STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:2229–2245
Published online 10 August 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2659

A nonlinear latent class model for joint analysis of multivariate longitudinal data and a binary outcome

Cécile Proust-Lima[1,2,*,†], Luc Letenneur[2,3] and Hélène Jacqmin-Gadda[1,2]

[1]*Institut National de la Santé et de la Recherche Médicale, Equipe de Biostatistique E0338, 33076 Bordeaux, France*
[2]*Université Victor Segalen, 33076 Bordeaux, France*
[3]*Institut National de la Santé et de la Recherche Médicale, Unité 593, 33076 Bordeaux, France*

SUMMARY

We consider a joint model for exploring association between several correlated longitudinal markers and a clinical event. A nonlinear growth mixture model exhibits the different latent classes of evolution of the latent quantity underlying the correlated longitudinal markers and a logistic regression models the probability of occurence of the clinical event according to the latent classes. By introducing a flexible nonlinear transformation including parameters to be estimated between each marker and the latent process, the model also deals with non-Gaussian continuous markers. Through an application on cognitive ageing, the two advantages of the model are underlined: (1) the latent profiles of evolution associated with the clinical event are described including covariate effects in the longitudinal model but also in the probability of class membership and in the probability of occurence of the event, and (2) a diagnostic and a prognostic tools are derived from the model for early detection of the clinical event using any available information about the longitudinal markers. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: cognitive ageing; dementia; early detection; growth mixture model; joint modelling

1. INTRODUCTION

In medical research, it is often of great interest to explore the association between evolution of longitudinal quantitative outcomes and a clinical event in order to both describe the evolution of markers during the course of a chronic disease, and to perform early detection of the disease using data on longitudinal evolution of the quantitative markers. In the screening of prostate cancer, modelling association between prostate-specific antigen (PSA) and prostate cancer showed that a

*Correspondence to: Cécile Proust-Lima, INSERM E0338, ISPED, Université de Bordeaux2, 146 rue Léo Saignat, 33076 Bordeaux cedex, France.
†E-mail: cecile.proust@isped.u-bordeaux2.fr

Received 19 December 2005
Accepted 13 June 2006

significant increase of PSA could be predictive of a cancer some years later [1]. In the treatment of Alzheimer's disease, it is now admitted that the earlier the treatment administrated, the better the evolution of the disease. The joint study of longitudinal markers of cognition and diagnosis of dementia could hence be useful for initiating the treatment earlier.

For studying the association between evolution of quantitative outcomes and a clinical event either in a descriptive perspective or a prognostic perspective, two kinds of joint models have been proposed: shared random-effect models and latent class models. A shared random-effect model [2, 3] consists in modelling the repeated quantitative outcome with a mixed model and including the individual random coefficients as covariates in the model for the event. In cognitive ageing context, Jacqmin-Gadda *et al.* [4] proposed a shared random-effect model with random change points in the specific aim of estimating a change point in the cognitive decline toward dementia. In contrast, a latent class model [1, 5, 6] assumes that the population is made of various subpopulations with different longitudinal evolutions modelled by a latent class variable; the risk of event depends on the longitudinal outcome only through the latent classes. Latent class models have some advantages over shared random-effect models. Indeed, in the latter, the assumption that the random-effects come from a common Gaussian distribution is quite unrealistic when the population consists of several subpopulations (at least ill and not ill subjects). Moreover, by exhibiting profiles of evolution associated with the clinical event, latent class models are simpler to interpret compared with the shared random-effect models which estimate correlations between the event and the random-effects. In particular, latent class models are an attractive tool for clinicians because profiles of evolution are easily drawn for each class, the impact of covariates on the probability of each profile are evaluated and the probability of the event in each latent class is estimated. At last, shared-random-effect models are computationally more demanding due to the numerical integrations across the random-effects required in the computation of the likelihood. In some latent class models, integrals across the random-effects may have a closed form and the dependency between the two parts of the model requires only a sum across the values of the latent class variable.

Latent class models for joint analysis of a longitudinal outcome and an event have already been developed. Muthén and Shedden [6] studied the relation between the shape of heavy drinking trajectory in the 18–25 year age range and the probability of alcohol dependency at age 30. Lin *et al.* [5] extended the model to irregularly spaced longitudinal readings and unequal number of measures for each subject in the context of the prediction of prostate cancer according to PSA profiles. Finally, Lin *et al.* [1] proposed to model the risk of prostate cancer using a semi-parametric survival model instead of a logistic model.

