History and new developments in estimation methods for nonlinear mixed-effects models*

Pr France Mentré

** My personal view with apologies for any omission*

INSERM U738

Department of Epidemiology, Biostatistics and Clinical Research
Bichat – Claude Bernard University Hospital,
University Paris 7, France

The beginning (70 - 80's)

- Nonlinear regression in pharmacokinetics
- Pharmacodynamic models

Holford & Sheiner (1981). Understanding the dose-effect relationship: clinical application of pharmacokinetic-pharmacodynamic models. *Clin Pharmacokinet*, 6:429-53.

- Linear mixed-effects models for cluster, correlated or longitudinal data…

Laird & Ware (1982). Random effects models for longitudinal data, *Biometrics*, 38:973-964

- EM algorithm for problem with missing data
  - Two steps algorithm
    - E-step: expectation of the log-likelihood of the complete data
    - M-step: maximisation of the log-likelihood of the complete data
  - Mixed-effects models: individual random-effects = missing data

Dempster, Laird & Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm, *JRSS B*, 1:1-38.

Lindstrom & Bates (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, *JASA*, 83:1014-22

Meanwhile at UCSF

NON linear Mixed Effects Model

- **1972**: The concept and the **FO method**

  Sheiner, Rosenberg & Melmon (1972). Modelling of individual pharmacokinetics for computer aided drug dosage. *Comput Biomed Res*, 5:441-59.

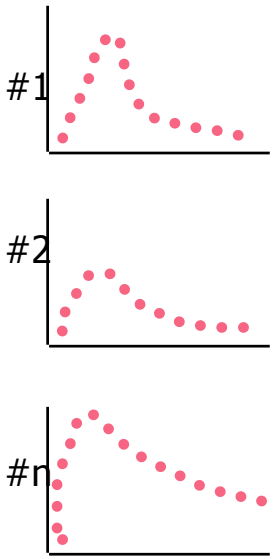
- **1977**: The first case study

  Sheiner, Rosenberg & Marathe (1977). Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokin Biopharm*, 5: 445-479.

- **1980**: NONMEM - An IBM-specific software

  Beal & Sheiner (1980). The NONMEM system. *American Statistician, 34:118-19*.
  Beal & Sheiner (1982). Estimating population kinetics. *Crit Rev Biomed Eng*, 8:195-222.

