Group Comparison with Fused LASSO Penalized Likelihood: an Alternative to Test Based Methods

Edouard Ollier[1], Xavier Delavenne[1],
Thierry Basset[1], Adeline Samson[2]

(1) Groupe de Recherche sur la Thrombose, EA3065, Université Jean Monnet, Saint-Etienne, France
(2) Université Grenoble-Alpes, Laboratoire Jean Kuntzmann, UMR CNRS 5224, Grenoble, France

Group Comparison with Non Linear Mixed Effect Model

Non Linear mixed effects model (NLME) with group of subjects ($g = 1, 2$) :

$$y_{i,j}^g = f(t_{i,j}^g, \phi_i^g) + h(t_{i,j}^g, \phi_i^g)\epsilon_{i,j}^g \text{ with } \epsilon_{i,j}^g \sim \mathcal{N}(0, 1)$$
$$\phi_i^g = \mu^g + \omega_i^g \text{ with } \omega_i^g \sim \mathcal{N}(0, \Omega^g)$$
$$\Omega^g \text{ diagonal}$$
$$h = af + b$$

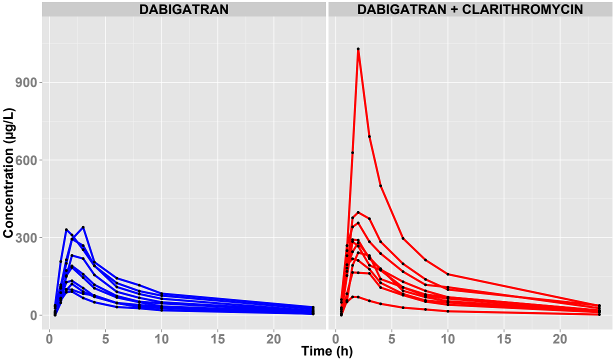
Hypothesis

$\boldsymbol{\mu^1 - \mu^2}$ **and** $\boldsymbol{\Omega^1 - \Omega^2}$ **are sparse**

Goal

Detect significant difference between group parameters

Dabigatran PK with or without Clarithromycin[1]



Is there a drug-drug interaction ?

Parameters influenced by clarithromycin ?

[1] Delavenne et al., 2013

Model Selection methods for NLME

- **Stepwise methods**:
  - Forward, Backward,...
  - Based on Likelihood Ratio Tests (LRT) or BIC

- **Greedy approach**:
  - All the possible model are evaluated

    $\hookrightarrow$Difficult with large number of parameter

- **Penalized likelihood**:
  - Preselect relevant models
  - example: the LASSO [2] (covariates selection)

[2] Bertrand J, Balding DJ, 2013

The Fused LASSO Penalty

$$\underset{\theta^1,\theta^2}{\text{ArgMin}} \sum_{g=1}^{2} -2LL(y^g, \theta^g) + \boldsymbol{\lambda} \|\theta^1 - \theta^2\|_1$$

Estimate a vector $\theta^1 - \theta^2$ that is sparse

λ tunes the level of sparsity :

- $\lambda = 0 \Rightarrow$ between group differences are estimated for each parameters
- high levels of λ values $\Rightarrow \theta^1 = \theta^2$

$\Rightarrow$ **We propose to use this penalty to preselect relevant models**

Fused LASSO for Non Linear Mixed Effect Models

Let θ^1 and θ^2 the following group parameters:

- $\theta^1 = (\mu^1, \Omega^1, a, b)$
- $\theta^2 = (\mu^2, \Omega^2, a, b)$

$$(\hat{\theta}^1, \hat{\theta}^2) = \underset{\theta^1, \theta^2}{\text{ArgMin}} \begin{cases} \sum_{g=1}^{2} -2LL(y^g; \theta^g) + \boldsymbol{\lambda_F} \|\boldsymbol{\mu^1} - \boldsymbol{\mu^2}\|_1 \\ \quad + \quad \boldsymbol{\lambda_V} \|\boldsymbol{\Omega^{1^{-1}}} - \boldsymbol{\Omega^{2^{-1}}}\|_1 \end{cases}$$

- $\boldsymbol{\lambda_F}$ tunes the level of sparsity of $\mu^1 - \mu^2$
- $\boldsymbol{\lambda_V}$ tunes the level of sparsity of $\Omega^1 - \Omega^2$

We modified the SAEM algorithm [3] [4] in order to solve this problem

[3] Kuhn E, Lavielle M, 2004
[4] Bertrand J, Balding DJ, 2013

The SAEM algorithm for Fused LASSO penalised likelihood

At iteration k :

