CSCanada
Progress in Applied Mathematics
Vol. 4, No. 2, 2012, pp. [148–166]
DOI: 10.3968/j.pam.1925252820120402.S0801

ISSN 1925-251X [Print]
ISSN 1925-2528 [Online]
www.cscanada.net
www.cscanada.org

Between-Subject and Within-Subject Model Mixtures for Classifying HIV Treatment Response

Cyprien Mbogning[a],[b],[c], Kevin Bleakley[a],[b] and Marc Lavielle[a],[b],*

[a] Inria Saclay, POPIX team.
[b] Laboratoire de Mathématiques d'Orsay, France.
[c] LIMSS, Ecole Nationale Supérieure Polytechnique, Yaoundé, Cameroun.

* Corresponding author.
Address: Inria Saclay, POPIX team; Laboratoire de Mathématiques d'Orsay, France; E-Mail: Marc.Lavielle@inria.fr

Received: August 17, 2012/ Accepted: October 8, 2012/ Published: October 31, 2012

Abstract: We present a method for using longitudinal data to classify individuals into clinically-relevant population subgroups. This is achieved by treating "subgroup" as a categorical covariate whose value is unknown for each individual, and predicting its value using mixtures of models that represent "typical" longitudinal data from each subgroup. Under a nonlinear mixed effects model framework, two types of model mixtures are presented, both of which have their advantages. Following illustrative simulations, longitudinal viral load data for HIV-positive patients is used to predict whether they are responding – completely, partially or not at all – to a new drug treatment.

Key words: Mixture models; HIV; SAEM; Classification

Mbogning, C., Bleakley, K., & Lavielle, M. (2012). Between-Subject and Within-Subject Model Mixtures for Classifying HIV Treatment Response. *Progress in Applied Mathematics, 4*(2), 148–166. Available from http://www.cscanada.net/index.php/pam/article/view/j.pam.1925252820120402.S0801 DOI: 10.3968/j.pam.1925252820120402.S0801

1. INTRODUCTION

For a variety of reasons – some known, some not – different patients respond differently to the same drug treatment. For certain patients, a drug does what it was

prescribed to do: kill bacteria, reduce blood pressure, decrease viral load, etc., but for others, the drug may be toxic or ineffective. When we collect response data on patients undergoing a treatment, it is useful to try to find patients for which the treatment is ineffective, and thus suggest modifications. We are particularly interested here in longitudinal response data in a population; the methods we present are generally applicable to this type of data.

The real-world example that motivates the approach is longitudinal HIV viral load data. For HIV-positive patients on a given drug regime, the evolution of the viral load in the blood can be measured over time. For some patients, the drug regime is ineffective and the viral load does not consistently drop; we call these *non-responders*, and it is of interest to detect them and provide alternative – hopefully more effective – treatments. Other patients react favourably to the treatment and the viral load drops to undetectable levels and stays there for a long period of time; these are called *responders*. Yet another group – *rebounders* – show an initial drop in viral load followed by an increase back towards the initial high viral load. They too will eventually require an alternative treatment.

Our goal is to use longitudinal data to infer the efficacy of the treatment, i.e., infer whether each patient is a non-responder, responder, rebounder or, as we will explain, some mixture of the above. In order to do this, we first model longitudinal HIV viral load data using recent additions to the nonlinear mixed-effects model (NLMEM) framework. Then, we extract relevant posterior probabilities or individual parameters to infer patient status. We now briefly introduce the NLMEM framework and model mixtures. To avoid potential confusion, note that *mixed-effects models* and *model mixtures* are not the same thing.

NLMEM – a special case of mixed-effects models – are statistical models which use both fixed and random effects in their construction (see [3,10,21,23] for more details). The model structure is hierarchical. At a first level, each individual has their own parametric regression model, known as the structural model, each identically defined up to a set of unknown individual parameters. At a second level, each set of individual parameters is assumed to be randomly drawn from some unknown population distribution. These models are particularly useful in population studies (e.g., population pharmacology – see [24]) where data is available for many individuals. Two types of variability are involved: *intra*- and *inter*-subject. We attempt to explain the latter using known *fixed-effects* (covariates) such as weight, blood type, etc. The non-explained part of the inter-subject variability is then modeled using *random effects*.

The introduction of a categorical covariate (e.g., sex, blood type, etc.) into such a model supposes that the population can be divided into subpopulations with respect to that covariate. However, there may be a categorical covariate which interests us but whose value is unknown for all individuals, such as the covariate "patient status" which interests us here. In this case, part of the goal becomes to infer the value of this covariate for each individual as part of the modeling process. One way to do this is to introduce *model mixtures*. There exist several types of model mixture which are useful in the context of mixed effects models; we will focus on two here:

- *Between-Subject Model Mixtures* (BSMM) assume that each individual's longitudinal data follows one of M "base" models, but we do not necessarily know *a priori* which one. Individual i thus has a label $z_i = m \in \{1, \ldots, M\}$

referring to the model that is supposed to have generated it. If the z_i are known, they can be treated as categorical covariates. We will show how to deal with the more challenging case of when they are unknown. Furthermore, for this z_i unknown case, we will show how to extract *a posteriori* estimates of the probability that each individual was generated by each of the base models; this will be used to predict which type of patient we have: non-responder, responder or rebounder. We note that BSMM were introduced as an example of a more general framework in [11] but were not developed further there.

- *Within-Subject Model Mixtures* (WSMM) make the hypothesis that the model mixture occurs *within each individual*. In the HIV example, this means that we consider that each patient is partially a non-responder, partially a responder and partially a rebounder. This is perhaps more biologically plausible than BSMMs in the sense that each individual's response may be due to their own particular combination of virus strains, cell populations, etc. Within the NLMEM framework, this means including individual "model proportion" parameters into the model and having to estimate them along with the other parameters of the NLMEM. It turns out that this does not require any mathematical extensions to a typical NLMEM. But as will be seen in the HIV example, we can use the estimated proportions to help categorize patients, especially those who do not naturally fall into one of the three "typical" categories.

BSMM and WSMM are new approaches in the context of NLMEM. We refer the reader to [1,12,13,15,16] for general details on mixture models in a standard context. We note also that there are fully Bayesian approaches to similar types of problems, in particular Bayesian nonparametric ones (see [7] for more details).

The paper is structured as follows. We introduce BSMMs in the NLMEM framework and calculate their log-likelihood, before briefly presenting WSMMs; they require no new mathematical framework. We then describe how to perform maximum likelihood estimation (MLE) for BSMMs using the Stochastic Approximation Expectation Maximization (SAEM) algorithm [4]. SAEM is implemented in the Monolix software and can be widely applied to various data types and real-life scenarios [2,6,9,18,22]. Next, we present an example for a simple BSMM case, and then a simulated example for mixtures of two models in the both the BSMM and WSMM cases. The simulations illustrate the quality of the parameter estimation of both methods and also their classification performance, i.e., how well they "predict" which model was used to generate each individual's data (in the BSMM case) and which model represented the biggest proportion (in the WSMM case). This is followed by a comprehensive modeling of HIV treatment response longitudinal data from a cohort of 578 patients using both BSMM and WSMM and a comparison of the quality and practical usefulness of each method. A discussion follows.

