



# Clustering of trajectories with mixed effects classification model: Inference taking into account classification uncertainties

Charlotte Dugourd, Amna Abichou-Klich, René Ecochard, Fabien Subtil

## ► To cite this version:

Charlotte Dugourd, Amna Abichou-Klich, René Ecochard, Fabien Subtil. Clustering of trajectories with mixed effects classification model: Inference taking into account classification uncertainties. *Statistics in Medicine*, 2023, 42 (25), pp.4570-4581. 10.1002/sim.9876 . hal-04384305

**HAL Id: hal-04384305**

**<https://hal.science/hal-04384305>**

Submitted on 22 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## RESEARCH ARTICLE

# Clustering of trajectories with mixed effects classification model: Inference taking into account classification uncertainties

Charlotte Dugourd<sup>1,2,3,4</sup> | Amna Abichou-Klich<sup>1,2,3,4</sup>  | René Ecochard<sup>1,2,3,4</sup> | Fabien Subtil<sup>1,2,3,4</sup>

<sup>1</sup>Service de Biostatistique, Hospices Civils de Lyon, Lyon, France

<sup>2</sup>Université de Lyon, Lyon, France

<sup>3</sup>Université Lyon 1, Villeurbanne, France

<sup>4</sup>CNRS, Laboratoire de Biométrie et Biologie Évolutive UMR 5558, Villeurbanne, France

## Correspondence

Amna Abichou-Klich, Service de Biostatistique-Bioinformatique, Hospices Civils de Lyon, 162, avenue Lacassagne - F-69003, Lyon, France.  
Email: [amna.klich01@chu-lyon.fr](mailto:amna.klich01@chu-lyon.fr)

Classifying patient biomarker trajectories into groups has become frequent in clinical research. Mixed effects classification models can be used to model the heterogeneity of longitudinal data. The estimated parameters of typical trajectories and the partition can be provided by the classification version of the expectation maximization algorithm, named CEM. However, the variance of the parameter estimates obtained underestimates the true variance because classification uncertainties are not taken into account. This article takes into account these uncertainties by using the stochastic EM algorithm (SEM), a stochastic version of the CEM algorithm, after convergence of the CEM algorithm. The simulations showed correct coverage probabilities of the 95% confidence intervals (close to 95% except for scenarios with high bias in typical trajectories). The method was applied on a trial, called low-cyclo, that compared the effects of low vs standard cyclosporine A doses on creatinine levels after cardiac transplantation. It identified groups of patients for whom low-dose cyclosporine may be relevant, but with high uncertainty on the dose-effect estimate.

## KEYWORDS

classification, confidence interval, longitudinal data, mixed effects model, SEM-CEM algorithms

## 1 | INTRODUCTION

The aim of unsupervised classification is to identify homogenous groups among a heterogeneous population without any prior knowledge on these groups. In medicine, it can involve a biomarker with repeated measurements (biomarker trajectory) and in this case the purpose is to gather biomarker trajectories into a small number of groups of similar trajectories. The classification provided is useful for exploring and understanding the heterogeneity of biomarker trajectories among patients.

Two different approaches can be applied to model heterogeneous data: the mixture model and the classification model. For both approaches, the principle is to suppose that there exists  $K$  latent typical trajectories underlying the individual trajectories. The mixture model for longitudinal data<sup>1,2</sup> considers that each individual trajectory is a mixture of these typical trajectories weighted by the posterior probabilities of belonging to them. After estimation of the parameters—the parameters of the typical trajectories and the prior probabilities of belonging to a group—a posterior classification may be

applied according to the maximum posterior probability. The initial objective of the mixture model is not to classify but to appropriately model trajectories, allowing heterogeneity; hence the typical trajectories may not necessarily correspond to patients, and can thus be difficult to interpret. An alternative is to use a classification model.<sup>3,4</sup> In the classification model, the population is supposed to be divided in  $K$  homogeneous subpopulations, and each patient belongs to a single group; the predicted trajectory of a given patient is the typical trajectory of the group to which he/she belongs. The groups have real meaning, because each group is associated with a set of patients. The parameters, which are jointly estimated, are the parameters of the typical trajectories, the prior probabilities to belong to each group, and the partition of patients into groups. This article focuses on the classification model. In order to take into account the between-individual variability within groups, it is possible to extend the classification model by incorporating random effects around the typical trajectories.<sup>5</sup>

This latter model, named mixed effects classification model, allows inference on the parameters through a classification algorithm that deals with the maximization of the classification likelihood. As the maximization is complex, it is performed iteratively with an alternation between a classification phase and a maximization phase, using an initial classification. The estimators of the parameters are assumed to be asymptotically distributed according to a multivariate normal distribution centered on the true values of the parameters and with a variance-covariance matrix equal to the inverse of the Fisher information matrix obtained maximization. This however, does not take into account the uncertainties of classification. Although the classification performed is assumed to be correct during the estimation of the parameters, there is an uncertainty regarding the obtained classification; the standard errors are underestimated and the inference is not valid.

This article presents a method that aims at taking into account classification uncertainties in the variance of the parameter estimates of mixed effects classification models. The objective is to be able to make correct inference, to provide the confidence intervals of the parameters of typical trajectories, and perform a statistical test. This article first introduces the model and the algorithm used. Simulations were then performed to evaluate the method. The model is finally applied to a clinical trial comparing the effects of a low vs standard dose of cyclosporine A on creatinine levels in the post-cardiac-transplant period. The estimate of the low-dose effect is provided with its confidence interval along with its test result.