- **E step**:
  1. **Simulation**:
     - simulation of ϕ^1 under $p(\phi|y^1;\theta_k^1)$ by MCMC
     - simulation of ϕ^2 under $p(\phi|y^2;\theta_k^2)$ by MCMC
  2. **Stochastic Approximation**: Stochastic approximation of $\mathbb{E}[logp(y^g,\phi^g,\theta^g)|y^g,\theta_k^g]$:
     - $Q_{k+1}^1(\theta) = \gamma_k log\ p(y^1,\phi^1,\theta_k^1) + (1-\gamma_k)Q_k(\theta)^1$
     - $Q_{k+1}^2(\theta) = \gamma_k log\ p(y^2,\phi^2,\theta_k^2) + (1-\gamma_k)Q_k(\theta)^2$

     with $(\gamma_k\ )$ sequence of decreasing step sizes

- **M step**: actualization of θ_k^1 and θ_k^2

$$(\theta_{k+1}^1,\theta_{k+1}^2) = \underset{\theta^1,\theta^2}{\mathrm{ArgMin}} \begin{cases} \sum_{g=1}^2 -Q_{k+1}^g(\theta^g) + \boldsymbol{\lambda_F}\|\boldsymbol{\mu^1 - \mu^2}\|_1 \\ \quad + \quad \boldsymbol{\lambda_V}\|\boldsymbol{\Omega^{1^{-1}} - \Omega^{2^{-1}}}\|_1 \end{cases}$$

M step in details

1. Fixed effects update[5]:

$$(\mu^1_{k+1}, \mu^2_{k+1}) \quad = \quad \underset{\mu^1, \mu^2}{\text{ArgMin}} \begin{cases} \sum_{g=1}^{2} -Q_k^g(\mu^g, \Omega_k^g, a_k, b_k) \\ \quad + \quad \boldsymbol{\lambda_F} \|\boldsymbol{\mu^1} - \boldsymbol{\mu^2}\|_1 \end{cases}$$

2. Random effects variance update[6]:

$$(\Omega^1_{k+1}, \Omega^2_{k+1}) = \underset{\Omega^1, \Omega^2}{\text{ArgMin}} \begin{cases} \sum_{g=1}^{2} -Q_k^g(\mu^g_{k+1}, \Omega^g, a_k, b_k) \\ \quad + \quad \boldsymbol{\lambda_V} \|\boldsymbol{\Omega^{1^{-1}}} - \boldsymbol{\Omega^{2^{-1}}}\|_1 \end{cases}$$

3. Error model's parameters update: same as in SAEM because they are not penalized.

[5] Boyd et al., 2011
[6] Danaher et al., 2011

BIC Selection of Optimal λ_F and λ_V Values

1. Preselect relevant models by solving the fused LASSO problem on a grid of λ_F and λ_V values:

$$(\ (\lambda_F^m, \lambda_V^m) \in \mathbb{R}^2, \ m \in \{1, ..., M\})$$

2. The optimal model (among the preselected) is the one that minimize the BIC [7]:

$$BIC(\lambda_F, \lambda_V) = \sum_{g=1}^{2} -2LL(y^g, \tilde{\theta}^g_{\lambda_F, \lambda_V}) + log(N) \times df_{\lambda_F, \lambda_V}$$

- N: total number of patients
- $\tilde{\theta}^g_{\lambda_F, \lambda_V}$: unpenalized reestimation of the selected model
- df_λ: number of distinct coefficients (degree of freedom)

[7] Delattre et al., 2014

Validation on Simulated Data Sets

- 50, 100 and 150 patients per group
- 30 simulated data sets for each group size
- Model : first order absorption, one compartment

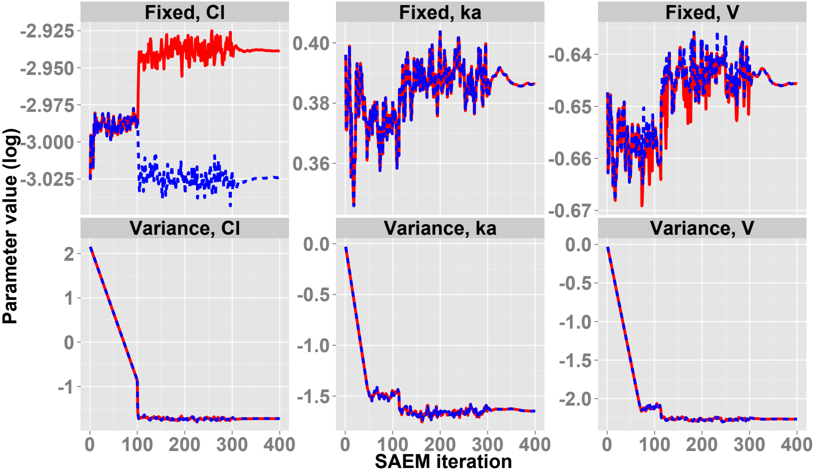
$$C(t) = \frac{D}{V} \frac{k_a}{k_a - Cl} (e^{-\frac{Cl}{V} t} - e^{-k_a t})$$

$$\mu^2_{Cl} = 0.7 \mu^1_{Cl} \quad \mu^2_{k_a} = \mu^1_{k_a} \quad \mu^2_V = 1.2 \mu^1_V$$