2. MODELS AND METHODS

2.1. Between-Subject Model Mixtures

Between-subject model mixtures (BSMMs) are a special case of NLMEMs that assume that the structural model is a mixture of M different structural models and

can be written, in the case of a continuous response, as:

$$y_{ij} = \sum_{m=1}^{M} \mathbb{1}_{\{z_i=m\}} \Big(f_m \left(x_{ij}; \psi_i\right) + g_m \left(x_{ij}; \psi_i, \xi\right) \varepsilon_{ij} \Big), \tag{1}$$

where

- $y_{ij} \in \mathbb{R}$ denotes the jth observation of the ith individual, $1 \leq i \leq N$ and $1 \leq j \leq n_i$.
- N is the number of individuals and n_i the number of observations of the ith individual.
- x_{ij} is a vector of regression variables (for longitudinal data, x_{ij} will generally be time t_{ij}).
- ψ_i is the d-vector of individual parameters of individual i. We assume that all the ψ_i are drawn from the same population distribution and are defined as Gaussian transformations:

  $$\psi_i = h(\mu, c_i, \eta_i), \tag{2}$$

  where h is a function which describes the covariate model, μ a vector of fixed-effects, c_i a vector of known covariates, $\eta_i \sim_{i.i.d} \mathcal{N}\left(0, \Sigma\right)$ a vector of random effects and Σ the inter-individual variance-covariance matrix.
- $z_i \in \{1, \ldots, M\}$ represents the (un)known group to which belongs the individual i. The proportion of individuals in group m is given by $\pi_m = \mathbb{P}\left(z_i = m\right)$ with $\sum_{m=1}^{M} \pi_m = 1$.
- $\varepsilon_{ij} \sim \mathcal{N}\left(0, 1\right)$ are the residual errors, and are independent of individual parameters ψ_i.
- f_m for $m = 1, \ldots, M$ are functions defining structural models in each group.
- g_m for $m = 1, \ldots, M$ are functions defining the (possibly heteroscedastic) residual error model. We will consider here error models of the form $g_m = a + b\, f_m$.

Furthermore, let $\theta = (\mu, \Sigma, \xi, \pi_1, \ldots, \pi_M)$ represent the complete set of population parameters.

BSMMs are particularly relevant in the domain of population pharmacology if we are aiming to distinguish between different classes of response to the same treatment. For example, if we can effectively model each class of longitudinal response data (specific mathematical function with parameters, etc.), *a posteriori* estimation of the label z_i for each individual leads to being able to assign each to a given "typical" response. We shall see later in the real HIV data that classes such as "non-responder", "responder", and "rebounder" could be used to categorize individuals' responses in this application.

2.2. Log-Likelihood of Between-Subject Model Mixtures

In this section, we briefly recall the initial exposition and notation in [11] for the log-likelihood of mixture models in general; it will be useful in the following. The complete data is noted (y, ψ, z) with y the observed data and (ψ, z) the unobserved data. For subject i, the log-likelihood of the complete data is

$$\mathcal{L}(y_i, \psi_i, z_i; \theta) = \sum_{m=1}^{M} \mathbb{1}_{z_i=m} \left(\mathcal{L}_m(y_i, \psi_i; \theta) + \log \mathbb{P}(z_i = m) \right), \tag{3}$$

where $\mathcal{L}_m(y_i, \psi_i; \theta)$, the log-likelihood of pairs of variables (y_i, ψ_i) in group $G_m := \{i, 1 \leq i \leq N \text{ such that } z_i = m\}$, is given by $\mathcal{L}_m(y_i, \psi_i; \theta) = \mathcal{L}_{Y,m}(y_i|\psi_i; \xi) + \mathcal{L}_\psi(\psi_i; \mu, \Sigma)$. The right-hand side terms are simple to calculate. $\mathcal{L}_m$ is assumed to belong to the exponential family, i.e., there exists a function ψ of θ and a minimal sufficient statistic $T(y_i, \psi_i)$ such that $\mathcal{L}_m(y_i, \psi_i; \theta) = \langle T(y_i, \psi_i), \theta \rangle - \psi(\theta)$.

In what follows, we will note $\mathbb{P}(z_i = m)$ as π_m or π_{im} for "proportion" for respectively BSMM or WSMM. We have that

$$\mathcal{L}(y_i, \psi_i, z_i; \theta) = \sum_{m=1}^{M} \mathbb{1}_{z_i=m} \left(\langle T(y_i, \psi_i), \theta \rangle + \log \pi_m - \psi(\theta) \right).$$

The likelihood of the complete data also belongs to the exponential family as it can be written $\mathcal{L}(y, \psi, z; \theta) = \langle S(y, \psi, z), \theta \rangle - \psi(\theta)$, where the mth row of S is given by

$$\left(\sum_{i=1}^{n} \mathbb{1}_{z_i=m} \ , \ \sum_{i=1}^{n} \mathbb{1}_{z_i=m} T(y_i, \psi_i) \right).$$

We will show later that this representation of the log-likelihood is helpful for implementing stochastic EM-like algorithms for the BSMM case.

2.3. Within-Subject Model Mixtures

It may be too simplistic to assume that each individual is represented by only one well-defined model from the mixture. For instance, in a pharmacological setting there may be subpopulations of cells, viruses (etc.) *within each patient* that react differently to a drug treatment. In this case, it makes sense to consider that the mixture of models happens *within* each individual. Such within-subject model mixtures (WSMMs) therefore require additional vectors of individual parameters $\pi_i = (\pi_{i1}, \ldots \pi_{iM})$ representing proportions of the M models within each individual i. These vectors are supposed independent of ψ, and naturally sum to 1 for each individual. Using the same notation as Section 2.1, observations are modeled by:

$$y_{ij} = \sum_{m=1}^{M} \pi_{im} \left(f_m(x_{ij}; \psi_i) + g_m(x_{ij}; \psi_i, \xi) \varepsilon_{ij} \right). \tag{4}$$

Since there are no latent categorical covariates, WSMMs actually fall under the framework of classical NLMEMs. Thus, no further specific methodology needs to be developed, and we refer to [4,8] for the standard treatment.

