Inserm

A comparison of estimation methods in nonlinear mixed effects models using a blind analysis

Pascal Girard, PhD
EA 3738 CTO, INSERM, University Claude Bernard Lyon I, Lyon
France Mentré, PhD, MD
Dept Biostatistics, INSERM U738, University Paris VII, Paris
France

Inserm

Project objectives

- Compare various “recently” developed
  and
- classical algorithms for non-linear mixed effect models using ML parametric estimate:
  - Bias
  - Precision
  - Standard Error estimates

Inserm

The simulation model: one compartment, 1st order absorption reparametrized to avoid flip-flop

$$CP(t) = \frac{Dose * Ka}{V * (Ka\text{-}Ke)} \left(\exp(-Ke \cdot t) - \exp(-Ka \cdot t)\right) * \exp(\varepsilon)$$

With

- $V = \theta_1 \times EXP(\eta_1)$
- $KE = \theta_2 \times EXP(\eta_2)$
- $KA = KE + \theta_3 \times EXP(\eta_3)$
- θ_1 , θ_2 , θ_3 fixed effect parameters, greater than 0, for *V*, *KA* and *KA-KE*;
- η_1, η_2, η_3 are random effects representing between individual variability normally distributed, with a block VARIANCE matrix Ω :

$$(\eta_1 , \eta_2 , \eta_3) \sim N\left((0,0,0), \begin{pmatrix} \varpi_1^2 & \varpi_{21} & \varpi_{31} \\ \varpi_{21} & \varpi_2^2 & \varpi_{32} \\ \varpi_{31} & \varpi_{32} & \varpi_3^2 \end{pmatrix}\right)$$

- ε random normally distributed effect representing within residual variability with $\varepsilon \sim N(0, \sigma^2)$.

Inserm

Simulation design

- 100 simulated datasets
- 100 individuals (ID) / dataset
- 4 samples / ID at times 1, 4, 12, 24 hrs + random shift:
  - $\varepsilon_1 \sim N(0, SD=0.25)$ at 1 and 4 hrs
  - $\varepsilon_2 \sim N(0, SD=2)$ at 12 and 24 hrs
- Missing points at random P(missing)=0.25

Inserm

Parameters used for simulation

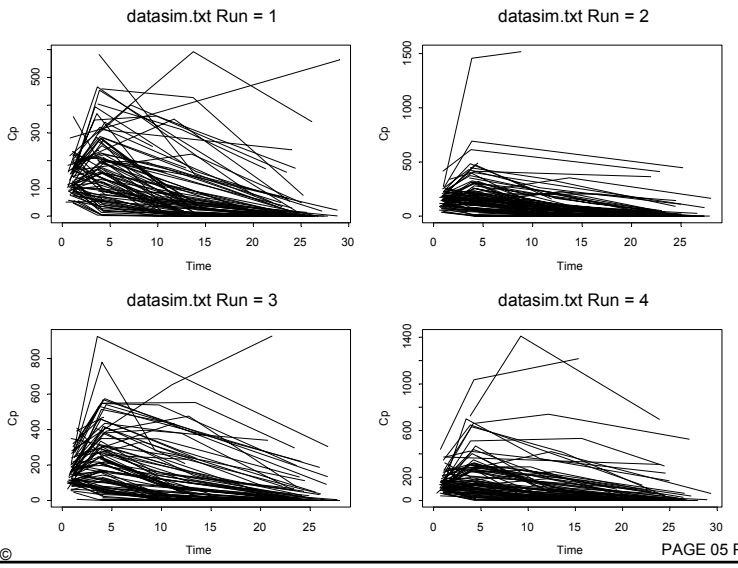
	V	Ke	Ka-Ke
Value	27.2	0.232	0.304
CV IIV	47%	81%	15%

Corr IIV	V	Ke	Ka-Ke
V	1	---	---
Ke	0.91	1	---
Ka-Ke	0.33	0.01	1

CV Resid = 25%

Inserm

Example of 4 simulated datasets



Inserm

Blinded analysis

- Exact model was given to each participant who ignored the true simulation parameters
- Suggested initial parameter estimates provided **only for fixed effects**:
  - V = 30 (true=27.2)
  - Ke = 0.3 (true = 0.232)
  - Ka-Ke =0.5 (true=0.304)
- Data sets blinded and send on 7th March 2005;
- Results received between March 25th and June 10th 05

Inserm

Algorithms involved in the comparison

Approximate Likelihood

- SAS proc nlmixed FO approximation
- NONMEM V and VI FOCE and nlme (Splus 6.2)
  - Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. Biometrics 1990; 46:673-87.
- ITBS (MW\Pharm & MulTifit)
  - Mentré F, Gomeni R. A two-step iterative algorithm for estimation in nonlinear mixed-effect models with an evaluation in population pharmacokinetics. J Biopharm Stat 1995; 5: 141-158.