Standard Two-Stage approach

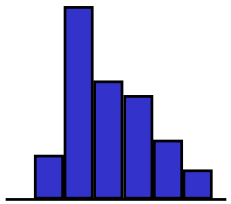


stage 1

Individual fitting Non Linear regression

Subject	Parameters estimate
#1	12.3
#2	21.9
#n	16.1

stage 2

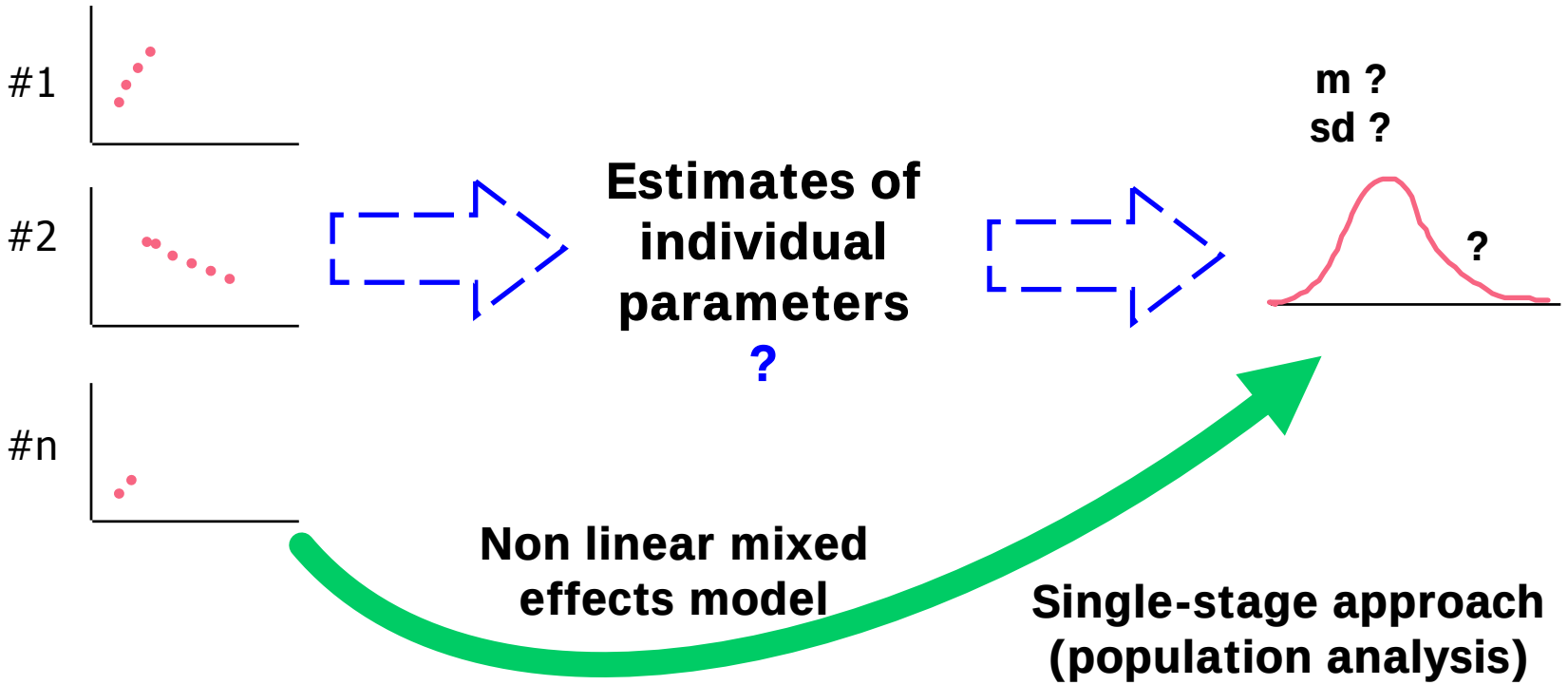


m , sd

Descriptive statistics, linear stepwise regression for covariate effect

From Steimer (1992): « Population models and methods, with emphasis on pharmacokinetics », in M. Rowland and L. Aarons (eds), *New strategies in drug development and clinical evaluation, the population approach*

Population approach



From Steimer (1992) : « Population models and methods, with emphasis on pharmacokinetics », in M. Rowland and L. Aarons (eds), *New strategies in drug development and clinical evaluation, the population approach*

Three seminal papers by Lewis Sheiner and Stuart Beal

1980: Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm*, 8:553-71.

1981: Evaluation of methods for estimating population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data. *J Pharmacokinet Biopharm*, 9:635-51.

1983: Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm*, 11:303-19.

The FO method (1)

- Estimation of population parameters by maximum likelihood
  - Find parameters that maximise the probability density function of the observations given the model
  - Good statistical properties of ML estimator

- Problem: No closed form of the likelihood
  - First order linearisation of the model around $\eta = 0$
  - Extended Least Square criterion

The FO method (2)

- Limitations
  - Assume that mean response = response for mean parameters -> *Not true for nonlinear models !!*
  - Bias if large inter-patient variability
  - Not real Maximum Likelihood Estimates (MLE)
    - good properties of MLE not always valid (LRT or Wald tests, standard errors from Fisher Information Matrix,...)
- Advantages
  - Better than STS in many cases
    - STS neglects estimation error : overstimation of inter-patient variability
      - Except for very rich design and small residual error
  - Takes into account correlation within subjects
    - better than all naive approaches (NAD, NPD)

More recent statistical developments in nonlinear mixed-effects models: three periods

1. 85 – 90: FOCE + other approaches: nonparametric, Bayesian

2. The 90's: new software, growing interest, new statistical developments, limitations of FOCE

3. Since 00: Stochastic methods for parametric ML estimation + …

85 to 90

- Linear mixed effects models in SAS (Proc MIXED in 1991)

- Rather limited interest of nonlinear mixed-effects in the statistical community

  Main statistical issue: nonlinearity with respect to the random effects
  - no close form to the likelihood
  - no easy implementation of the E-step of EM
  - FO approach: pseudo linear-random effects model

  => Three main extensions

85 to 90

1. Parametric and ML

- Method(s) based on first-order linearisation (FOCE)
  1. Estimation of individual random effects given current estimates (MAP)
  2. Linearisation of the model around the current estimates of the random effects
  3. Newton-Raphson iterative solution to a linear mixed-effects estimation problem

Lindstrom & Bates (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46:673-87.