3. MAXIMUM LIKELIHOOD ESTIMATION ALGORITHMS FOR BETWEEN-SUBJECT MODEL MIXTURES

A method such as the Stochastic Approximation EM (SAEM) algorithm [4] needs to be used to replace the E-step of the EM algorithm in the NLMEM framework. The stochastic step of SAEM is performed in practice using an MCMC procedure. SAEM has been successfully applied to a wide number of data types and real-life situations including bioequivalence crossover trials [6], estimation of population pharmacokinetic-pharmacodynamic viral dynamics parameters [2], longitudinal ordered categorical data [20], population models for count data [19] and group comparison tests in longitudinal data analysis [18].

We now develop this technique in the case of BSMMs. Iteration k consists of a number of MCMC iterations with $p(\psi, z|y; \theta^{(k)})$ as the stationary distribution. More precisely, the Gibbs algorithm is combined with the Metropolis-Hastings algorithm, with various proposal kernels. Here, the N subjects are assumed to be independent and the same procedure is used for each of the N subjects.

For subject i, draw $z_i^{(k)} \in \{1, \ldots, M\}$ from the multinomial distribution

$$\mathcal{M}_M \left(\frac{\pi_m^{(k-1)} p_m\left(y_i, \psi_i^{(k-1)}; \theta^{(k-1)}\right)}{\sum_{r=1}^{M} \pi_r^{(k-1)} p_r\left(y_i, \psi_i^{(k-1)}; \theta^{(k-1)}\right)} \right)_{m=1,\ldots,M} .$$

A first possible kernel uses the marginal distribution $p(\psi_i; c_i, \theta^{(k)})$ for generating a candidate ψ_i^c; more precisely, $\eta_i \sim \mathcal{N}\left(0, \Sigma^{(k)}\right)$ and ψ_i^c is as in (2). The probability of acceptance, i.e., the probability to move from ψ_i to ψ_i^c, becomes

$$\alpha\left(\psi_i, \psi_i^c\right) = \min\left(1, \frac{p\left(y|\psi_i^c, z_i^{(k)}; \theta^{(k)}\right)}{p\left(y|\psi_i, z_i^{(k)}; \theta^{(k)}\right)}\right).$$

A random walk can also be used as a possible kernel: $\psi_i^c \sim \mathcal{N}\left(\psi^{(k-1)}, \Omega\right)$. The diagonal matrix Ω can be adaptively adjusted to get a chosen acceptance rate. Setting different elements of the diagonal of Ω to 0 during iterations can be done to use different directions. The probability of acceptance is

$$\alpha\left(\psi_i, \psi_i^c\right) = \min\left(1, \frac{p\left(y, \psi_i^c, z_i^{(k)}; \theta^{(k)}\right)}{p\left(y, \psi_i, z_i^{(k)}; \theta^{(k)}\right)}\right).$$

For details on how to choose parameters such as the number of iterations of the MCMC procedure during the simulation step, the step-size sequence (δ_k), etc., we refer the reader to [11].

We remark that this version of SAEM for BSMMs is now implemented in the MONOLIX software.

BSMM Example. In order to illustrate MLE for the BSMM model, let us consider a simple example. Suppose first that we have a constant residual model

$g_m(x_{ij};\psi_i,\xi) = a$. Then, the conditional log-likelihood of the observations in the group G_m is:

$$\mathcal{L}_{Y,m}(y_i|\psi_i;\xi) = -\frac{1}{2a^2}\sum_{j=1}^{n_i}\left(y_{ij} - f_m(x_{ij},\psi_i)\right)^2 - n_i\log(a) - \frac{n_i}{2}\log(2\pi).$$

Furthermore, assuming a Gaussian distribution without covariates (i.e., $\psi_i = \mu + \eta_i$) for individual parameters, the likelihood of the individual parameters is

$$\mathcal{L}_{\psi}(\psi_i;\mu,\Sigma) = -\frac{1}{2}(\psi_i-\mu)'\Sigma^{-1}(\psi_i-\mu) - \frac{d}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma|).$$

In the first step of iteration k, we must approximate the following minimal sufficient statistics:

$$\begin{aligned}
s_{k,1,m} &= s_{k-1,1,m} + \delta_k\left(\sum_{i=1}^{N}\mathbb{1}_{z_i^{(k)}=m} - s_{k-1,1,m}\right)\\
s_{k,2} &= s_{k-1,2} + \delta_k\left(\sum_{i=1}^{N}\psi_i^{(k)} - s_{k-1,2}\right)\\
s_{k,3} &= s_{k-1,3} + \delta_k\left(\sum_{i=1}^{N}\psi_i^{(k)}\psi_i^{(k)'} - s_{k-1,3}\right)\\
s_{k,4} &= s_{k-1,4} + \delta_k\left(\sum_{i,j,m}\mathbb{1}_{z_i^{(k)}=m}\left(y_{ij} - f_m\left(x_{ij},\psi_i^{(k)}\right)\right)^2 - s_{k-1,4}\right).
\end{aligned}$$

Then, in the M-step, we update parameters according to:

$$\begin{aligned}
\pi_m^{(k)} &= \frac{s_{k,1,m}}{N}\\
\mu^{(k)} &= \frac{s_{k,2}}{N}\\
\Sigma^{(k)} &= \frac{s_{k,3}}{N} - \left(\frac{s_{k,2}}{N}\right)\left(\frac{s_{k,2}}{N}\right)'\\
a^{(k)} &= \sqrt{\frac{s_{k,4}}{\sum_{i=1}^{N} n_i}}.
\end{aligned}$$

3.1. Estimation of Individual Parameters

The overall goal of this paper is not to provide complicated viral dynamics models (ordinary differential equations, etc.) but rather to show how the model mixture framework can be used to predict classes of individuals. For our real-life application, this means being able to decide whether HIV patients are non-responders, responders, rebounders or some mixture of the above. For BSMM, the latent categorical covariate z contains the unknown "class" labels that need to be estimated. For a given set of population parameters θ, we can use each individual's conditional distribution $p(z_i,\psi_i|y_i,\theta)$ to estimate the latent variable z_i and the vector of individual parameters ψ_i. A first estimate is the Maximum a Posteriori (MAP) which is

obtained by maximizing this joint conditional distribution with respect to (z_i, ψ_i):

$$\left(\hat{z}_i, \hat{\psi}_i\right) = \arg \max_{(z_i,\psi_i)} p\left(z_i, \psi_i | y_i, \theta\right).$$

This maximization is not straightforward and we refer the reader to [11] for a complete methodology. Another way to estimate the latent covariate z_i is to maximize the marginal conditional distribution:

$$\hat{z}_i = \arg \max_m \mathbb{P}\left(z_i = m | y_i; \theta\right). \tag{5}$$

The value of (5) can be estimated using a stochastic approximation during the SAEM iterations.