## 2 | THE MIXED EFFECTS CLASSIFICATION MODEL

### 2.1 | The model

Let  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT_i})$  be the response vector of the  $T_i$  successive measurements of individual  $i$ , with  $i = 1, \dots, N$ , and  $Y_{it}$  correspond to the  $t$ th measurement of individual  $i$ , with  $t = 1, \dots, T_i$ . Let  $z_{ik} = 1$  if the  $i$ th individual belongs to the  $k$ th group and 0 otherwise, with  $k = 1, \dots, K$ . The corresponding partition is denoted by  $\mathbf{P}$ , and  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$  is the vector of prior probabilities of belonging to each group ( $0 \leq \pi_k \leq 1$  for  $k = 1, \dots, K$ , and  $\sum_{k=1}^K \pi_k = 1$ ).

If the  $i$ th individual trajectory belongs to the  $k$ th group, the mixed effects classification model may be written as:

$$\mathbf{Y}_i = \mathbf{X}_{i1}\boldsymbol{\gamma} + \mathbf{X}_{i2}\boldsymbol{\beta}_k + \mathbf{Z}_i\mathbf{b}_{ik} + \boldsymbol{\epsilon}_i,$$

with  $\mathbf{b}_{ik} \sim N(0, \mathbf{D}_k)$  and  $\boldsymbol{\epsilon}_i \sim N(0, \sigma^2 \mathbf{I}_{T_i})$ .

$\mathbf{X}_{i1}$  is the design matrix with the  $p_1$  fixed effect covariates associated with the vector of coefficients  $\boldsymbol{\gamma}$ , identical across groups.  $\mathbf{X}_{i2}$  is the design matrix including the time effect, with the  $p_2$  group-dependent fixed effect covariates associated with the vector of coefficients  $\boldsymbol{\beta}_k$ . Depending on the model used, there may be no parameter with identical effects across groups, but only different ones.  $\mathbf{Z}_i$  is the design matrix with  $q$  covariates—which can include the time effect—and their effects are assumed to be random across individuals. The coefficient vectors of random effects,  $\mathbf{b}_{ik}$ , are independent and identically distributed according to a normal distribution with mean zero and variance-covariance matrix  $\mathbf{D}_k$  that may depend on the group or be identical across groups.

Conditional on group membership, the marginal distribution is:

$$f(\mathbf{y}_i, \boldsymbol{\gamma}, \boldsymbol{\beta}_k, \mathbf{D}_k, \boldsymbol{\pi}, \sigma^2) \sim N(\mathbf{X}_{i1}\boldsymbol{\gamma} + \mathbf{X}_{i2}\boldsymbol{\beta}_k, \mathbf{Z}_i\mathbf{D}_k\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{T_i}).$$

This distribution allows calculation of the classification likelihood.<sup>6</sup> The parameters to estimate are  $\boldsymbol{\gamma}, \boldsymbol{\beta}_k, \mathbf{D}_k, \boldsymbol{\pi}, \sigma^2$ , and the partition  $\mathbf{P}$ .

## 2.2 | The algorithm

Simultaneously maximizing the classification likelihood according to the parameters and the partition is difficult, so the CEM algorithm is applied. The CEM algorithm<sup>7</sup> is the classification version of the expectation maximization (EM)<sup>8,9</sup> algorithm. It is used to estimate the partition and the parameters of the model. A classification step is incorporated between steps E and M of the EM algorithm. Starting from an initial partition, the  $m$ th iteration is defined as follows:

- **E-step:** Estimation of the posterior probability of belonging to group  $k$  for each individual  $i$ , according to the parameters obtained in the previous iteration:

$$\text{post.prob}_{ik}^{(m+1)} = \frac{\pi_k f(\mathbf{y}_i, \boldsymbol{\beta}_k^{(m)}, \mathbf{D}_k^{(m)}, \sigma_k^{2(m)})}{\sum_{k=1}^K \pi_k f(\mathbf{y}_i, \boldsymbol{\beta}_k^{(m)}, \mathbf{D}_k^{(m)}, \sigma_k^{2(m)})}.$$

- **C-step:** Assignment of each trajectory  $i$  to the group providing the maximum  $\text{post.prob}_{ik}^{(m+1)}$  ( $1 \leq k \leq K$ ), leading to  $z_{ik}^{(m+1)}$  and the partition  $\mathbf{P}^{(m+1)}$ .
- **M-step:** Maximization of log classification likelihood conditional on  $\mathbf{P}^{(m+1)}$ , and re-estimation of the parameters of typical trajectories (using potentially numerical algorithms) and prior probabilities. For prior probabilities, the estimate is given simply by:

$$\pi_k^{(m+1)} = \frac{1}{N} \sum_{i=1}^N z_{ik}^{(m+1)}.$$

This method is classically used for models without random effects, but it was evaluated and validated for mixed effects models.<sup>5</sup> In the M step, the estimates of  $\boldsymbol{\beta}_k$ ,  $\mathbf{D}_k$ , and  $\sigma_k^2$  can be obtained using classical algorithms for mixed effects models, such as the ones included in the *lme* function from the *nlme* R package<sup>10</sup> (R version 3.6.2). The R programs is available in the Supporting Information. These three steps are repeated until the difference in the log classification likelihood between two consecutive iterations is less than a given threshold value or when there is no change in the partition. As the solution provided is dependent on the initial partition,<sup>7</sup> the CEM algorithm is repeated several times from different initial partitions and the partition that provides the highest value of the log classification likelihood is kept.