- Attempt to use stochastic simulation to avoid linearisation

Mentré, Mallet & Steimer (1988). Hyperparameter estimation using stochastic aaproximation with application to population pharmacokinetics. *Biometrics*, 44:673-83.

85 to 90

2. Nonparametric and ML (NPML)

- No assumption on the distribution of the random effects (but parametric model !)

- Estimated distribution for the random effects = discrete distribution (any shape)
  - integrals -> sums
  - no approximation of the likelihood, no linearisation

Mallet (1986). A maximum likelihood estimation method for random coefficient regression models. *Biometrika,* 41:1015-23.

85 to 90

3. Bayesian parametric approaches

- Three levels of hierarchy: prior distribution on the population parameters

- Use of Monte Carlo Markov Chain (MCMC)
  - Inferences based on full posterior distributions
  - Evaluation using Gibbs sampler

Gelfand & Smith (1990). Sampling-based approaches to calculating marginal densities. *JASA*, 85:398-409.

Gelfand, Hills, Racine-Poon &Smith (1990). Illustration of bayesian inference in normal data models using Gibbs sampling. *JASA*, 972-975.

The 90's

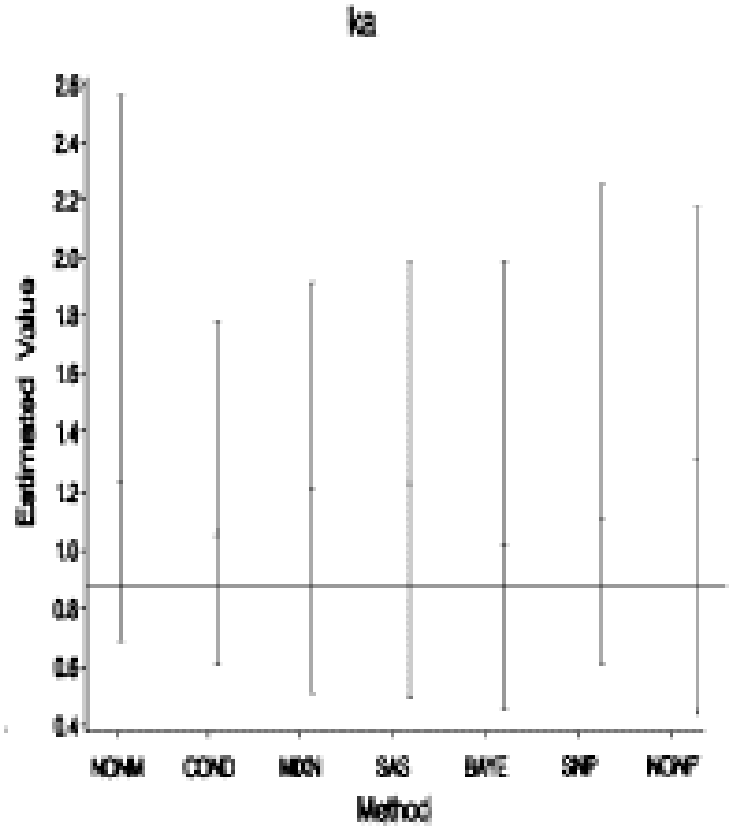
- Growing interest in the statistical field
  - Several developments "around" FOCE approach
  - New methods, new software
  - A Population Pharmacokinetic Modeling Workgroup formed by the Biopharmaceutical section of ASA (1991)
  - Review papers (Yuh, Beal, Davidian, Harrison, Hester, Kowalski, Vonesh & Wolfinger, *Biometrics,* 1994; Davidian & Giltinan, *J Biopharm Stat*, 1993)
  - Books

Davidian & Giltinan (1995). *Nonlinear models for repeated measurement data.* Chapman and Hall.

Vonesh & Chinchilli (1997). *Linear and nonlinear models for the analysis of repeated measurements.* Marcel Dekker.

Pinheiro & Bates (2000). *Mixed-effect models in S and Splus.* Springer Verlag.