Exact Likelihood

3 Stochastic methods

- PEM
- MCPEM
- SAEM

One method based on approximated integral of exact likelihood

- SAS proc nlmixed Adaptive Gaussian Quadrature:
  - 3 points per dimension (27 points in all) used to integrate the multidimensional exact likelihood

Inserm

Compilation of results:

Relative error, bias and Root Mean Square Error

- For Ω, computation of $\sqrt[3]{\mathrm{Det}\left(\Omega_{est}\Omega_{true}^{-1}\right)}$ with true=1
- Relative Estimation eRror (RER) for each simulation presented as boxplots: RER= (Est – True) / True × 100
  - Wilcoxon test for testing **H0:** RER=0 vs. **H1:** RER≠0
- Bias % = 100 × mean(Est – True) / True
- **Precision**: relative Root Mean Square Error (RMSE) :

$$\%RMSE = 100\sqrt{\mathrm{mean}\left(\left(Est - True\right)^2\right)/\,True^2}$$

➔ Results send out to participants on June 6th

Inserm

RESULTS

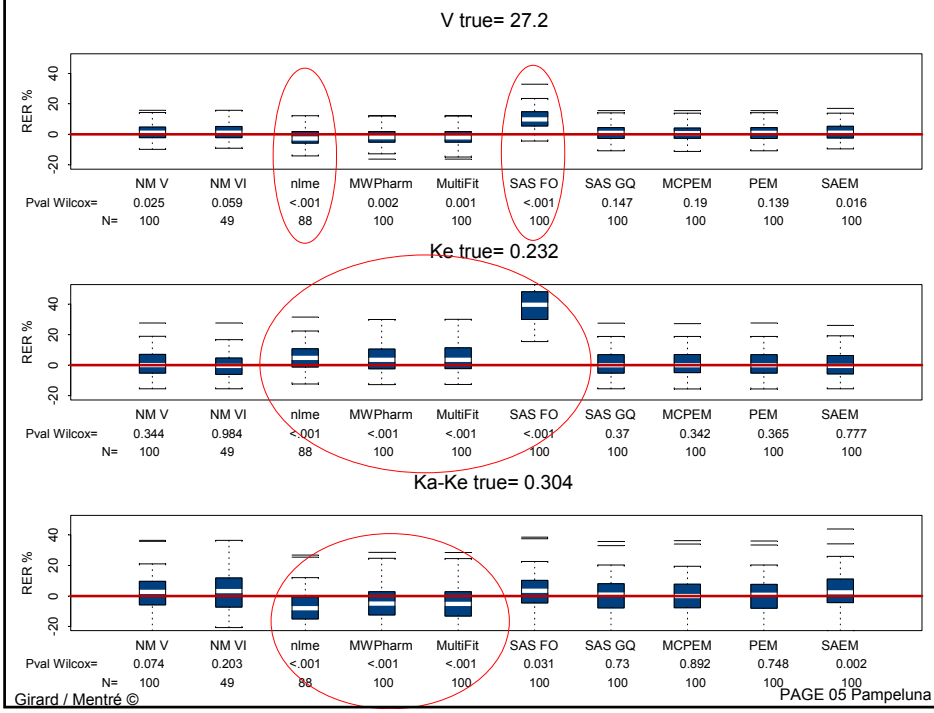
Inserm

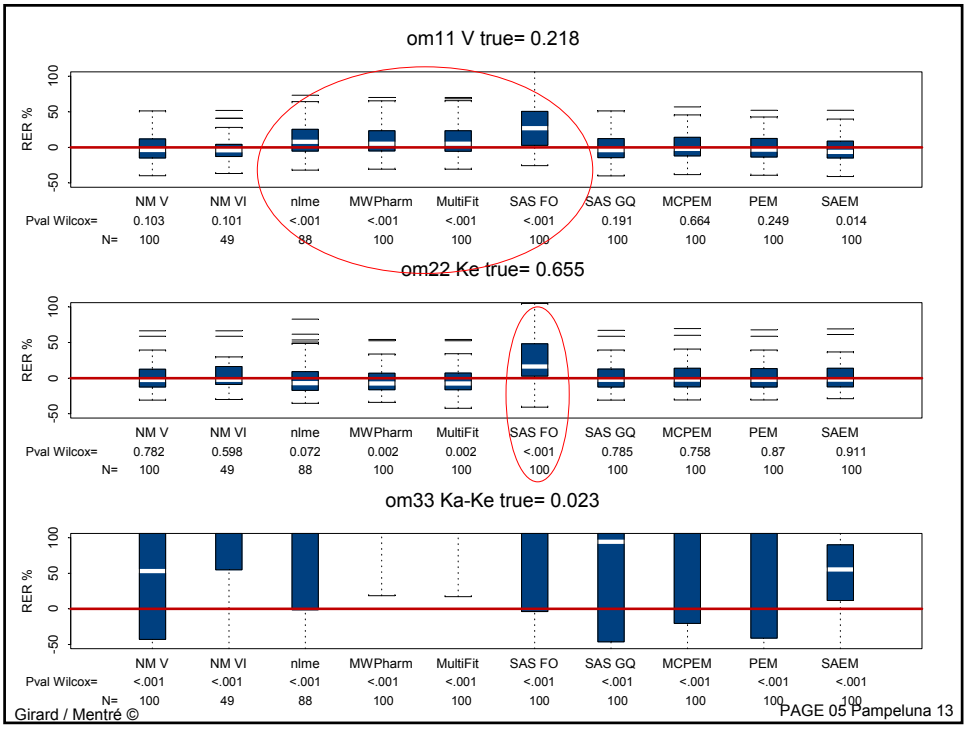
All algorithms converged and provided estimates for all simulated data sets except FOCE (NONMEM and nlme)

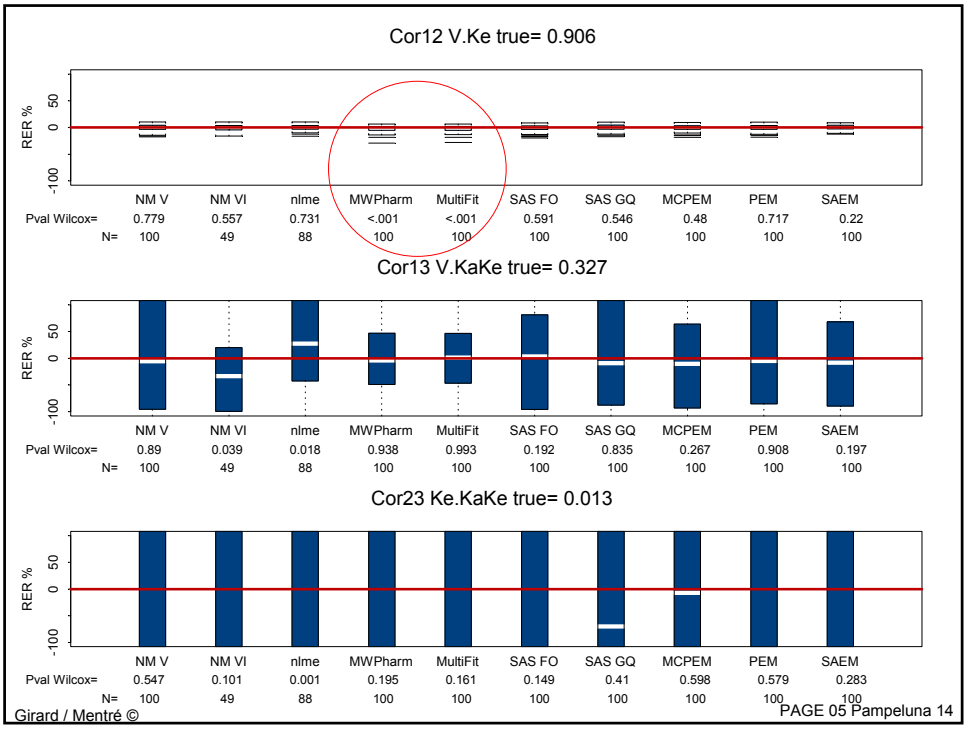
	nlme 1st try [1]	nlme 2nd try	NONMEM Version V [2]	NONMEM Version VI [2] (Beta)
Only Est. step successful	100%	88%	100%	49%
Est and COV successful	100%	70%	47%	49%