## 2.3 | Inference method

For inference, the usual estimators of standard errors of mixed effects models can be used, but these do not take into account the uncertainty in the obtained partition. The method proposed in this article uses a stochastic algorithm, named SEM, to generate randomness on the partition obtained using the CEM algorithm. The randomness is then included in the variance of the estimators using Rubin's method.<sup>11</sup>

The SEM (stochastic EM) algorithm is a stochastic version of the CEM algorithm.<sup>7</sup> The main idea is to replace the C-step by an S-step of random classification according to the posterior probabilities. The value of  $z_{ik}^{(m)}$  is no longer determined by the maximum posterior probability but is drawn according to a multinomial distribution for a given posterior probability of belonging to each group. Usually, this stochastic algorithm, proposed by Celeux et al is used to overcome the limitations of EM and CEM algorithms and avoid the suboptimal solutions linked to the initial partition.<sup>7,12,13</sup> The authors propose to apply the SEM algorithm before applying the CEM one, because generating randomness permits to investigate likelihood in more depth, and allows to provide a better initial partition. Other solutions can be used to avoid suboptimal solutions, for example, by running the CEM algorithm from different initial partitions, then keeping the solution that lead to the maximum likelihood. The present article does not focus on the way to obtain the optimal solution, but on an approach for parameter inference that takes into account the uncertainty on the classification. The SEM algorithm was applied after the CEM one to correctly estimate the variance. Several iterations of the SEM were performed on the solution obtained with the CEM algorithm. During each iteration, randomness is generated around the classification according to the posterior probabilities, then parameter are estimated with their standard errors. The randomness is then included in the variance of the estimators using Rubin's method; the variance of the pooled parameter estimate is calculated from the within and between SEM iteration variance described as follows.

Let  $\beta$  be the parameter to be estimated and  $V_\beta$  its variance. The realization of  $W$  classifications from the SEM algorithm leads to  $W$  estimations of  $\beta$  and  $V_\beta$ , denoted by  $\hat{\beta}^m$  and  $\hat{V}_\beta^m$ . The computation of variance is therefore based on the following Rubin's formula<sup>11</sup>:

$$\hat{V}_\beta^* = \frac{1}{W} \sum_{m=1}^W \hat{V}_\beta^m + \frac{W+1}{W} \times \frac{\sum_{m=1}^W (\hat{\beta}^m - \hat{\beta}^*)^2}{W-1}.$$

The estimator of the parameter asymptotically follows a normal distribution centered on the true value. Therefore, a Wald confidence interval can be obtained.

### 3 | SIMULATIONS

#### 3.1 | Design

Five types of simulations were performed to analyze the effect of different factors on the results. First, the effect of the heterogeneity inside groups was analyzed. Then the effects of the sample size and the number of measurements were analyzed. The type I error rate was also assessed. Finally, the sensitivity of the results according to the number of SEM iterations was analyzed.

**General design:** Simulations were defined as follows:  $N$  patients,  $T$  repeated measurements,  $K = 3$  latent groups with the following proportions, 22%, 31%, and 47%. In each group, there were two different treatment arms: half of the patients received the experimental treatment and the other half received the control treatment. Due to the randomization, there was no treatment effect at baseline, but only on the slope of the trajectories ( $\beta_{\text{Trt}}$ ). Typical trajectories, shown in Figure 1, were defined by a second-order polynomial:

$$\beta_{0k} + (\beta_{01} + \beta_{\text{Trt}} \times 1_{\text{Trt}}) \times \mathbf{t} + \beta_{2k} \times \mathbf{t}^2,$$

where  $\beta_0 = (10 \ 7 \ 2)$ ,  $\beta_1 = (-0.4 \ 0.01 \ 0.6)$ ,  $\beta_2 = (0.02 \ 0.08 \ 0.02)$ , and  $\beta_{\text{Trt}} = -0.1$  (treatment effect on the slope).

In each group, the individual trajectories were simulated by considering a residual SD  $\sigma = 2$ , and by taking into account Gaussian random effects on the intercept and slope. These effects were centered on zero, with variance-covariance matrix  $\mathbf{D}_k = \mathbf{D}$  (which was considered identical across groups).

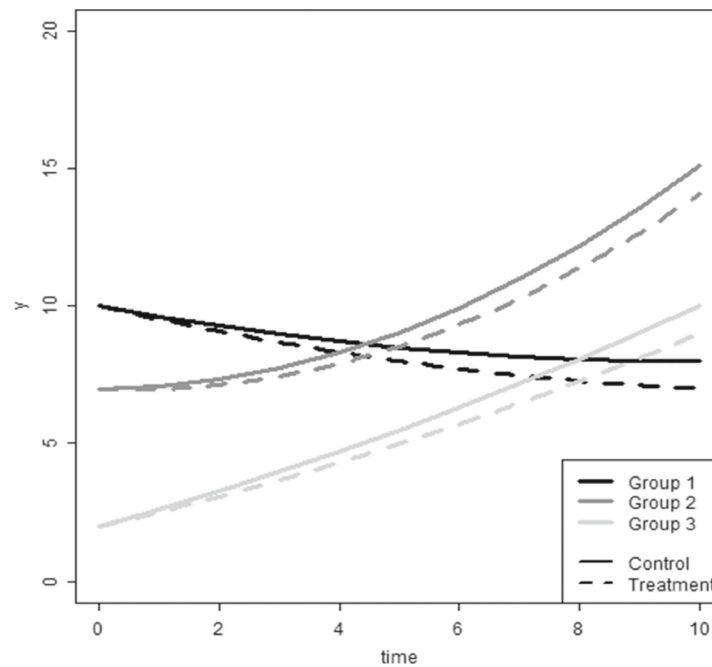
**Types of simulations:** To analyze the effect of the heterogeneity between trajectories inside groups, the variance-covariance matrix  $\mathbf{D}$  of the random effects was changed from one scenario to another. There were 320 trajectories ( $N = 320$ ) and 11 repeated measurements ( $T_i = 11$ ). Four different scenarios were defined. Except for scenario S1, the correlation between the intercept and the slope was close to 0.5 (S1: correlation of 0.32). For the four scenarios, the matrix  $\mathbf{D}$  was chosen according to the ratio of the between-individual variance ( $\text{trace}(\mathbf{t}_i \mathbf{D} \mathbf{t}_i')$ ) over the total variance within groups ( $\text{trace}(\mathbf{t}_i \mathbf{D} \mathbf{t}_i' + \sigma^2 \mathbf{I}_{T_i})$ ). The first scenario S1 had a very low heterogeneity inside groups that increased from S2 to S4.