Blind evaluation of several estimation algorithm on one simulated data set



Roe (1997). Comparison of population pharmacokinetic modeling methods using simulated data: results from the Population Modeling Workgroup. *Stat Med*, 16:1241-57. (meeting of the American Statistical Association in 1994)

The 90's

1. Parametric and ML

- New approximations of the likelihood
  - FOCE and GEE2

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

  - Laplacian (in NONMEM and SAS)

Wolfinger (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80:791-5.

Wolfinger (1997). An example of using mixed models and PROC MIXED for longitudinal data. *J Biopharm Stat*, 7:481-500.

  - Adaptive Gaussian quadrature (in SAS)

Pinheiro & Bates (1995). Approximations to the Log-Likelihood function in the nonlinear mixed-effects model. *J Comput Graph Stat*, 1:12-35.

See updated review by Davidan & Giltinan (2003), *J Agric Biol Enviro Stat*, 8:387-419

The 90's

1. Parametric and ML (cont'd)

- Several software with FOCE and/or these new approximations
  - NONMEM 4 in 1992 (Beal & Sheiner)
  - Macro MIXNLIN for SAS in 1995 (Vonesh & Carter)
  - Macro NLINMIX then Proc NLMIXED in SAS in 1999 (Wolfinger)
  - nlme in Splus as a macro in 1993 then in Splus in 1996 (Pinheiro & Bates)
  - WinNonMix in 1998 (Pharsight©)
  - MicroPharm–K in 1995 (Urien)
  - …

The 90's

1. Parametric and ML (cont'd)

- Avoding linearisation of the likelihood: EM- like
  - Solve the E-step using MAP (in P-PHARM in 1992 - ITBS)

Mentré & Gomeni (1995). A two-step algorithm for estimation on non-linear mixed-effects with an evaluation in population pharmacokinetics. *J Biopharm Stat*, 5:141-158.

Aarons (1993). The estimation of population pharmacokinetic parameters using an EM algorithm. *Comput Methods Programs Biomed*, 41:9-16.

  - EM with Monte Carlo integration for the E -step

Walker (1996). An EM algorithm for non-linear random effects models. *Biometrics*, 52:934-944.

The 90's

2. Non Parametric and ML

- NPEM: ideas of EM algorithm to estimate the discrete ditribution as in NPML (in USC*PACK)
  Schumitzky (1991). Nonparametric EM algorithm for estimating prior distributions. *Applied Math Comput*, 45:141-57.
- Smooth nonparametric: Mixture of normal distribution for the random effects
  Davidian & Gallant (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, 80:475-88.
- Splines for the random effects or for the longitudinal model
  Park, Verotta, Blaschke & Sheiner (1997). A semiparametric method for describing noisy population pharmacokinetic data. *J Pharmacokinet Biopharm,* 25:615-42

The 90's

3. Bayesian approaches

- Other development of MCMC using Metropling-Hasting (POPKAN software)

Smith & Wakefield (1994). The hierarchical Bayesian approach to population pharmacokinetic modelling. *Int J Biomed Comput,* 36:35-42.

Wakefield (1996). The Bayesian analysis of population pharmacokinetic models. *JASA*, 91:61-76.

The 90's

3. Bayesian approaches (cont'd)

- Development of BUGS (96) and PKBUGS (99)

Spiegelhalter, Thomas, Best & Gilks (1996). *BUGS 0.5: Bayesian Inference Using Gibbs Sampling - Manual*. MRC Biostatistics Unit, Cambridge.

Best, Tan, Gilks &Spiegelhalter (1995). Estimation of population pharmacokinetics using the Gibbs sampler. *J Pharmacokin Biopharm*, 23: 407-424.

Lunn & Aarons (1997). Markov chain Monte Carlo techniques for studying interoccasion and intersubject variability: application to pharmacokinetic data. *Applied Stat*, 46:73-91.

Lunn, Wakefield, Thomas, Best & Spiegelhalter (1999). *PKBugs User Guide*. Dept. Epidemiology & Public Health, Imperial College School of Medicine, London.

The 90's

3. Bayesian approaches (cont'd)

- Nonparametric methods (Dirichlet processes)

Wakefield & Walker (1997). Bayesian nonparametric population models: formulation and comparison with likelihood approaches. J *Pharmacokinet Biopharm*, 1997 25:235-53.

Rosner & Muller (1997). Bayesian population pharmacokinetic and pharmacodynamic analyses using mixture models. *J Pharmacokinet Biopharm,* 25:209-33.