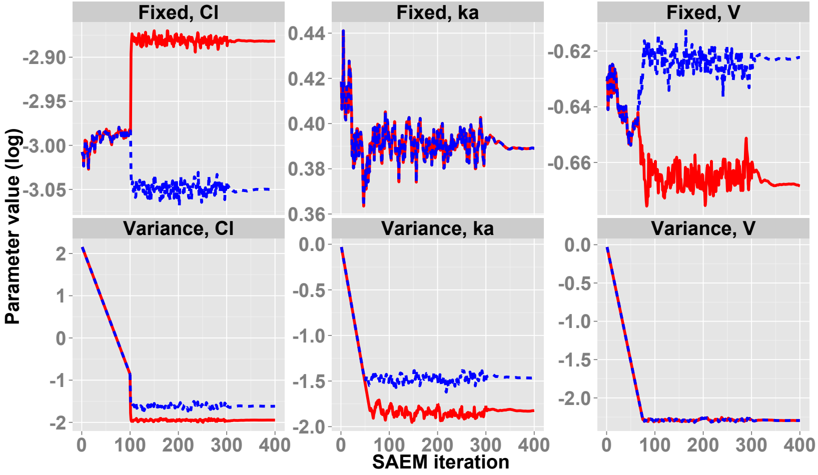
$$\omega^2_{Cl} = 2 \omega^1_{Cl} \quad \omega^2_{k_a} = 2 \omega^1_{k_a} \quad \omega^2_V = \omega^1_V$$

6 parameters
each could be equal or different in the 2 groups
$\Downarrow$
2^6 (= 64) possible models

SAEM Estimates: High λ_F and λ_V values



SAEM Estimates: Lower λ_F and λ_V values



BIC Selection Performance

- The fused LASSO preselect $\sim$ 18 models over the 64
- Percentage of correctly selected models:

	Subjects per group		
	50	100	150
Forward (LRT)	13%	33%	40%
Forward (BIC)	23%	36%	53%
Fused + BIC	33%	63%	80%

Dabigatran PK with or without Clarithromycin

Inverse Gaussian absorption model:

$$IG(t) = Dose \times F \times \sqrt{\frac{MAT}{2\pi CV^2 t^3}} \times e^{\frac{-(t-MAT)^2}{2CV^2 MATt}}$$

$$\frac{dD_c}{dt} = IG(t) - D_c(\frac{Q}{V_c} + \frac{Cl}{V_c}) + \frac{Q}{V_p} D_p$$

$$\frac{dD_p}{dt} = \frac{Q}{V_c} D_c - \frac{Q}{V_p} D_p$$

- Individual parameters: Log-Normal distribution
- Random effects on all the parameters

Dabigatran PK with or without Clarithromycin: Fused LASSO Analysis

First analysis[8]:

- Performed only on fixed effects
- Forward stepwise
- Significant difference only for μ_F

Fused LASSO analysis:

- Effect of clarithromycin tested on 8 parameters:
  - Absorption:

$$\mu_F, \mu_{MAT}, \mu_{CV} \text{ and } \omega_F, \omega_{MAT}, \omega_{CV}$$

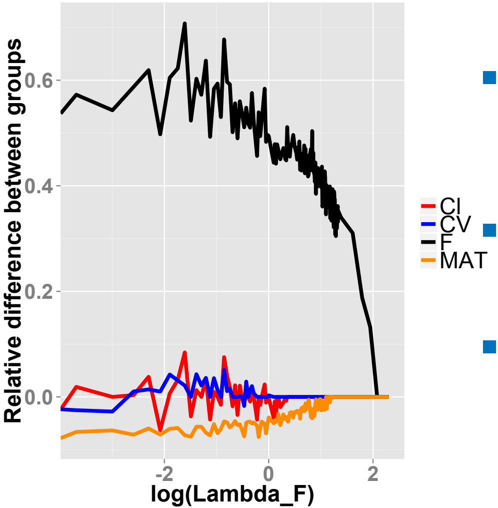
  - Elimination:

$$\mu_{Cl} \text{ and } \omega_{Cl}$$

$\Rightarrow$ $\mathbf{2^8 (= 256)}$ **possible models**

[8] Delavenne et al., 2013

Dabigatran PK with or without Clarithromycin: Results



- The fused LASSO preselect 77 models over the 256 possibilities
- Effect of Clarithromycin detected (BIC) on : μ_F, ω_F
- Possible effect on μ_{MAT} but not detected

Conclusion

- Extension of the SAEM algorithm for fused LASSO penalized likelihood on both fixed effects and random effects variances

- Fused LASSO penalty is well suited to do group comparison

- Not shown during the talk:
  - any number of groups (example: dose ranging study $\Rightarrow$ non linear PK)
  - covariates selection (example: genomic data)

- Perspectives:
  - Proof of convergence for the algorithm
  - Strategy for building the grid of λ_F and λ_V values
  - Take into account nested random effects (inter-occasion variability)