# A comparison of bootstrap approaches for estimating standard error of parameters in linear mixed effects models


H.T. THAI [1], C. Veyrat-Follet [2], F. Mentré [1], E. Comets [1]

(1) INSERM UMR738, University Paris Diderot-Paris 7, Paris, France
(2) Modeling and Simulation Group, Drug Disposition Department, Sanofi, Paris, France


## CONTEXT

- Standard errors (SE) of parameters are usually obtained by the classical asymptotic approach via the inverse of Fisher information matrix
- The bootstrap, introduced by Efron (1979)[1] is an alternative approach to obtain the distribution of estimators such as SE or confidence intervals
- The principle of bootstrap is to resample with replacement from the original data to create replicate datasets with the same sample size
- In PK/PD, case bootstrap (paired nonparametric bootstrap) has been frequently used, but never compared to other bootstrap alternatives which better take into account the structure of longitudinal data[2,3,4]

## OBJECTIVES

- Study and propose appropriate bootstrap methods in mixed effects models, focusing first on linear models (LME)
- Evaluate the performance of proposed bootstrap methods by simulation

## METHODS

### Bootstrap methods for mixed-effects models

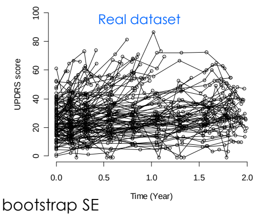
- Resample two levels of variability in the longitudinal data
  - between-subject variability: resample the entire subjects (case) or the random effects ($\eta$)
  - residual variability: resample the residuals from all subjects (global residual, GR) or the residuals within each subject (individual residual, IR)
- Two versions of bootstrap
  - nonparametric bootstrap: resample from the empirical distribution
    - correction of random effects & residuals by the ratio between the empirical and estimated variance-covariance matrix[4]
  - parametric bootstrap: simulate within the estimated distribution

| | | Variability related to subject | | |
|---|---|---|---|---|
| | | None | Resample the subjects | Resample the random effects |
| Variability related to observations | None | Original dataset | $B_{case,none}$ | |
| | Resample the residuals globally | $B_{none,GR}$ $B_{none,PR}$ | $B_{case,GR}$ $B_{case,PR}$ | $B_{\eta,GR}$ $B_{P\eta,PR}$ |
| | Resample the residuals individually | $B_{none,IR}$ | $B_{case,IR}$ | $B_{\eta,IR}$ |

GR: Global residual
PR: Parametric residual
IR: Individual residual
$\eta$: Random effects
P$\eta$: Parametric random effects

### Simulation study

- Motivating example (courtesy of Prof. Nicholas Holford):
  - subset of placebo group with UDPRS (Unified Parkinson's Disease Rating Scale) score from entry to 2 years
  - linear disease progression model describing natural evolution of Parkinson's disease[5]: $S(t) = S_0 + \alpha.t$



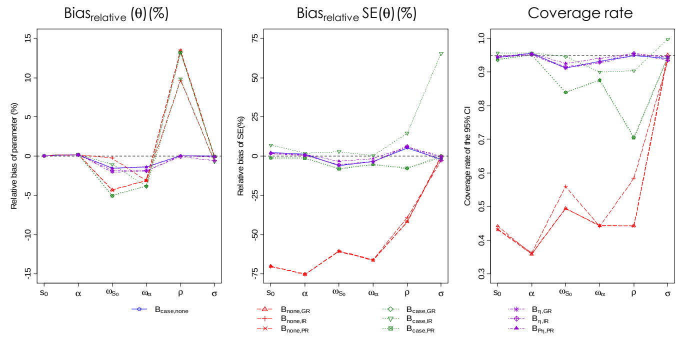

- Three designs:
  - rich design: N=100, n=7, $\sigma$=5.86
  - sparse design: N=30, n=3, $\sigma$=5.86
  - design with large error: N=100, n=7, $\sigma$=17.5
- Number of simulated replication K=1000
- Number of bootstrap per replication B=1000
- Evaluation criteria:
  - empirical SE: "true" SE to calculate relative bias of bootstrap SE
  - relative bias of bootstrap parameter estimates and their SE: no bias (±5%), moderate (±5% to ±10% ), important (>±10%)
  - coverage rate of the 95% bootstrap CI: good (90-100%), low (80-90%), poor (<80%)

### Application to real dataset

- Application of proposed bootstrap methods to the real dataset
  - compare mean and SE of bootstrap estimates