Publications with these methods in PUBMED for the last 10 years

NONMEM	425
P-Pharm	28
NPEM	25
NPML	10
Proc NLMIXED	6
WinNonMix	5
nlme	3
PK-Bugs	1

Early 00's: limitations of FOCEs

1. Simulation studies

- Increased type I error of LRT and Wald test using NONMEM and/or nlme

Wahlby, Jonsson & Karlsson (2001). Assessment of actual significance levels for covariate effects in NONMEM. *J Pharmacokinet Pharmacodyn*, 28:231-52.

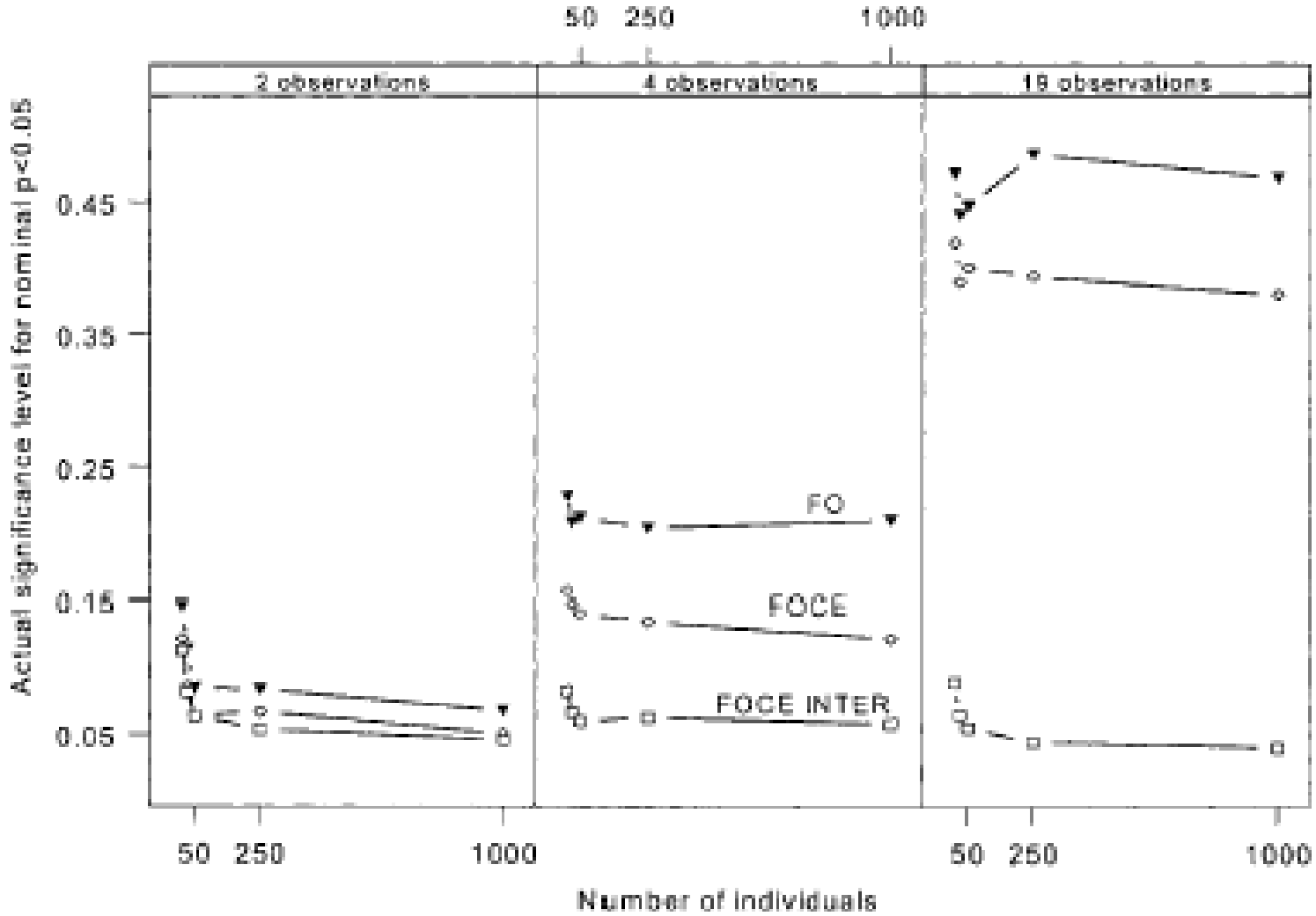
Wahlby, Bouw, Jonsson & Karlsson (2002). Assessment of type I error rates for the statistical sub-model in NONMEM. *J Pharmacokinet Pharmacodyn*, 29:251-69.