As for WSMM, since there are no latent categorical covariates z, estimation of individual proportions π_i as well as individual parameters ψ_i is straightforwardly obtained by maximizing the joint conditional distribution $p\left(\pi_i, \psi_i | y_i, \theta\right)$ with respect to (π_i, ψ_i):

$$\left(\hat{\pi}_i, \hat{\psi}_i\right) = \arg \max_{(\pi_i,\psi_i)} p\left(\pi_i, \psi_i | y_i, \theta\right).$$

Once we have estimates of individual parameters, individual predictions for BSMM are obtained using $\hat{y}_{ij} = f_{\hat{z}_i}\left(x_{ij}, \hat{\psi}_i\right)$. Similarly, for WSMM, the individual predictions are calculated as

$$\hat{y}_{ij} = \sum_{m=1}^{M} \hat{\pi}_{im} f_m\left(x_{ij}, \hat{\psi}_i\right).$$

4. SIMULATED DATA EXAMPLE

4.1. Modeling with Between-Subject Model Mixtures

We performed a simulation study to evaluate the performance of the proposed BSMM algorithm for estimating parameters and classifying individuals using a mixture of two simple models defined as follows:

$$\begin{array}{ll} \text{if } z_i = 1, & y_{ij} = f_1\left(\psi_i, t_{ij}\right) + a\varepsilon_{ij} \\ \text{if } z_i = 2, & y_{ij} = f_2\left(\psi_i, t_{ij}\right) + a\varepsilon_{ij}, \end{array}$$

where

$$f_1\left(\psi_i, t_{ij}\right) = A_i\,, \qquad f_2\left(\psi_i, t_{ij}\right) = A_i e^{-L_i t_{ij}}\,,$$

and the vectors of individual parameters $\psi_i = (\log(A_i), \log(L_i))$ are such that A_i and L_i are log-normally distributed, $\log(A_i) \sim \mathcal{N}\left(\log(A), \sigma_A^2\right)$ and $\log(L_i) \sim \mathcal{N}\left(\log(L), \sigma_L^2\right)$. Note that this model can also be seen as a parameter mixture for L_i: $L_i = 0$ in group 1 and L_i log-normally distributed in group 2. For the experiments, we set $A = 10$, $L = 0.2$, $\sigma_A^2 = 0.5$ and $\sigma_L^2 = 0.5$. Furthermore, we fixed $\mathbb{P}\left(z_i = 1\right) = \pi_1 = 1/3$ and $a = 1$. t_{ij} is the jth measurement time for subject i, and we used the same set of times $t = 0, 1, \ldots, 8$ for all N subjects. $K = 1000$ datasets were simulated and the parameters were estimated using the proposed algorithm. Figure 1 shows examples of longitudinal data for 10 subjects with (a) $z_i = 1$ and (b) $z_i = 2$.

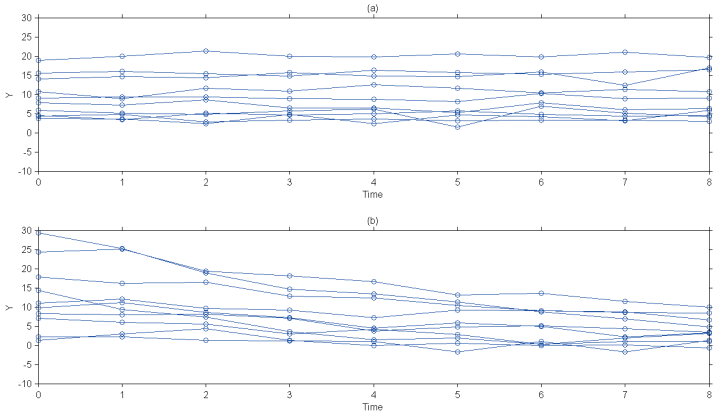

Figure 1
Spaghetti Plots of Data for Ten Individuals Belonging to Group 1 (a) and 10 Others Belonging to Group 2 (b).

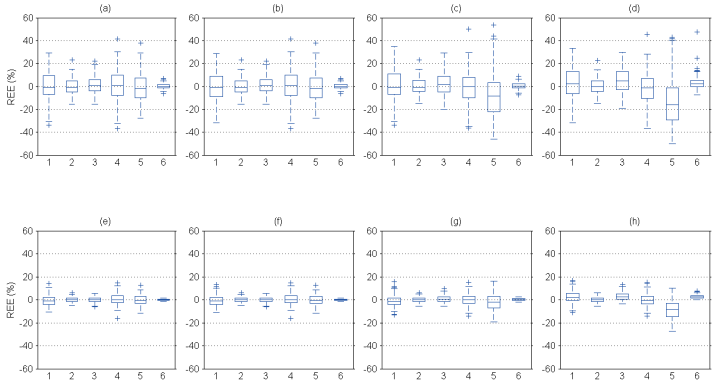

Figure 2
Empirical Distribution of the Relative Estimation Error (REE_k (%)) in the BSMM Scenario with Different Sample Sizes: (a-d) $N = 100$, (e-h) $N = 1000$ and Various Cases: (a)&(e) z and ψ Are Known, (b)&(f) z Is Unknown and ψ Is Known, (c)&(g) z Is Known and ψ Is Unknown, (d)&(h) z and ψ Are Unknown. The Estimated Parameters Are 1: π_1; 2: A; 3: L; 4: σ_A^2; 5: σ_L^2; 6: a.

Let $\theta^\star$ be the true value of any parameter and $\hat{\theta}_k$ the estimated value obtained with the kth simulated dataset. The relative estimation error (in %) REE_k was used as a quality criteria:

$$\text{REE}_k = \frac{\hat{\theta}_k - \theta^\star}{\theta^\star} \times 100.$$

Figure 2 shows the distribution of the REE_k for each parameter when (a-d) $N = 100$ and (e-h) $N = 1000$ with various experimental situations, i.e., when ψ and/or z are supposed known/unknown.

It suggests that most parameters are estimated with little or no bias, except perhaps the variance of the individual parameter L_i. Cases (a-b) and (e-f) are quite similar, suggesting that the EM algorithm is efficient with respect to bias (mixture parameters are estimated in cases (b-f) with the EM algorithm). Furthermore, cases (c-d) and (g-h) are quite similar and we see that there is little degradation compared with (a-b) and (e-f) respectively. Thus, the SAEM algorithm for mixtures appears to be efficient with respect to bias. As expected, we see that parameters are better estimated with $N = 1000$, but even with $N = 100$ the results are acceptable.

Quantitative results are presented in Table 1, which gives means as well as standard errors for each of the estimated parameters when $N = 100$ and $N = 1000$.

Table 1
Mean of Parameter Estimates and Standard Errors for the BSMM Scenario with $N = 100$ or $N = 1000$.

		N=100		N=1000	
θ	θ^*	Mean of estimates	SE of $\hat{\theta}$	Mean of estimates	SE of $\hat{\theta}$
π_1	0.33	0.334	0.049	0.331	0.015
A	10	10.03	0.713	10.01	0.226
L	0.20	0.202	0.02	0.20	0.006
σ_A^2	0.50	0.497	0.075	0.50	0.023
σ_L^2	0.50	0.49	0.099	0.50	0.033
a	1	1.004	0.026	1.002	0.008

The parameter estimates, overall, match the population values, and the standard errors seems reasonable and are quite low, especially with large sample size. In addition, Figure 3, representing the relative difference between the estimated standard error and the empirical standard error, shows that the standard errors are well estimated and match quite well the empirical ones.