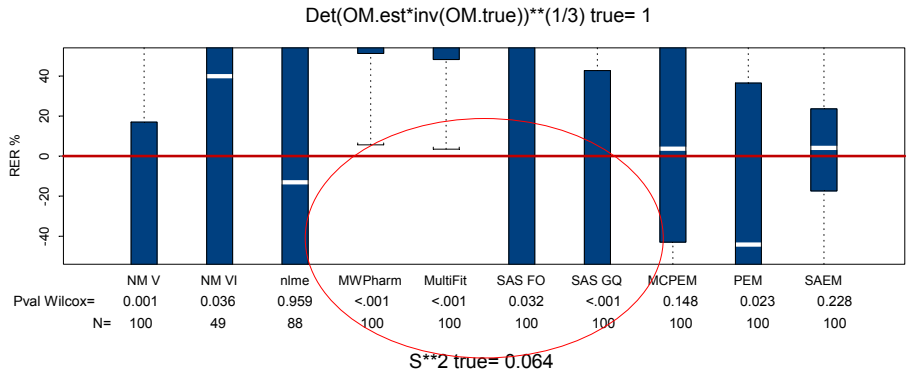
1. Biased estimates with 1st set of results ➔ 2nd set of results were
2. repeated search with perturbed initial estimates

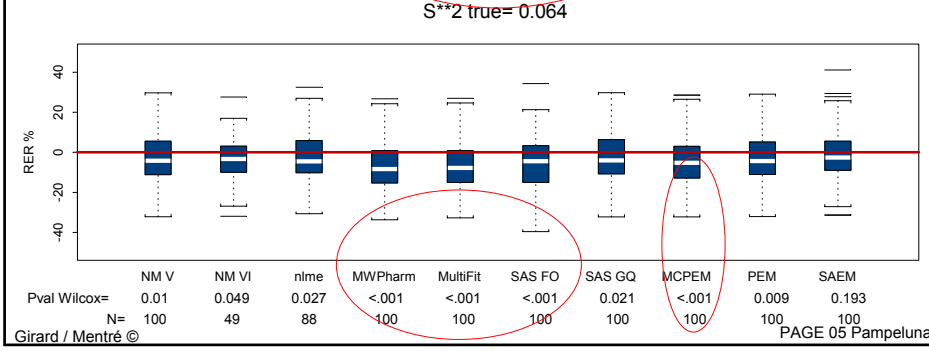
Almost perfect matching on failed run# 7, 9, 28, 45, 47, 53, 68, 69, 76, 80, 90




om11 V true= 0.218
RER %
100
50
0
-50
NM V NM VI nlme MWPharm MultiFit SAS FO SAS GQ MCPEM PEM SAEM
Pval Wilcox= 0.103 0.101 <.001 <.001 <.001 <.001 0.191 0.664 0.249 0.014
N= 100 49 88 100 100 100 100 100 100 100
om22 Ke true= 0.655
RER %
100
50
0
-50
NM V NM VI nlme MWPharm MultiFit SAS FO SAS GQ MCPEM PEM SAEM
Pval Wilcox= 0.782 0.598 0.072 0.002 0.002 <.001 0.785 0.758 0.87 0.911
N= 100 49 88 100 100 100 100 100 100 100
om33 Ka-Ke true= 0.023
RER %
100
50
0
-50
NM V NM VI nlme MWPharm MultiFit SAS FO SAS GQ MCPEM PEM SAEM
Pval Wilcox= <.001 <.001 <.001 <.001 <.001 <.001 <.001 <.001 <.001 <.001
N= 100 49 88 100 100 100 100 100 100 100


Cor12 V.Ke true= 0.906
RER %
50
0
-100
NM V NM VI nlme MWPharm MultiFit SAS FO SAS GQ MCPEM PEM SAEM
Pval Wilcox= 0.779 0.557 0.731 <.001 <.001 0.591 0.546 0.48 0.717 0.22
N= 100 49 88 100 100 100 100 100 100 100
Cor13 V.KaKe true= 0.327
RER %
50
0
-100
NM V NM VI nlme MWPharm MultiFit SAS FO SAS GQ MCPEM PEM SAEM
Pval Wilcox= 0.89 0.039 0.018 0.938 0.993 0.192 0.835 0.267 0.908 0.197
N= 100 49 88 100 100 100 100 100 100 100
Cor23 Ke.KaKe true= 0.013
RER %
50
0
-100
NM V NM VI nlme MWPharm MultiFit SAS FO SAS GQ MCPEM PEM SAEM
Pval Wilcox= 0.547 0.101 0.001 0.195 0.161 0.149 0.41 0.598 0.579 0.283
N= 100 49 88 100 100 100 100 100 100 100