$$\begin{aligned} \text{S1: } \mathbf{D} &= \begin{pmatrix} 0.01 & 0.001 \\ 0.001 & 0.001 \end{pmatrix}, & \text{S2: } \mathbf{D} &= \begin{pmatrix} 0.4 & 0.1 \\ 0.1 & 0.1 \end{pmatrix}, \\ \text{S3: } \mathbf{D} &= \begin{pmatrix} 0.8 & 0.25 \\ 0.25 & 0.3 \end{pmatrix}, & \text{S4: } \mathbf{D} &= \begin{pmatrix} 2 & 0.6 \\ 0.6 & 0.7 \end{pmatrix}. \end{aligned}$$

The classification became more difficult when the variances were higher, likely due to more overlap between trajectories, inducing more uncertainty on the classification.<sup>14,15</sup>

For the analysis of the effect of sample size, performed only for the scenario S2, different numbers of trajectories were simulated:  $N = 320$ ,  $N = 160$ ,  $N = 96$ , and  $N = 64$  (with  $T_i = 11$ ).

To analyze the effect of the number of measurements per trajectory, again only for scenario S2,  $T_i$  was decreased to 5 (the time interval was the same from 0 to 10, the timepoints 0, 2.5, 5, 7.5, and 10 were choose). The effect of the number of measurements was assessed for different numbers of trajectories:  $N = 320$ ,  $N = 160$ ,  $N = 96$ , and  $N = 64$ .



**FIGURE 1** Typical trajectories of the three groups in the control and treatment arms, used for simulations. Solid lines: control arms. Dashed lines: treatment arms. Black lines: group 1; dark grey lines: group 2. Light grey lines: group 3

Then the type I error rate was evaluated using the S2 scenario,  $N = 320$ ,  $T_i = 11$ , without treatment effect ( $\beta_{\text{Tt}} = 0$ ).

Finally, a sensitivity analysis was performed by changing the number of iterations of the SEM algorithm for the S2 scenario,  $N = 320$ ,  $T_i = 11$ , and  $\beta_{\text{Tt}} = -0.1$ . The number of iterations tested was  $W = 10$ ,  $W = 50$ ,  $W = 75$ ,  $W = 100$ ,  $W = 150$ , and  $W = 200$ .

For each simulation design, 1000 datasets were simulated. The classification model applied to the dataset was defined as the model used for data simulation. The CEM algorithm was initialized from the true partition because the performances of the CEM algorithm were already evaluated in a previous paper.<sup>5</sup> Except for the sensitivity analyses on the number of SEM iterations, all simulations were performed with a number of SEM iterations equal to 100.

**Evaluation criteria:** The validity of the inference method was assessed by two criteria. First the coverage probability of the 95% confidence interval, defined by the proportion of datasets for which the confidence interval contained the theoretical value, along with the width of the confidence interval. Second the type I error rate was defined by the proportion of datasets for which the null treatment hypothesis was rejected (the  $P$ -value was below 0.05) under null treatment hypothesis. For this latter criterion, only the scenario S2 with 320 trajectories, 11 measurements, and  $\beta_{\text{Tt}} = 0$  was applied.

Other criteria evaluating the performance of the model were given: the misclassification percentage and the average bias of predicted trajectories (which was defined as the average, in all individuals, of the difference between the true typical trajectories and the estimated ones calculated in each individual).

### 3.2 | Results

**Heterogeneity inside groups:** The coverage probability of the 95% confidence intervals decreases with the increase in ratio of the between-individual variance over the total variance within groups; the percentage of misclassification as well as the bias increase with the increase in that ratio (Table 1). The coverage probability of scenarios S1 and S2 are close to 95% while those of scenarios S3 and S4 are generally lower than 95%, the lowest being 87.5%. The width of the confidence intervals for the different estimates of the model parameters increases with the increase in the random effects variance-covariance matrix (Appendix, Table A1). As shown in Table A2 and Figure A1, for each parameter in each group, the variances of the parameter estimates for the method that takes into account the uncertainty in the classification are

TABLE 1 Coverage probability of each parameter's 95% confidence interval, misclassification percentage (total and per group), average bias of predicted trajectories by group and arm, according to the scenario