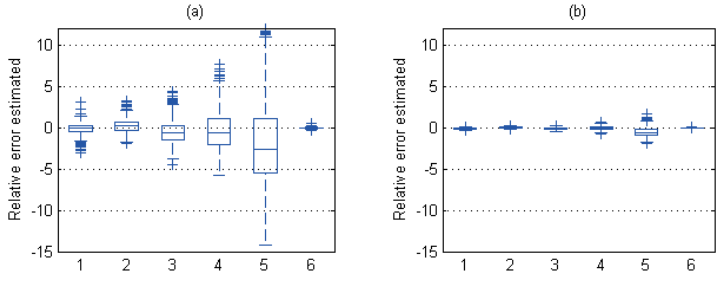

Figure 3
Relative Difference (in %) Between Estimated and Empirical Standard Errors for the BSMM Scenario with (a): $N = 100$, (b): $N = 1000$. The Estimated Parameters Are 1: π_1; 2: A; 3: L; 4: σ_A^2; 5: σ_L^2; 6: a.

Figure 4 provides a graphical illustration of the probability of correct classification in both groups for $N = 100$ and $N = 1000$ subjects. Note that the number of individuals in each group is considered as fixed here ($N_1 = N\pi_1$) during the Monte Carlo simulation. For each of the $K = 1000$ runs, the probabilities of correct classification for the N subjects were computed and ranked in increasing order. Then, the empirical *median* sequence of these 1000 sequences was computed in each group. This median is represented by the solid line in Figure 4. These graphs are more informative than the distribution of the number of subject misclassified over the simulations. Indeed, we see for instance that, with $N = 100$, less than 3 (resp. 4) subjects among 33 (resp. 67) of group 1 (resp. group 2) have a probability smaller than 0.8 to be correctly classified in half of the cases.

As expected, the probability of correct classification is greater when (ψ_i) is known, but it is interesting to note that the difference is relatively small.

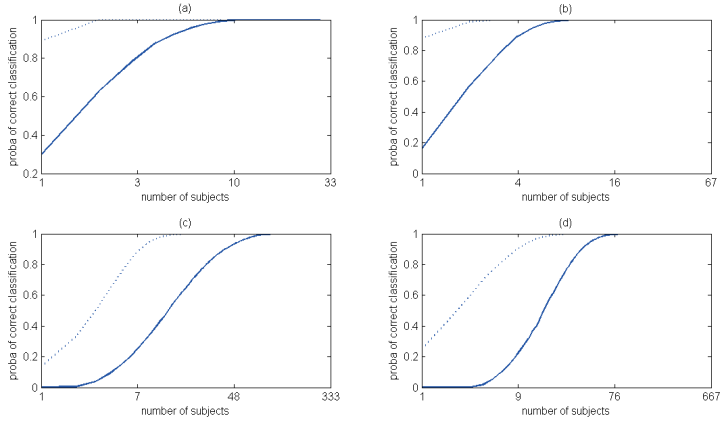

Figure 4
Medians of the Probabilities of Correct Classification Ranked in Increasing Order in Both Groups with (a & b): $N = 100$ ($N_1 = 33$, $N_2 = 67$) and (c & d): $N = 1000$ ($N_1 = 333$, $N_2 = 667$) Subjects ; (a & c): Group 1 and (b & d): Group 2. Solid Line: the Individual Parameters (ψ_i) Are Unknown; Dotted Line: the Individual Parameters (ψ_i) Are Known.

4.2. Modeling with Within-Subject Mixture Models

We used the same model but now assumed that individual proportions π_{i1} were model parameters:

$$f(\psi_i, t_{ij}) = \pi_{i1} f_1(\psi_i, t_{ij}) + (1 - \pi_{i1}) f_2(\psi_i, t_{ij}),$$

where the structural models f_1 and f_2 were the same as for those for the BSMM, and individual proportions modeled as:

$$\pi_{i1} = \frac{1}{1 + s\, e^{\eta_i}}, \qquad \eta_i \sim \mathcal{N}\left(0, \sigma_s^2\right),$$

with $s = 2$ and $\sigma_s^2 = 0.2$. Population parameters were fixed at $A = 10, L = 0.2, \sigma_A^2 = 0.5$ and $\sigma_L^2 = 0.5$. We used the same measurement times for the N subjects as in the BSMM scenario.

Figure 5 shows the distribution of the REE_k for each parameter (a & b) and the relative difference between the estimated standard error and the empirical standard error (c & d), when (a & c): $N = 100$ and (b & d): $N = 1000$ subjects.

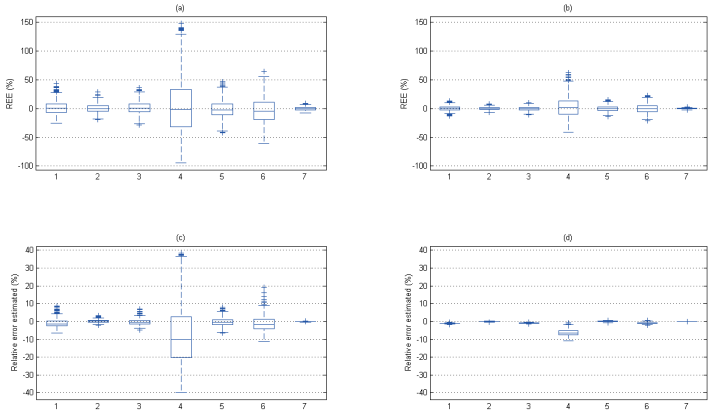

Figure 5
Empirical Distribution of the Relative Estimation Error (REE_k (%)) in the WSMM Scenario (a & b) and the Relative Difference Between the Estimated Standard Error and the Empirical Standard Error (c & d) with Two Sample Sizes: (a & c): $N = 100$, and (b & d): $N = 1000$. The Estimated Parameters Are 1: s; 2: A; 3: L; 4: σ_s^2; 5: σ_A^2; 6: σ_L^2; 7: a.

We see in (a) and (b) that parameters are estimated with very little bias, and with high precision when the sample size increases to $N = 1000$. In (c) and (d), we see that the standard errors are generally well estimated and close to the empirical ones when the sample size increases to $N = 1000$.

Quantitative results are presented in Table 2, which gives means as well as the standard errors for each of the estimated parameters when $N = 100$ and $N = 1000$.

The parameter estimates, overall, match the population values and the standard errors seem reasonable and are quite small with large sample size.