Inserm

Bias % for fixed effects

±<5% ±5-10% ±10-20% ±>20%

	TH1 (V=27.2)	TH2 (Ke=0.23)	TH3 (Ka-Ke=0.30)
NM V	1.2	1.1	2.1
NM VI *	1.8	0.5	2.7
nlme**	-2.1	5.2	-7.0
MWPharm	-1.8	4.3	-4.6
MultiFit	-1.9	4.4	-5.1
SAS FO	9.8	39.6	2.5
SAS GQ	0.9	1.0	0.4
MCPEM	0.8	1.1	0.3
PEM	0.9	1.0	0.3
SAEM	1.4	0.5	3.7

*** NM VI based on 49 successful runs out of 100**
**** nlme "2nd chance" and 88 successful runs out of 100**

Inserm

Bias % for variance omega**2

±<20% ±20-50% ±50-100% ±>100%

	om V (tr=0.22)	om Ke (tr=0.65)	om33 (tr=0.02)
NM V	-2.8	0.5	185.0
NM VI *	-2.7	3.0	342.4
nlme**	12.2	1.2	266.5
MWPharm	8.8	-4.6	434.7
MultiFit	9.1	-4.7	433.1
SAS FO	31.9	29.3	433.4
SAS GQ	-2.1	0.6	177.6
MCPEM	-0.1	1.6	229.8
PEM	-1.7	0.8	197.9
SAEM	-4.1	1.4	56.8

*** NM VI based on 49 successful runs out of 100**
**** nlme "2nd chance" and 88 successful runs out of 100**

Inserm

Bias % for correlation coefficients

±<20% ±20-100% ±100-500% ±>500%

	V-Ke (tr=0.91)	V-KaKe (tr=0.33)	Ke-KaKe (tr=0.01)
NM V	-0.4	-9.4	374.0
NM VI *	-0.7	-30.0	-497.4
nlme**	0.0	27.4	1330.1
MWPharm	-2.3	-2.9	300.4
MultiFit	-2.1	-2.1	319.7
SAS FO	-0.3	-32.1	539.1
SAS GQ	0.1	-9.8	483.9
MCPEM	-0.6	-17.7	248.5
PEM	-0.4	-7.7	325.5
SAEM	0.5	-23.0	284.5

*** NM VI based on 49 successful runs out of 100**
**** nlme "2nd chance" and 88 successful runs out of 100**

Inserm

Bias % for det(Ω.inv(Ω_{tr})) and S^2

±<10% ±10-20% ±20-50% ±>50%

	det(OM.inv(OMtr)) (true=1)	S**2 (true=0.06)
NM V	-25.9	-3.3
NM VI *	29.2	-3.7
nlme**	6.9	-2.7
MWPharm	84.4	-7.5
MultiFit	80.3	-7.0
SAS FO	34.1	-5.6
SAS GQ	-35.1	-2.8
MCPEM	15.9	-4.5
PEM	-12.2	-3.1
SAEM	4.0	-1.3

*** NM VI based on 49 successful runs out of 100**

**** nlme "2nd chance" and 88 successful runs out of 100**

Inserm

RMSE Results

Inserm

RMSE % for fixed effects

<10% 10-20% 20-50% >50%

	TH1 (V=27.2)	TH2 (Ke=0.23)	TH3 (Ka-Ke=0.30)
NM V	5.5	8.2	11.4
NM VI *	5.9	8.5	12.7
nlme**	5.7	9.7	12.7
MWPharm	5.7	9.2	12.1
MultiFit	5.8	9.3	12.6
SAS FO	11.7	41.7	12.3
SAS GQ	5.4	8.1	11.6
MCPEM	5.4	8.1	11.7
PEM	5.4	8.1	11.7
SAEM	5.4	8.3	11.6