Scenario	Coverage probability of the 95% confidence intervals (%)										% of mis-classification		Average bias of predicted typical trajectories					
	Gr. 1			Gr. 2			Gr. 3			Trt		Control arm			Treatment arm			
	$\beta_{01}$	$\beta_{11}$	$\beta_{21}$	$\beta_{02}$	$\beta_{12}$	$\beta_{22}$	$\beta_{03}$	$\beta_{13}$	$\beta_{23}$	$\beta_{Trt}$		Gr. 1	Gr. 2	Gr. 3	Gr. 1	Gr. 2	Gr. 3	
$\begin{pmatrix} 0.01 & 0.001 \\ 0.001 & 0.001 \end{pmatrix}$	96.1	95.1	95.5	96.0	96.3	95.3	93.3	93.6	93.4	94.8	0.2 (0.4, 0.3, 0.1)	−0.003	0.002	<0.001	<0.001	0.005	0.002	
S1																		
$\begin{pmatrix} 0.4 & 0.1 \\ 0.1 & 0.1 \end{pmatrix}$	93.9	94.3	94.4	93.3	93.5	93.4	94.5	94.8	95.0	95.2	5.1 (3.4, 8.3, 3.7)	−0.022	0.061	−0.045	−0.010	0.074	−0.033	
S2																		
$\begin{pmatrix} 0.8 & 0.25 \\ 0.25 & 0.3 \end{pmatrix}$	93.0	94.1	93.7	91.1	92.6	92.8	93.0	93.8	94.3	95.7	7.5 (5.7, 12.0, 5.4)	−0.004	0.084	−0.063	−0.007	0.081	−0.066	
S3																		
$\begin{pmatrix} 2 & 0.6 \\ 0.6 & 0.7 \end{pmatrix}$	90.2	93.1	93.7	88.2	92.1	90.2	87.5	91.1	93.4	94.9	12.0 (11.7, 18.3, 8.8)	0.067	0.133	−0.097	0.008	0.074	−0.156	
S4																		

Abbreviations: Gr. 1, group 1; Gr. 2, group 2; gr. 2, group 3; S1, S2, S3, S4: scenarios for the random effects variance–covariance matrix.

**TABLE 2** Comparison of the coverage probability of the 95% confidence intervals obtained with the methods that take into account or not classification uncertainty, for scenarios S1 to S4

Scenario	Classification uncertainty taken into account	Coverage probability of the 95% confidence intervals (%)									
		Group 1			Group 2			Group 3			Trt
		$\beta_{01}$	$\beta_{11}$	$\beta_{21}$	$\beta_{02}$	$\beta_{12}$	$\beta_{22}$	$\beta_{03}$	$\beta_{13}$	$\beta_{23}$	$\beta_{Trt}$
S1	Yes	96.1	95.1	95.5	96.0	96.3	95.3	93.3	93.6	93.4	94.8
	No	96.1	95.1	95.5	95.8	96.3	95.3	93.2	93.6	93.4	94.2
S2	Yes	93.9	94.3	94.4	93.3	93.5	93.4	94.5	94.8	95.0	95.2
	No	91.0	93.7	93.5	85.4	91.2	91.1	89.4	93.9	93.8	94.0
S3	Yes	93.0	94.1	93.7	91.1	92.6	92.8	93.0	93.8	94.3	95.7
	No	89.0	93.4	93.0	78.3	88.9	89.5	85.1	92.7	92.5	94.8
S4	Yes	90.2	93.1	93.7	88.2	92.1	90.2	87.5	91.1	93.4	94.9
	No	80.5	91.1	91.7	64.2	86.0	80.7	73.4	89.6	89.7	90.4

Abbreviations: S1, S2, S3, S4, scenarios for the random effects variance–covariance matrix; Trt, treatment effect.

close to the ones of the method that does not take into account the uncertainty for scenario S1, which is the scenario with low between-individual variance. However, for the scenarios with higher between-individual variance, variances with the new method are higher than the ones obtained using the classical method, especially for group 2. Since the typical trajectory of group 2 is between the typical trajectories of groups 1 and 3, the uncertainty in classification is higher for that group. These results are reflected in the results of the coverage probability of the 95% confidence intervals presented in Table 2. For scenario S1, the coverage probabilities of the two methods are similar. For the other scenarios, the coverage probability of the classical confidence intervals decreases much more than the one of the confidence intervals that take into account the classification uncertainty. For example, for group 2 and scenario S2, the coverage probability of the intercept is 93.3% for the one that takes into account the classification uncertainties compared to 85.4% for the one that does not.

**Sample size:** The misclassification percentages and the coverage probabilities remain stable when the number of patients decreases (total number of measurements by trajectory  $T_i = 11$ , Table 3). As expected, the width of the confidence intervals increases when the number of trajectories decreases (Appendix, Table A3).

**Number of measurements per trajectory:** The misclassification percentages, and consequently the bias, increases with the decrease in the number of measurements, leading to wider confidence intervals. This however, has low impact on the coverage probabilities (Table 3—Appendix, Table A3).

**Type I error rate:** For simulations under the null treatment hypothesis, the estimated type I error rate was 4.6%.

**Number of SEM iterations:** For the last simulations, the coverage probabilities remain stable whatever the number of iterations of the SEM algorithm, even for  $W = 10$  (Table 4).

## 4 | APPLICATION

### 4.1 | Context

In the “low-cyclo” trial,<sup>15</sup> 95 heart transplant patients aged 18 to 65 years were centrally randomized to receive either low-dose or standard-dose cyclosporine A (CsA concentrations of 130–200  $\mu\text{g/L}$ ,  $N_1 = 47$  vs 2002–300  $\mu\text{g/L}$ ,  $N_2 = 48$ ) during the first 3 months after transplantation. The primary endpoint was serum creatinine using eight measurements: one every 50 days for the first 3 months, one per month until 6 months, and then one per quarter. A mixed model was used by Boissonat et al,<sup>15</sup> but no significant difference was found in creatinine over time between the two arms. Figure 2 shows high heterogeneity of the individual trajectories of creatinine over time. The hypothesis was the presence of groups within the population, with an effect of low-dose CsA only for some of them. A mixed effects classification model thus seemed well-suited in this example for the estimation of low-dose CsA on each latent group.