Comets E & Mentre F. (2001). Evaluation of tests based on individual versus population modeling to compare dissolution curves. *J Biopharm Stat*, 11:107-23

Ding & Wu (2001). Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics*., 2:13-29.

Type I errors of LRT with NONMEM

Wahlby, Jonsson & Karlsson, *J Pharmacokin Biopharm*, 2001



Early 00's: limitations of FOCEs

2. Estimator inconstistencies

- Inconsitency of the FOCE estimators when N increases with fixed n_i

  - need that both N and min(n_i) increases

Ramos & Pantula (1995). Estimation of nonlinear random coefficient models. *Statist Probab Letter,* 24: 49-56.

Vonesh (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models *Biometrika*, 83:447-52.

- Not that bad when sigma "small" compare to omega

Ko & Davidian (2000). Correcting for measurement error in individual-level covariates in nonlinear mixed effect models. *Biometrics*, 56:368-75.

2000 and after in parametric ML

Stochastic and MCMC methods in mixed effects model for maximum likelihood estimation! ☺

Gu & Kong (1998). A stochastic approximation algorithm with MCMC for incomplete data estimation problems. *PNAS*, 95: 7270-4.

Delyon, Laveille & Moulines (1999). Convergence of a stochastic approximation version of the EM procedure. *Ann Stat*, 27: 94-128.

Chen, Zhang & Davidian (2002). A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution.*Biostatistics*, 3:347-60.

Four new stochastic methods in parametric ML

Presented at the symposium *New algorithm developments in the field of non-linear mixed effect model,* Lyon, Sep. 2004

1. SPML: Didier Concordet (Toulouse, France)

Concordet & Nunez (2002). A simulated pseudo-maximum likelihood estimator for nonlinear mixed models. *Comput Stat Data Analysis*, 39: 187-201.

2. SAEM - MONOLIX: Marc Lavielle (Paris, France)

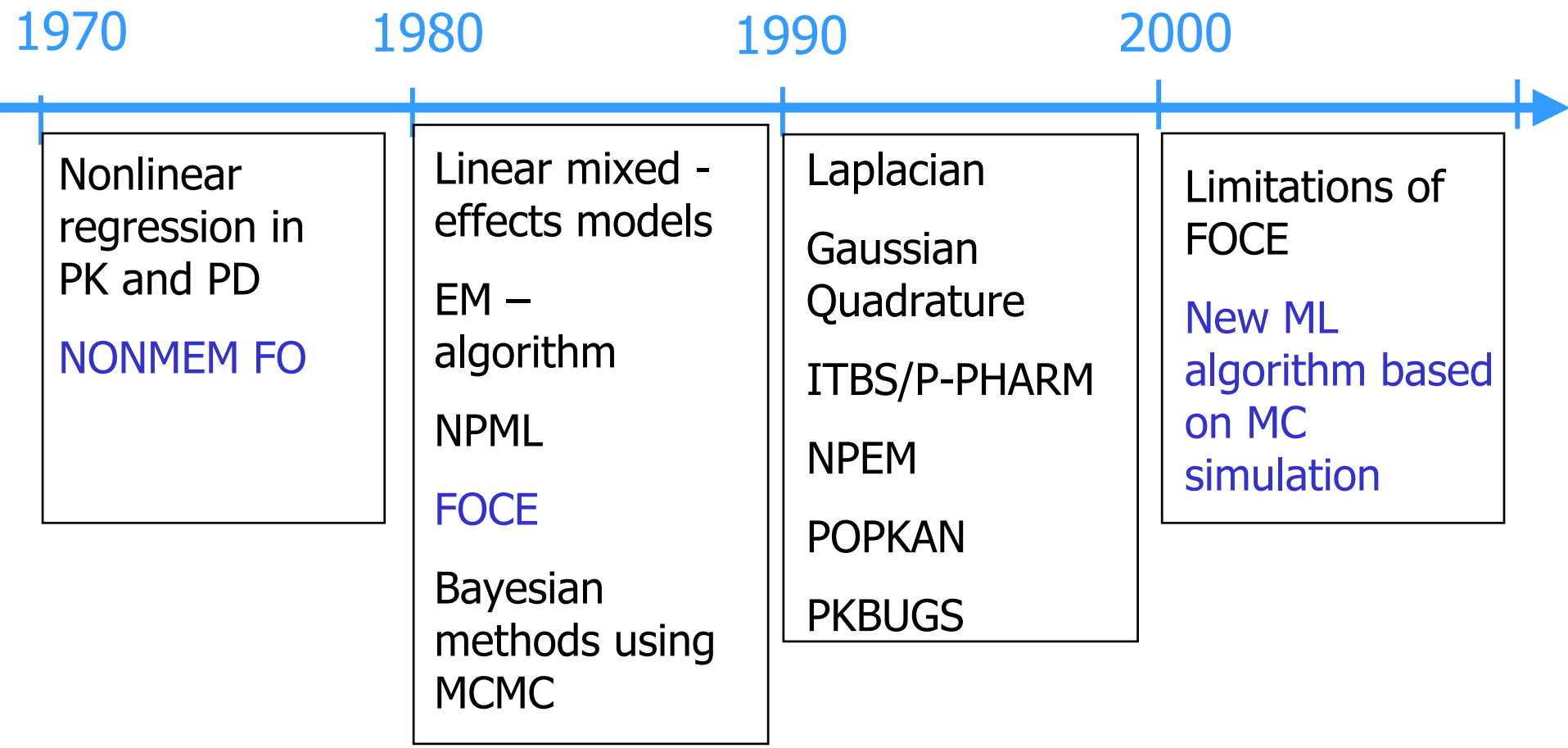
Kuhn E, Laveille M (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Analysis,* 49:1020-1038.