5. APPLICATION TO REAL DATA

5.1. Description of the Data

The randomized, controlled and partially blinded POWER project conducted by TIBOTEC comprises 3 studies performed in highly treatment-experienced HIV-infected patients using Darunavir/Ritonavir or an investigator-selected control protease inhibitor, combined with an optimized background regimen of nucleotide re-

Table 2
Mean of Parameter Estimates and Standard Errors for the WSMM Scenario with $N = 100$ or $N = 1000$.

		N=100		N=1000	
θ	θ^*	Mean of estimates	SE of $\hat{\theta}$	Mean of estimates	SE of $\hat{\theta}$
s	2	2.02	0.2183	2.00	0.053
A	10	10.04	0.714	10.00	0.226
L	0.20	0.202	0.021	0.199	0.006
σ_s^2	0.20	0.212	0.11	0.205	0.022
σ_A^2	0.50	0.495	0.073	0.50	0.023
σ_L^2	0.50	0.481	0.112	0.50	0.032
a	1	1.00	0.027	1.00	0.008

verse transcriptase inhibitors with or without the fusion inhibitor enfuvirtide. The output data is the viral load evolution for 578 patients. Figure 6 gives examples of patients with one of three "characteristic" viral load progressions:

- *Non-responders* (1) show no decline in viral load.
- *Responders* (2) exhibit a sustained viral load decline.
- *Rebounders* (3 and 4) exhibit an initial drop in viral load, then a rebound to higher viral load levels.

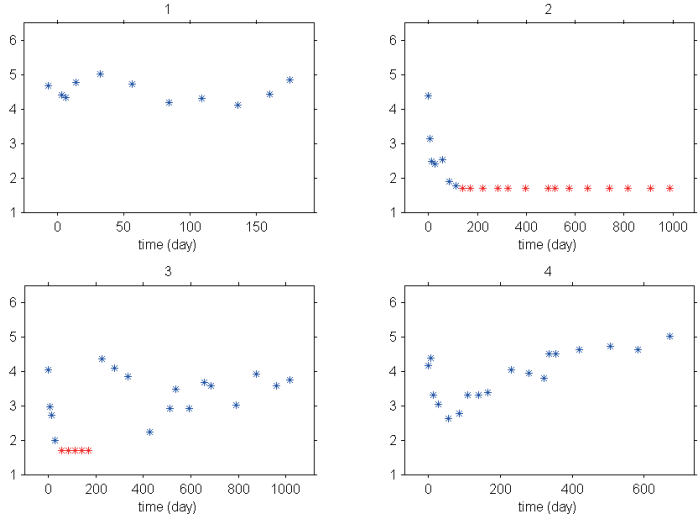

Figure 6
Viral Load Progression for 4 HIV-Infected Patients. (1) Non-Responder; (2) Responder; (3) and (4) Are Rebounders. Red Points Indicate Below Level of Quantification Data.

Remark. There is a detection limit at 50 HIV RNA copies/ml, corresponding to a log-viral load of 1.7; i.e., data are left censored. These points are shown in red in Figure 6. Censoring is taken into account in the following analysis as described in [17].

5.2. Class Prediction Using Between-Subject Model Mixtures

Within a few months of HIV infection, patients typically enter a steady state of chronic infection and have a stabilized concentration of HIV-1 in blood plasma. When the anti-retroviral treatment starts, the viral load of patients who respond shows an initial rapid exponential decay, usually followed by a slower second phase of exponential decay, see [14]. It is shown in [5] that the biphasic decay in viral load can be approximated by a bi-exponential model $A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}$.

After the decrease in viral load levels, some subjects show a rebound, which can be due to several factors (non-adherence to the therapy, emergence of drug-resistant virus strains, etc.). We propose to extend the bi-exponential model to these patients by adding a third phase described by a logistic growth process $A_3/(1 + e^{-\lambda_3(t-\tau)})$, where τ is the inflection point of this growth process.

We then propose to describe the log-transformed viral load with a BSMM with three simple models, corresponding to each of three characteristic viral load progressions:

- Non-responder-like data can be described using a simple horizontal line. The structural model is given by:

$$z_i = 1, \quad f_1(\psi_i, t_{ij}) = A_{1i} + A_{2i}\,.$$

- As described above, the drop in viral load in responder-like data can be described using a bi-exponential mixed-effects model:

$$z_i = 2, \quad f_2(\psi_i, t_{ij}) = A_{1i} e^{-\lambda_{1i} t_{ij}} + A_{2i} e^{-\lambda_{2i} t_{ij}},$$

  where λ_{1i} and λ_{2i} describe the rate of exponential decay and A_{1i} and A_{2i} are intercept parameters for individual i. As in [3], these parameters are considered to be strictly positive.

- Rebounder-like data show a rebound after a biphasic decrease in viral load levels:

$$z_i = 3, \quad f_3(\psi_i, t_{ij}) = A_{1i} e^{-\lambda_{1i} t_{ij}} + A_{2i} e^{-\lambda_{2i} t_{ij}} + \frac{A_{3i}}{1 + e^{-\lambda_{3i}(t_{ij} - \tau_i)}}\,.$$

The log-transform viral load is then modeled by:

$$\log(y_{ij}) = \sum_{m=1}^{3} \mathbb{1}_{z_i = m} \log\left(f_m(\psi_i, t_{ij})\right) + \varepsilon_{ij}, \tag{6}$$

where y_{ij} is the viral load for subject i at time t_{ij} and $\psi_i = (A_{1i}, A_{2i}, A_{3i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}, \tau_i)$ the vector of individual parameters.

These parameters are positive and distributed according to log-normal distributions. Thus,

$$\log A_{li} = \log A_l + \eta_{li}, \qquad \eta_{li} \sim \mathcal{N}\left(0, \sigma^2_{A_l}\right), \qquad l = 1, 2, 3$$

$$\begin{aligned} \log \lambda_{mi} &= \log \lambda_m + \eta_{(m+3)\,i}, \qquad \eta_{(m+3)\,i} \sim \mathcal{N}\left(0, \sigma^2_{\lambda_m}\right), \qquad m = 1, 2, 3 \\ \log \tau_i &= \log \tau + \eta_{7i}, \qquad \eta_{7i} \sim \mathcal{N}\left(0, \sigma^2_{\tau}\right). \end{aligned}$$

By setting $\pi_1 = \mathbb{P}(z_i = 1)$, $\pi_2 = \mathbb{P}(z_i = 2)$ and $\pi_3 = \mathbb{P}(z_i = 3)$ with $\sum_{m=1}^{3} \pi_m = 1$, the complete set of population parameters to be estimated is given by $\theta = (A_1,\ A_2,\ A_3,\ \lambda_1,\ \lambda_2,\ \lambda_3,\ \tau,\ \pi_1,\ \pi_2,\ \pi_3)$.