**TABLE 3** Coverage probability of each parameter's 95% confidence interval, misclassification percentage (total and per group), average bias of predicted trajectories by group and arm, according to the number of patients and the number of measurements

Number of measures	Number of patients	Coverage probability of the 95% confidence intervals (%)										% of misclassification		Average bias of predicted typical trajectories					
		Group 1			Group 2			Group 3			Trt			Control arm			Treatment arm		
		$\beta_{01}$	$\beta_{11}$	$\beta_{21}$	$\beta_{02}$	$\beta_{12}$	$\beta_{22}$	$\beta_{03}$	$\beta_{13}$	$\beta_{23}$				Gr. 1	Gr. 2	Gr. 3	Gr.1	Gr. 2	Gr. 3
11	320	93.9	94.3	94.4	93.3	93.5	93.4	94.5	94.8	95.0	95.2	5.1 (3.4, 8.3, 3.7)		-0.022	0.061	-0.045	-0.010	0.074	-0.033
	160	95.3	94.6	95.5	94.6	94.6	94.5	92.0	95.3	95.3	94.4	5.0 (3.3, 8.2, 3.7)		-0.008	0.073	-0.043	0.005	0.086	-0.031
	96	94.3	94.2	94.3	93.7	93.6	94.5	94.0	94.0	94.3	94.9	5.0 (3.3, 8.2, 3.7)		0.014	0.060	-0.033	0.031	0.077	-0.016
	64	95.6	94.8	93.7	94.8	94.6	95.6	94.7	94.5	94.7	93.6	5.0 (3.4, 8.2, 3.6)		-0.006	0.088	-0.045	0.002	0.093	-0.040
5	320	91.3	93.1	91.5	86.8	90.6	91.3	90.5	94.3	94.0	93.7	10.1 (9.6, 16.2, 6.2)		-0.063	0.206	-0.086	-0.051	0.218	-0.074
	160	92.8	92.4	93.6	92.1	91.1	91.5	92.2	93.7	93.4	93.5	10.3 (9.5, 16.4, 6.6)		-0.049	0.194	-0.112	-0.019	0.224	-0.082
	96	96.4	94.2	95.2	93.2	93.6	92.7	93.1	94.1	95.3	93.0	10.5 (9.8, 16.8, 6.6)		-0.085	0.158	-0.132	-0.015	0.228	-0.062
	64	95.2	94.2	94.2	93.8	94.6	93.9	94.2	94.4	94.4	93.2	10.5 (10.0, 16.9, 6.4)		-0.060	0.206	-0.106	-0.034	0.232	-0.080

Abbreviations: Gr. 1, group 1; Gr. 2, group 2; Gr. 3, group 3; Trt, treatment effect.

TABLE 4 Coverage probability of the 95% confidence intervals according to the number of SEM iterations

Number of SEM iterations	Coverage probability of the 95% confidence intervals (%)								
	Group 1			Group 2			Group 3		
	$\beta_{01}$	$\beta_{11}$	$\beta_{21}$	$\beta_{02}$	$\beta_{12}$	$\beta_{22}$	$\beta_{03}$	$\beta_{13}$	$\beta_{23}$
10	95.0	95.6	94.8	92.0	94.4	94.3	94.1	93.5	94.9
50	95.0	95.5	95.1	91.9	94.1	94.5	94.4	93.8	94.9
75	94.9	95.5	95.1	92.1	94.2	94.5	94.6	93.7	95.0
100	95.1	95.5	94.9	92.8	94.4	94.5	94.6	93.7	94.9
150	95.0	95.5	95.1	92.9	94.3	94.5	94.7	93.7	94.7
200	95.0	95.5	95.0	92.6	94.3	94.5	94.6	93.7	94.7

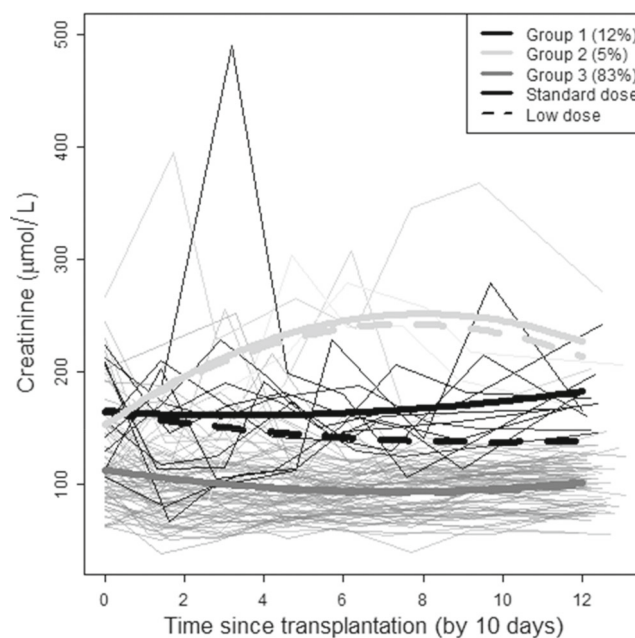


FIGURE 2 Observed trajectories (thin lines) colored according to the classification obtained via the CEM algorithm and typical trajectories (thick lines) obtained with the mixed effects classification model. Solid lines: standard dose. Dashed lines: low-dose. For group 3 the two typical trajectories overlapped. Dark grey line: mean trajectory of group 1; light grey line: mean trajectory of group 2; medium grey line: mean trajectory of group 3

## 4.2 | Mixed effects classification model

A mixed effects classification model was applied to classify the 95 patients into three different groups, to be consistent with previous work on these data.<sup>3</sup> To model the creatinine trajectories, a second-order polynomial relationship was used. Only the intercept and linear slope were considered random. These effects were centered on zero, with variance-covariance matrix  $D$  (which was considered identical across groups).