*** NM VI based on 49 successful runs out of 100**

**** nlme "2nd chance" and 88 successful runs out of 100**

Inserm

RMSE % for variance omega**2

<20% 20-50% 50-100% >100%

	om V (tr=0.22)	om Ke (tr=0.65)	om33 (tr=0.02)
NM V	19.4	18.1	358.8
NM VI *	18.3	19.8	471.1
nlme**	28.5	33.8	536.7
MWPharm	22.3	17.6	524.7
MultiFit	23	17.8	528.7
SAS FO	48.2	51.3	738.8
SAS GQ	19.2	18.2	331.3
MCPEM	18.9	18.5	381.9
PEM	19.2	18.3	360.9
SAEM	18.6	18.3	88.3

*** NM VI based on 49 successful runs out of 100**

**** nlme "2nd chance" and 88 successful runs out of 100**

Inserm

RMSE % for correlation coefficients

<20% 20-100% 100-500% >500%

	V-Ke (tr=0.91)	V-KaKe (tr=0.33)	Ke-KaKe (tr=0.01)
NM V	5.5	146.2	4012.8
NM VI *	5.9	101.3	2763.8
nlme**	5.1	120.1	3608.0
MWPharm	5.9	70.6	2042.2
MultiFit	5.8	73.2	2124.6
SAS FO	5.6	160.3	4095.7
SAS GQ	5.3	144.4	4012.4
MCPEM	5.2	118.0	3298.4
PEM	5.3	133.1	3724.8
SAEM	4.5	116.7	2863.7

*** NM VI based on 49 successful runs out of 100**

**** nlme "2nd chance" and 88 successful runs out of 100**

Inserm

RMSE % for det(Ω.inv(Ω_{tr})) and S^2

<20% 20-50% 50-100% >100%

	det(OM.inv(OMtr)) (true=1)	S**2 (true=0.06)
NM V	83.5	13.6
NM VI *	90.3	12.9
nlme	97.0	13.3
MWPharm	95.1	14.5
MultiFit	91.5	14.1
SAS FO	153.8	15.1
SAS GQ	91.9	13.4
MCPEM	72.4	13.5
PEM	74.2	13.3
SAEM	29.6	13.3

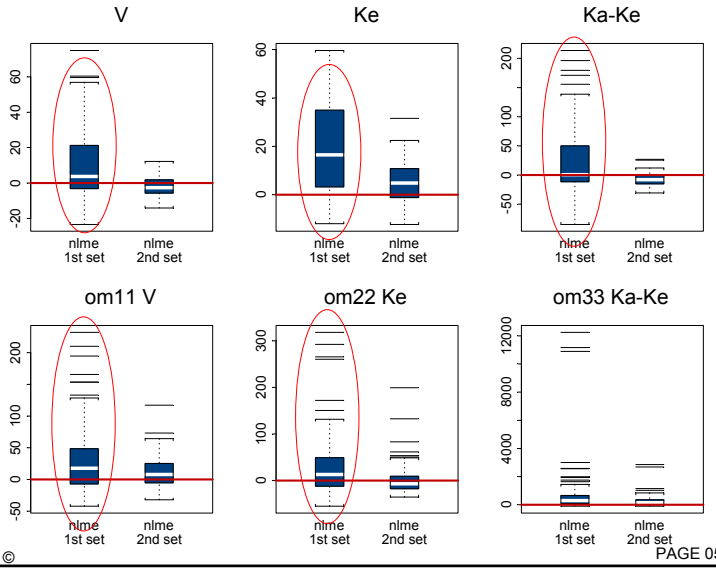
*** NM VI based on 49 successful runs out of 100**

**** nlme "2nd chance" and 88 successful runs out of 100**

Inserm

Comparison between 1st set and 2nd one for nlme:

«Because we were surprised with the blinded results regarding NLME, we went back to the code we used to produce the fits and realized that we had specified the model in a unstable way that created convergence problems in NLME. We re-defined the model formulation in S+ and re-ran the analyses. The fits were considerably more stable than before. »
Hsu & Pinheiro