Parameter estimation was performed using the SAEM algorithm for BSMM implemented in Monolix. This algorithm combined with a previous version [17] (which is an extension of the SAEM algorithm [8] to left censored data) takes properly into account the censored viral load data below the limit of quantification. Figure 7 shows the individual fits for the 4 patients, the vector of estimated posterior probabilities $\hat{p}_i$ where $p_{im} = \mathbb{P}\left(z_i = m | y_i; \hat{\theta}\right)$, i.e., the probabilities for each subject to belong to each of the three classes, and the class z_i to which they are assigned (1 = non-responder, 2 = responder, 3 = rebounder) corresponding to the maximum of the estimated posterior probabilities. We see that in all four cases, there is little ambiguity in the results, i.e., the correct class has a posterior probability very close to 1.

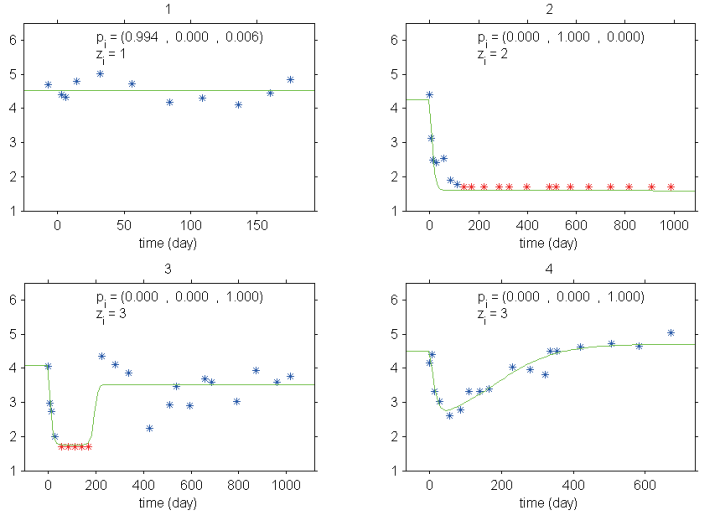

Figure 7
Viral Load Time Series and Individual Fits for the Four Patients in Figure 6. z_i is the Predicted Class of the Patient (1 = Non-Responder, 2 = Responder, 3 = Rebounder) and p_i the Posterior Probability that the Patient Is in Classes 1 to 3; the Index of Its Maximum Component Is Used to Predict z_i.

5.3. Class Prediction Using Within-Subject Mixture Models

Not all observed viral load progressions fall so easily into one of the three classes, as for example the patients shown in Figure 8.

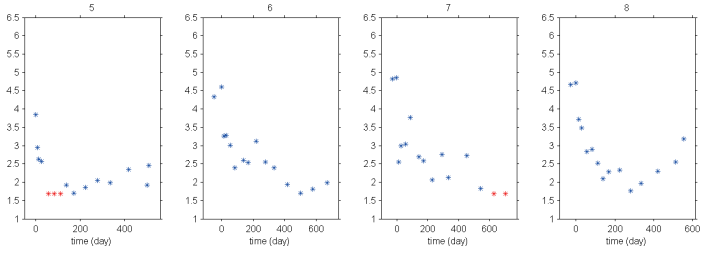

Figure 8
Viral Load Data for 4 Patients with Ambiguous Progressions. Red Points Indicate Below Level of Quantification Data.

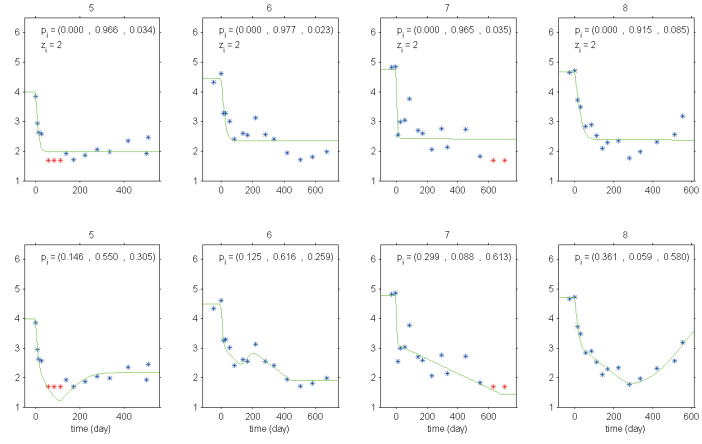

Figure 9
Viral Load Time Series and Individual Fits for Patients 5–8 from Figure 8 When Using BSMM (row 1) and WSMM (row 2). For BSMM, z_i Is the Predicted Class of the Patient (1 = Non-Responder, 2 = Responder, 3 = Rebounder) and p_i the Posterior Probability that the Patient Is in Classes 1 to 3. For WSMM, p_i Are the Mixture Parameters of the Classes Estimated as Part of the Model.

In these cases, it does not seem quite so reasonable to model the data under the BSMM assumption that each patient must belong uniquely to one class; instead, it is perhaps more natural to suppose that each patient is partially responding, partially non-responding and partially rebounding with respect to the given drug treatment. The goal becomes to find the relative strength of each process in each patient, and a WSMM is an ideal tool to do this. The proportions π_i are now individual parameters in the model and the problem is transformed into a standard NLMEM. Since these proportions are assumed to be positive and summing to 1 for each patient, we assumed a logit-normal distribution on the $\pi_{im}, m = 1, 2, 3$ as

follows:

$$\pi_{i1} = \frac{\gamma_{1i}}{1+\gamma_{1i}+\gamma_{2i}} \qquad \pi_{i2} = \frac{\gamma_{2i}}{1+\gamma_{1i}+\gamma_{2i}} \qquad \pi_{i3} = \frac{1}{1+\gamma_{1i}+\gamma_{2i}} \ ,$$

with

$$\log \gamma_{ki} = \log \gamma_k + \eta_{ki}, \qquad \eta_{ki} \sim \mathcal{N}\left(0, \sigma^2_{\gamma_k}\right) \quad \text{for } k = 1, 2.$$

We performed parameter estimation for WSMM using the SAEM algorithm in MONOLIX, taking properly into account the censored viral load data below the limit of quantification as before. Results for the four patients are presented in Figure 9; the first row gives the results which would be obtained using BSMM for these four patients, the second row the WSMM results.

We see that for all four patients, BSMM predicts that the patient was a responder. The visually poor individual fits in all four cases give rise to suspicion in the validity of the result. In particular, in plot 8 – top row – what appears to be a "late" rebounder is hardly "seen" by the algorithm; the posterior probability that the patient is a rebounder is only 0.085. Clearly, forcing each patient to belong to one class in the interior of the algorithm is a disadvantage of the method for real-world data.

In contrast to this, the first thing that is immediately obvious with the WSMM modeling (Figure 9 – bottom row) is that the individual fits are significantly better than the BSMM ones. This can be confirmed using the BIC criteria which clearly selects the WSMM model: BIC(WSMM)=14668, whereas BIC(BSMM)=15029. The estimated parameters p_i are generally consistent with what we "see" in the graphs, and when they are not, it helps us to look closer.

