linear models in the second simulation study. Again, the asymptotic method performed well in most cases. With balanced designs, the case bootstrap works slightly better than the non-parametric

bootstrap of random effects and residuals but less well than the parametric bootstrap. Compared to the asymptotic approach, the bootstrap methods provided better description of uncertainty of some parameters, especially parameters which enter the model most non-linearly in the frequent sampling designs. However, they performed less well in the sparse design, in particular for Q which was poorly estimated. In the unbalanced design, the case bootstrap overestimated the SE of Q, while the non-parametric residual bootstrap overestimated the SE of variance parameters even with stratification while the asymptotic method and the parametric residual bootstrap showed adequate estimates of SE. Stratification improved the estimates of SE for some parameters but degraded others, even in this simple design with only two groups.

Conclusion

In this work, we extended and evaluated different bootstrap methods for mixed-effect models. The asymptotic method performed well in most cases while the bootstrap methods provided better estimates of uncertainty for parameters with high non-linearity. However, the bootstrap methods face several practical problems. They can generate a wrong estimate of the SE of a parameter with a skewed distribution when the parameters are not well estimated, meaning that bootstrap cannot correct for poor estimation. Our simulation study also showed that bootstrap may improve the estimates of standard errors of the parameters in some settings, but perform worse than the asymptotic approaches under different conditions, which makes it difficult to assess the added value of bootstrap in practical studies. In addition, bootstrapping in unbalanced designs is much more challenging, and stratification may be insufficient to correct for heterogeneity especially in very unbalanced designs.

References

[1] Ette, E. et Williams, P. (2007), *Pharmacometrics: the science of quantitative pharmacology*, Wiley-Interscience, Hoboken, New Jersey.

[2] Boeckmann, A., Sheiner, L. et Beal, S. (1994), *NONMEM Users Guide: Part V*, University of California, NONMEM Project Group, San Francisco.

[3] Lixoft (2013), *MONOLIX (MOdèles NOn LInéaires à effets miXtes)*, INRIA, Orsay, France.

[4] Comets, E., Lavenu, A. et Lavielle, M. (2011), SAEMIX, an R version of the SAEM algorithm, *20^{th} meeting of the Population Approach Group in Europe, Athens, Greece*, abstr 2173.

[5] Ette, E.I. (1997), Stability and performance of a population pharmacokinetic model, *Journal of Clinical Pharmacology*, 37, 486–495.

[6] Efron, B. et Tibshirani, R.J. (1994), *An introduction to the bootstrap*, Chapman & Hall, New York.

[7] Shao, J. et Tu, D. (1995), *The jacknife and bootstrap*, Springer, New York.

[8] Parke, J., Holford, N.H.G. et Charles, B.G. (1999), A procedure for generating bootstrap samples for the validation of non-linear mixed-effects population models, *Computer Methods and Programs in Biomedicine*, 59, 19–29.

[9] Das, S. et Krishen, A. (1999), Some bootstrap methods in non-linear mixed-effects models, *Journal of Statistical Planning and Inference*, 75, 237–245.

[10] Ocana, J., El-Halimi, R., Ruiz de Villa, M.C. et Sanchez, J.A. (2005), *Bootstrapping repeated measures data in a non-linear mixed-models context*, Mathematics Preprint Series.

[11] Thai, H., Mentré, F., Holford, N.H., Veyrat-Follet, C. et Comets, E. (2013), A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models, *Pharmaceutical Statistics*, 12, 129–40.

[12] Carpenter, J.R., Goldstein, H. et Rasbash, J. (2003), A novel bootstrap procedure for assessing the relationship between class size and achievement, *Applied Statistics*, 52, 431–443.

[13] Holford, N.H.G., Chan, P.L.S., Nutt, J.G., Kieburtz, K., Shoulson, I. et al. (2006), Disease progression and pharmacodynamics in Parkinson disease-evidence for functional protection with levodopa and other treatments, *Journal of Pharmacokinetics and Pharmacodynamics*, 33, 281–311.

[14] 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.

[15] Thai, H., Mentré, F., Holford, N.H., Veyrat-Follet, C. et Comets, E. (2013), 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*, *in press*.

[16] Thai, H., Veyrat-Follet, C., Vivier, N., Dubruc, C., Sanderink, G. et al. (2011), A mechanism-based model for the population pharmacokinetics of free and bound aflibercept in healthy subjects, *British Journal of Clinical Pharmacology*, 72, 402–14.