Assuming that the effect of the low-dose CsA on creatinine is progressive, an interaction between time and treatment was added to the model. For group  $k$  the model may be written as follows:

$$\text{Creat}_i = \beta_{0k} + b_{0ki} + (\beta_{1k} + b_{1ki} + \beta_{\text{dose } k} \times \text{dose}) \times \text{time} + \beta_{2k} \times \text{time}^2 + \epsilon_i$$

with  $b_{ki} \sim N(0, D)$  and  $\epsilon_i \sim N(0, \sigma^2 I)$ .

**TABLE 5** Low-Cyclo parameter estimates with their corresponding 95% confidence interval and *P*-value for the different groups

		Estimate	95% CI	<i>P</i> -value
Group 1	$\beta_1$	164.24	[147.68, 180.80]	0.000
	$\beta_2$	−1.80	[−17.94, 14.34]	0.827
	$\beta_3$	0.28	[−1.26, 1.81]	0.724
	$\beta_{\text{Trt}}$	−3.64	[−7.99, 0.71]	0.101
Group 2	$\beta_1$	152.53	[134.29, 170.77]	0.000
	$\beta_2$	24.71	[10.43, 38.99]	0.001
	$\beta_3$	−1.54	[−2.88, −0.20]	0.024
	$\beta_{\text{Trt}}$	−1.15	[−4.92, 2.63]	0.552
Group 3	$\beta_1$	112.72	[106.90, 118.53]	0.000
	$\beta_2$	−5.27	[−7.99, −2.55]	0.000
	$\beta_3$	0.36	[−0.19, 0.92]	0.203
	$\beta_{\text{Trt}}$	−0.15	[−1.85, 1.54]	0.860

The CEM algorithm was repeated 100 times from different initial partitions of the patients. The confidence interval and test result were given for fixed parameters according to the new method presented in this article.

### 4.3 | Results

The three groups obtained appear to have clinical meaning (Figure 2). Group 1, which includes 11.6% of the population, corresponds to patients with low kidney function immediately after heart transplantation but who somewhat improved with time after transplantation. These patients probably had a renal insufficiency related to the heart disease. Group 2, in which 5.3% of the population is included, corresponds to patients with low kidney function and whose creatinine levels were high and even increased at the beginning of the follow-up. These patients probably had organic renal insufficiency. At last, group 3 corresponds to patients with normal levels of creatinine at the time of heart transplantation that remained stable after surgery. These patients, who represent 83.2% of the population, had good kidney function.

According to Figure 2, the treatment has a visible effect on groups 1 and 2 as found in another article<sup>3</sup> that used the same data. The estimations of the coefficients in Table 5 also reflect a more important impact of low-dose CsA for groups 1 and 2. The parameter estimates are provided with their confidence interval and *P*-value. According to these last results, the dose effect is not statistically significant. Figures A2 and A3 show the typical trajectories in each treatment arm and the corresponding pointwise confidence intervals.

## 5 | DISCUSSION

The present article presents a method which takes into account classification uncertainties in the variance of the parameter estimates of a mixed effects classification model. Overall, simulations showed that the coverage probabilities of the 95% confidence intervals are close to 95%, except in some scenarios where the variances of the random effects are high. Furthermore, the coverage is better (up to 24% higher) than the one obtained by the classical method that does not take into account classification uncertainties. The ratio of the two estimate variances was always greater than 1 and reached 2.910. Trajectory classification may be performed using classification models without random effects; in this case, the uncertainty in the classifications may be lower, and the difference in coverage probabilities between the two methods for confidence interval construction (the one taking into account classification uncertainties and the one that does not) is much lower.

The simulation results showed that the probability of coverage is generally lower than 95% when the misclassification, and consequently the bias, of typical trajectories are high, for example, in case of high heterogeneity inside groups. An

appropriate variance for the estimated parameters will not overcome the bias in the estimate, and hence will result in poor coverage probabilities.

The width of the confidence intervals increased with the decrease in the number of trajectories/measurements and the increase in the random effect variance. The typical trajectory of a group is all the more uncertain—and may not be relevant—when the heterogeneity inside the group increases; it is hence logical that the width of the confidence intervals increases. Of note, with respect to the scale, the widths of the confidence intervals for the slope parameters were generally larger than the ones of the other parameters; this is explained by the fact that the treatment had an effect on the slope of the trajectories.

Since there was enough information on each trajectory due to a high number of measurements, the coverage probabilities were not greatly modified when the number of patients decreased reasonably.

The variations in the number of SEM algorithm iterations did not greatly modify the coverage probability of the confidence intervals and their width. This can be explained by the good characteristics of the scenario S2 ( $N = 320$ ,  $T_i = 11$ ). For multiple imputation, it is agreed that the number of imputations does not need to be high (usually less than 20).<sup>16</sup> However, more work would be needed on different scenarios to be sure that less than 20 SEM iterations are sufficient to achieve appropriate coverage probabilities, and that is why for most of the simulations and for the application, 100 SEM iterations were used.

In this work, the proposed method to obtain the confidence interval of the parameters was evaluated only on the less complex model; that is, the classification model with a between individual variance equal across groups. It would be interesting to evaluate also this method on the classification model with distinct between individual variance across groups.<sup>14,17</sup>

The constructed confidence intervals for low-cyclo were very wide and the effect of low-dose CsA was not significant for groups 1 and 2. This lack of significance may be due to a lack of power since groups 1 and 2 contained very few patients.

Studies on low-cyclo have shown negative results with regard to the treatment effect,<sup>18</sup> explained mostly by the high heterogeneity between patients. A mixture mixed model could have been used to take into account the heterogeneity. However, due to biological reasons, the low-dose CsA may be efficient only for some patients, mostly patients that have an impaired kidney function. Hence, the treatment effect may be group-dependent. It is therefore necessary to have well-defined groups of patients that are clinically relevant, and the use of a classification model instead of a mixture one thus seems more appropriate to identify patients who may benefit from the dose reduction. Generally, classification approaches are used only to identify groups, and in that case confidence intervals are not really necessary. However, in this specific example, an inference method is necessary to estimate the effect of a parameter—here the treatment—for each latent group.