For instance, in plots 5 and 6 – bottom row – we see evidence of responding and a hint of rebound; this is confirmed in the estimated parameters p_i. In plot 7 – bottom row – the patient would seem to be a very slowly-but-surely responder. However, modeling happens population-wise, and this patient's reponder curve is far from the "normal" (steep drop followed by flatlining) responder; consequently the responder proportion in the mixture is small (0.088). Another reason may be that the 4th, 5th and 6th data points indicate a steep rebound from the 3rd data point (even though subsequent points drop) possibly influencing the weight of the rebounder model in the mixture (0.613). This is probably also accentuated by the fact that the rebounder model "includes" the responder model's biexponential decay terms. This remark also goes some way to explaining plot 8 – bottom row; the rebounder mixture coefficient (0.580) dominates and the responder coefficient (0.059) is unintuitively small. Again this is probably because the steep drop that we associate with the responder model is already included as the first two terms in the rebounder model; hence the small responder coefficient.

6. DISCUSSION

We have presented a classification methodology to interpret longitudinal data in a population context using model mixtures in the NLMEM framework. Two classes of model mixtures were introduced: between-subject and within-subject (BSMM and WSMM), and it was shown how to perform maximum likelihood estimation in the BSMM case. These algorithms are now available in the software MONOLIX. In simulations with mixtures of two models, we saw generally good parameter estimation performance and prediction performance, i.e., prediction of which model

had been used to generate each individual's longitudinal data. In real longitudinal HIV viral load data, we found that BSMMs were very efficient both in modeling and predicting the type of patient (non-responder, responder, rebounder) whenever the patient's viral load evolution was a "model case" of one of the three classes. However, when the viral load evolution was not so "typical", the fact that BSMMs force the allocation of once class to each patient lead to poor individual model fits and what's more, dubious class prediction.

On the other hand, WSMMs allow for more flexibility in modeling and are consistent with the biologically plausible hypothesis that each individual may be only partially responding/non-responding/rebounding due to their own unique virus strains, cell populations, etc. Allowing a mixture of biologically relevant models *internally* to each patient clearly improved the individual model fits and permitted a more nuanced reply to the classification question of whether the patient was responding adequately to the treatment or not. Based on these results, we would have more confidence in suggesting a modification to HIV drug regimes based on WSMM modeling than BSMM.

ACKNOWLEDGMENTS

We wish to thank Janssen Research & Development, a division of Janssen Pharmaceutica NV., for providing the POWER data.

REFERENCES

[1] Bryant, P.G. (1991). Large sample results for optimization based clustering methods. *Journal of Classiffication, 8*, 31–44.

[2] Chan, P., Jacqmin, P., Lavielle, M., McFadyen, L., & Weatherley, B. (2011). The use of the SAEM algorithm in MONOLIX software for estimation of population pharmacokinetic-pharmacodynamic-viral dynamics parameters of maraviroc in asymptomatic HIV subjects. *Journal of Pharmacokinetics and Pharmacodynamics , 38*, 41–61.

[3] Davidian, M., & Giltinan, D.M. (1995). *Nonlinear models for repeated measurements data*. London: Chapman & Hall.

[4] Delyon, B., Lavielle, M., & Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics, 27*, 94–128.

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

[6] Dubois, A., Lavielle, M., Gsteiger, S., Pigeolet, E., & Mentré, F. (2011). Model-based analyses of bioequivalence crossover trials using the SAEM algorithm. *Statistics in Medicine, 30*, 582–600.

[7] Hjort, N.L., Holmes, C., Mller, P., & Walker, S.G. (Eds.) (2010). *Bayesian nonparametrics*. Cambridge: Cambridge University Press.

[8] Kuhn, E., & Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Statist. Data Anal., 49*, 1020–1038.

[9] Lavielle, M., Samson, A., Fermin, A.K., & Mentré, F. (2011). Maximum likelihood estimation of long term HIV dynamic models and antiviral response. *Biometrics, 67*, 250–259.

[10] Lindstrom, M.J., & Bates, D.M. (1990). Nonlinear mixed-effects models for repeated measures. *Biometrics, 46*, 673–687.

[11] Mbogning, C., & Lavielle, M. (n.d.). Inference in mixtures of non linear mixed effects models. Submitted.

[12] McLachland, G.J., & Basford, K.E. (1988). *Mixture models: inference and applications to clustering*. New York: Marcel Dekker Inc.

[13] McLachland, G.J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley-Interscience.

[14] Perelson, A., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., ..., Ho, D. (1997). Decay characteristics of HIV−1 infected compartments during combination therapy. *Nature, 387*, 188–191.

[15] Redner, R.A., & Walker, H.F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review, 26*, 195-239.

[16] Roeder, K., & Wasserman, L. (1997). Practical bayesian density estimation using mixtures of normals. *Journal of the American Statistician Association, 92*, 894–902.

[17] Samson, A., Lavielle, M., & Mentré, F. (2006). Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: application to HIV dynamics model. *Comput. Statist. Data Anal., 51*, 1562–1574.

[18] Samson, A., Lavielle, M., & Mentré, F. (2007). The SAEM algorithm for group comparison tests in longitudinal data analysis based on nonlinear mixed-effects model. *Statistics in Medicine, 26*, 4860-4875.

[19] Savic, R., & Lavielle, M. (2009). A new SAEM algorithm: performance in population models for count data. *Journal of Pharmacokinetics and Pharmacodynamics, 36*, 367–379.

[20] Savic, R., Mentré, F., & Lavielle, M. (2011). Implementation and evaluation of an SAEM algorithm for longitudinal ordered categorical data with an illustration in pharmacometrics. *The AAPS Journal, 13*(1), 44–53.

[21] Sheiner, L.B., & Beal, S.L. (1985). Pharmacokinetic parameter estimates from several least squares procedures: superiority of extended least squares. *J. Pharmacokinet. Biop., 13*, 185–201.

[22] Snoeck, E., Chan, P., Lavielle, M., Jacqmin, P., Jonsson, N., Jorga, K., ..., Frey, N. (2010). Hepatitis C viral dynamics explaining breakthrough, relapse or response after chronic treatment. *Clinical Pharmacology and Therapeutics, 87*(6), 706–713.

[23] Vonesh, E.G., & Chinchilli, V.M. (1997). *Linear and nonlinear models for the analysis of repeated measurments*. New York: Marcel Dekker.

[24] Wakefield, J.C., Aaron, L., & Racine-Poon, A. (1998). The bayesian approach to population pharmacokinetic/pharmacodynamic modelling. *In case studies in bayesian statistics*, In B.P. Carlin *et al.* (Eds.). New York: Springer.