3. MCPEM: Serge Guzy (USA)

Bauer & Guzy (2004). Monte Carlo Parametric Expectation Maximization Method for Analyzing Population PK/PD Data. In: D'Argenio DZ, ed. *Advanced Methods of PK and PD Systems Analysis*. pp: 135-163.

4. PEM: Bob Leary (LA, USA)

Leary, Jelliffe, Schumitzky & Port (2004). Accurate Maximum Likelihood Estimation for Parametric Population Analysis. *PAGE*, 2004.


1970
1980
1990
2000
Nonlinear regression in PK and PD
NONMEM FO
Linear mixed - effects models
EM – algorithm
NPML
FOCE
Bayesian methods using MCMC
Laplacian
Gaussian Quadrature
ITBS/P-PHARM
NPEM
POPKAN
PKBUGS
Limitations of FOCE
New ML algorithm based on MC simulation

CONCLUSION

- Nonlinear mixed-effects models and ML increasingly used
- NONMEM used mostly in drug companies
- FOCE developed almost 15 years ago and have several drawbacks
- New ML methods based on stochastic simulation developed by statisticians

=> New software or new algorithm in current software needed

Blind comparison of several algorithms and/or software for parametric ML

P. Girard & F.Mentré, 2004-2005

1. PD example: 50 data sets – 7 methods
   - Results presented during the meeting in Lyon, France (Sep 2004)
2. PK example: 100 data sets - 10 methods

Procedure

1. Data sets sent blindly to all participants with given model and starting value for the fixed effects
2. Results received by Pascal Girard and compiled
3. Comparison of results sent blindly to each author
4. Sometimes second set of results sent by the participant

		PD	PK
FOCE- NONMEM V	N. Jonsson	☺	☺
FOCE- NONMEM VI	N. Jonsson		☺
FO – SAS NLMIXED	R. Wolfinger		☺
Adaptative Gaussian – SAS NLMIXED	A. Maloney R. Wolfinger	☺	☺
FOCE – Splus nlme	J. Pinheiro & C.H. Hsu	☺	☺
ITBS - MW\Pharm (Visual Basic)	H. Proost		☺
ITBS - MultiFit (Turbo Pascal)	H. Proost		☺
SPML	D. Concordet	☺	
SAEM – MONOLIX (Matlab)	M. Lavielle	☺	☺
MCPEM	S. Guzy	☺	☺
PEM	B. Leary	☺	☺