Inserm

Use of non-parametric ranks to classify all methods

	Rank	TH1 (V=27.2)
NM V	7	1.2
NM VI *	4	1.8
nlme**	2	-2.1
MWPharm	5	-1.8
MultiFit	3	-1.9
SAS FO	1	9.8
SAS GQ	9	0.9
MCPEM	10	0.8
PEM	8	0.9
SAEM	6	1.4

$$Quote = \sum_{k=1}^{10} rank_k$$

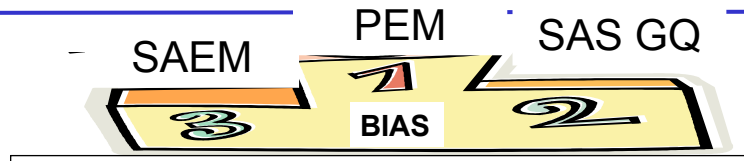
$$10 \le Quote \le 100$$

☹ ☺

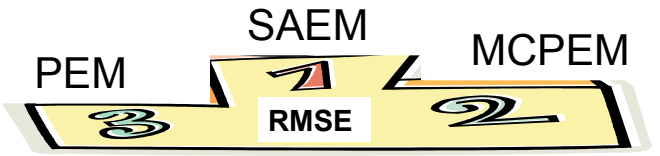
* NM VI is beta version: only 49% convergence
** nlme: only 88% convergence with 2nd try

Inserm

Classification of methods based on their sum of ranks for bias and RMSE



1	2	3	4	5	6	7	8	9	10
PEM	SAS GQ	SAEM	MCPEM	NM V	NM VI*	nlme**	MWPharm	MultiFit	SAS FO
84	80	79	77	74	51	51	40	39	30

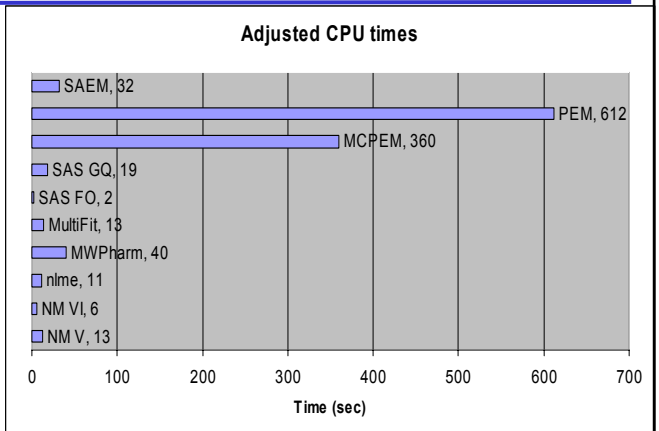


10	9	8	7	6	5	4	3	2	1
SAEM	MCPEM	PEM	SAS GQ	NM V	NM VI	MWPharm	MultiFit	nlme	SAS FO
89	78	73	70	64	59	58	53	44	17

Inserm

CPU Times
Adj_time=time_one_run × GHz

	GHz	CPU time for 1 run (sec)
SAEM[1]	1.6	20
PEM	1.7	360
MCPEM	1	360
SAS GQ	3.1	6
SAS FO	3.1	1
MultiFit	1.1	12
MWPharm	1.1	36
Nlme [2]	1.6	7
NM VI [3]	2.13	3
NM V [3]	2.13	6



1. A compiled version would probably be considerably faster than actual implementation in Matlab for PEM and SAEM
2. For a successful nls + nlme run
3. For a successful NONMEM run (EST & COV)

Inserm

Conclusions

- Methods without approximation of the likelihood (either stochastic methods or GQ) show better performances than others with an approximation.
- Reduction of BIAS and RMSE with stochastic methods is obtained only with slight CPU time increase for SAEM
- FO in SAS shows, as expected, large bias
- FOCE methods, either implemented in nlme or in NONMEM have either convergence problems or computational ones for covariance matrix with these data sets
- Future NONMEM VI does not solve NM V issues but rather seems to be even more stringent than version V
- Datasets will be made available soon on a web site
- there are a number of buts and ifs when you make comparisons of this sort (Niclas)

Inserm

Acknowledgements for accepting to participate into this blind comparison (in alphabetical order)

- **Serge Guzy** (MCPEM, Xoma corp., USA)
- **Niclas Jonsson** (NONMEM FOCE V & VI. Univ. Uppsala, Sweden)
- **Marc Lavielle** (SAEM, Univ. Paris Sud Orsay, FR)
- **Bob Leary** (PEM; Pharsight, USA)
- **José Pinheiro and Chyi-Hung Hsu** (Splus nlme. Novartis, USA)
- **Hans Proost** (MW\Pharm and MultiFit. Univ. Groningen, NL)
- **Russ Wolfinger** (SAS PROC NLMIXED, USA)