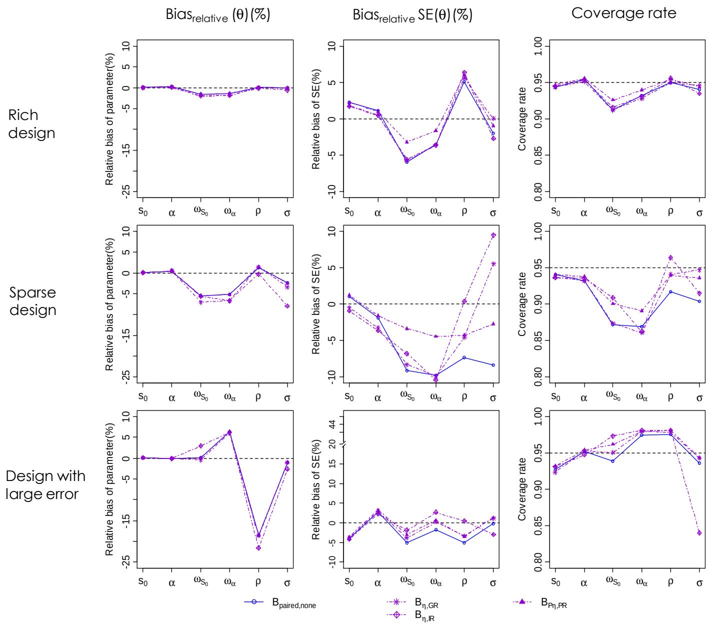
## RESULTS

### Performance of bootstrap methods in the rich design




- As expected, bootstrapping only residuals underestimates greatly SEs of parameters except for $\sigma$ and provides poor coverage rate
- Case bootstrap ($B_{case,none}$) works well although only the between-subject variability is resampled
- Case bootstrap and bootstrap of both random effects & residuals ($B_{case,none}$, $B_{\eta,GR}$, $B_{\eta,IR}$, $B_{P\eta,PR}$) perform well and are selected as the bootstrap candidates
  - Correction of random effects and residuals improves the estimates for variance parameters and their SEs, particularly for $\sigma$

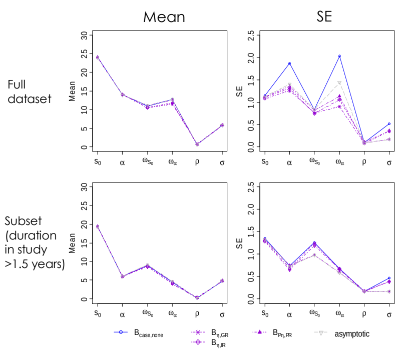
### Performance of bootstrap candidates




- The bootstrap candidates perform well in the rich and large error designs but less well in the sparse design
- $B_{P\eta,PR}$ works slightly better than $B_{case,none}$
- $B_{\eta,GR}$ performs slightly better than $B_{\eta,IR}$

## APPLICATION TO REAL DATASET

- The bootstrap candidates give similar estimates for all parameters, which are also close to asymptotic estimates
- $B_{case,none}$ gives different results for SE of $\alpha$ and $\omega_\alpha$ in the full dataset
- In the subset with patients staying until 1.5 years, similar performance of bootstrap candidates and the asymptotic approach is observed
  - drop-out influences $B_{case,none}$ more than other bootstraps




## CONCLUSIONS

- The four bootstrap methods ($B_{case,none}$, $B_{\eta,GR}$, $B_{\eta,IR}$, $B_{P\eta,PR}$) are selected as the bootstrap candidates due to their good performance in the evaluated designs
- Case bootstrap works well in linear-mixed effects models although only the between-subject variability is resampled
- Parametric bootstrap of random effects and residuals works slightly better than the case bootstrap in our simulations, but may not be as robust to model or distributional misspecifications

## REFERENCES


1. Efron B. *Annals of Statistics* 1979 ;7 :1-26
2. Das S, Krishen A. *Journal of Statistical Planning and Inference* 1999;75: 237-245
3. Ocana J et al. *Mathematics Preprint Series* 2005.
4. Wang J et al. *Computer Methods and Programs in Biomedicine* 2006;82:130-143
5. Holford NHG et al. *Journal of pharmacokinetics and pharmacodynamics* 2006; 33(3): 281-311



*Acknowledgments*: We would like to thank Prof. Nicholas Holford for providing us the data from Parkinson's study and for his valuable advices.