It could be interesting to investigate factors that may be associated with a group membership and that may predict the latent groups (such as age, sex, medical history, etc). A multinomial logistic distribution could be used to model the prior probabilities according to these factors. Once again, an inference method would be necessary to test the association between such factors and the latent groups, and the one proposed in this article could be applied.

The plotted confidence intervals for the typical trajectories are pointwise ones (this means that, at a given time, 95% of the confidence intervals contained the true value). In practice, the question is often: does a new trajectory belong to a specific group. To answer this question, one should use fluctuation intervals, which are constructed using random effects and random fluctuations, and are defined on the whole trajectory set, that is to say confidence bands (a region containing the whole trajectory, therefore at any time).<sup>19</sup> Further work is needed to define confidence bands.

Although the results obtained from the simulations were good, they concerned a very specific example, and might thus not be generalizable to other examples. Moreover, the number of measurements herein was relatively high, which is not always the case with clinical data, and the simulation scenarios did not include missing data. The validity of the method, and mostly its limitations, need to be further investigated.

## 6 | CONCLUSION

For inference after a classification based model, it is recommended to take into account the classification uncertainty in the variance of the parameter estimates. This can be performed by using the SEM algorithm after convergence of the CEM algorithm, combined with the Rubin's formula. The method leads to an appropriate coverage probability of the confidence intervals, except in some situations, notably when the heterogeneity inside groups is high.

## ACKNOWLEDGMENTS

The authors would like to thank to Pascale Boissonnat for providing Low-cyclo data. They also thank Mrs Véréna Landel for helpful comments, suggestions, and revisions of the final manuscript.

## CONFLICT OF INTEREST STATEMENT

The authors declare having no conflict of interest.

## DATA AVAILABILITY STATEMENT

The program for the mixed effects classification with the associated inference and the simulation datasets generated for the current study are available on request from the corresponding author. However, the clinical dataset are not publicly available.

## ORCID

Amna Abichou-Klich  <https://orcid.org/0000-0002-3052-7335>

## REFERENCES

1. Nagin DS, Odgers C. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol*. 2010;6:109-138.
2. Pickels A, Croudace T. Latent mixture models for multivariate and longitudinal outcomes. *Methods Med Res*. 2010;19(3):271-289. doi:10.1177/0962280209105016
3. Klich A, Ecochard R, Subtil F. Unequal intra-group variance in trajectory classification. *Stat Med*. 2018;37(28):4155-4166. doi:10.1002/sim.7921
4. Subtil F, Boussari O, Bastard M, Etard JF, Ecochard R, Génolini C. An alternative classification to mixture modeling for longitudinal counts or binary measures. *Stat Methods Med Res*. 2017;26:453-470. doi:10.1177/0962280214549040
5. Klich A, Ecochard R, Subtil F. Trajectory clustering using mixed classification models. *Stat Med*. 2021;40:1-15.
6. Celeux G, Goavert G. Comparison of the mixture and the classification maximum likelihood in cluster analysis. *J Stat Comput Simul*. 1993;47:127-146.
7. Celeux G, Goavert G. A classification EM algorithm for clustering and two stochastic versions. *Comput Stat Data Anal*. 1992;14:315-332.
8. Dempster AP, Laird N, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B*. 1977;39(1):1-38.
9. McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions*. 2nd ed. New York: John Wiley & Sons; 2007.
10. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. New York: Springer; 2000.
11. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons; 2004.
12. Celeux G, Diebolt J. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comput Stat Q*. 1985;2:73-82.
13. Biernacki C, Celeux G, Goavert G. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Comput Stat Data Anal*. 2003;41:561-575.
14. Diallo TMO, Morin AJS, Lu HZ. Impact of misspecifications of the latent variance-covariance and residual matrices on the class enumeration accuracy of growth mixture models. *Struct Equ Model*. 2016;23(4):507-531.
15. Kim ES, Wang Y. Class enumeration and parameter recovery of growth mixture modeling and second-order growth mixture modeling in the presence of measurement noninvariance between latent classes. *Front Psychol*. 2017;8:1499.
16. Soullier N. Traitement de la non-réponse par imputation multiple; 2012. [http://www.jms-insee.fr/2012/S06\\_3\\_ACTE\\_SOULLIER\\_JMS2012.PDF](http://www.jms-insee.fr/2012/S06_3_ACTE_SOULLIER_JMS2012.PDF)
17. Heggeseth BC, Jewell NP. The impact of covariance misspecification in multivariate Gaussian mixtures on estimation and inference: an application to longitudinal modeling. *Stat Med*. 2013;32:2790-2803.
18. Boissonnat P, Gaillard S, Mercier C, et al. Impact of the early reduction of cyclosporine on renal function in heart transplant patients: a French randomised controlled trial. *Trials*. 2012;13(231):231. doi:10.1186/1745-6215-13-231
19. Maringwa JT, Geys H, Shkedy Z, et al. Application of semiparametric mixed models and simultaneous confidence bands in a cardiovascular safety experiment with longitudinal data. *J Biopharm Stat*. 2008;18:1043-1062.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Dugourd C, Abichou-Klich A, Ecochard R, Subtil F. Clustering of trajectories with mixed effects classification model: Inference taking into account classification uncertainties. *Statistics in Medicine*. 2023;42(25):4570-4581. doi: 10.1002/sim.9876