In medical research, various markers are often collected repeatedly and it can be of interest to make use of the information from all the markers [7–9]. In many cases, the markers are highly correlated and may be viewed as various measures of a common underlying quantity which is modelled using a latent process. When the markers are continuous, they are modelled using the Gaussian assumption [8–10] although in many applications, they have a distribution far from a Gaussian distribution. For instance, in neuropsychology, the latent cognitive level is measured by several correlated psychometric tests which often exhibit a non-Gaussian continuous distribution. Recently, Ganiayre *et al.* [11] extended the latent process approach of Hashemi *et al.* [12] to model jointly ordinal psychometric tests and dementia assuming dementia was defined as the crossing of an estimated level by the latent process. However this approach does not take into account subpopulation structure and is numerically untractable for multiple longitudinal outcomes. Proust *et al.* [13] proposed a nonlinear model with a latent

DOI: 10.1002/sim

process to analyse multivariate and non-Gaussian longitudinal data by using flexible parametrized nonlinear transformations to model the relationship between the longitudinal outcomes and the latent process.

In this paper, our objective is to extend the joint model developed by Lin *et al.* [5] by using the nonlinear multivariate approach described in Reference [13] in order to propose a latent class model for the joint analysis of non-Gaussian multivariate longitudinal data and a binary outcome. This work was mostly motivated by the study of cognitive ageing. We aimed at: (i) describing profiles of cognitive evolution associated with dementia, and (ii) developing a tool for early diagnostic making use of information from repeated measures of several psychometric tests.

In the following section, we define the joint nonlinear latent class model. In Section 3, we compute the log-likelihood of the joint model and the posterior probabilities stemmed from the mixture model. We then propose a diagnostic and a prognostic tool for detecting the event. Section 4 focuses on an application to data from the French prospective cohort study PAQUID [14]. We conclude in Section 5 with a discussion.

2. METHODOLOGY

2.1. Nonlinear model for multivariate longitudinal data

Consider K correlated continuous outcomes. Each outcome k, $k=1,\ldots,K$ is measured on each subject i, $i=1,\ldots,N$ at n_{ik} occasions. For outcome k and subject i, the vector of measurements is $y_{ik}=(y_{i1k},\ldots,y_{ijk},\ldots,y_{in_{ik}k})$, j representing the occasion. The times of measurements denoted t_{ijk} may be different for each subject and each outcome.

Consider now a latent process $\Lambda_i=(\Lambda_i(t))_{t\geqslant 0}$ defined in continuous time for each subject i and representing the quantity underlying the K outcomes. We assume that the measurement y_{ijk} is related to the latent process at time t_{ijk} through a flexible monotone increasing transformation h_k depending on an outcome-specific vector of parameters η_k to be estimated. This measurement model is specified as follows:

$$h_k(y_{ijk};\eta_k)=\Lambda_i(t_{ijk})+\alpha_{ik}+\varepsilon_{ijk} \tag{1}$$

where α_{ik} are random intercepts independently distributed according to a $N(0,\sigma^2_{\alpha_k})$ distribution and ε_{ijk} are independent Gaussian errors with mean zero and variance $\sigma^2_{\varepsilon_k}$.

The random coefficient α_{ik} introduces variability between the markers conditionally on the value of the latent process. For example, in ageing context, α_{ik} takes into account the fact that for a same value of the latent cognition, subjects can score differently in cognitive domains associated with the psychometric tests.

For the transformation h_k, we chose the Beta cumulative distribution function (CDF) which depends only on two parameters $\eta_k=(\eta_{1k},\eta_{2k})^{\mathrm{T}}$ and offers a large flexibility in the shapes [13]. As the Beta CDF is defined in [0, 1], a preliminary step consists in rescaling the tests to the unit interval.

2.2. Mixture model for the latent process

We assume the population consists of G subpopulations represented by G latent classes. Within each latent class g, $g=1,G$, the latent process follows an homogeneous linear mixed model. By

DOI: 10.1002/sim

defining c_{ig}, the latent variable which equals one if subject i belongs to the latent class g and zero otherwise (the sum of c_{ig} over the G classes equals one), the latent process evolution for a subject i given that he belongs to latent class g is written as follows:

$$\Lambda_i(t)|_{c_{ig}=1} = Z(t)^{\mathrm{T}}u_{ig} + X_{1i}(t)^{\mathrm{T}}\beta + X_{2i}(t)^{\mathrm{T}}\gamma_g, \quad t \geqslant 0 \tag{2}$$

where the $(p+1)$-vector $Z(t)^{\mathrm{T}} = (1, t, \ldots, t^p)$ is a time polynomial of degree p and the random-effect vector u_{ig} is distributed according to the class-specific Gaussian distribution $\mathrm{N}(\mu_g, \omega_g^2 B)$ with $\omega_1 = 1$. The vectors $X_{1i}(t)$ and $X_{2i}(t)$ are, respectively, a q_1-vector and a q_2-vector of possibly time-dependent covariates; $X_{1i}(t)$ is associated with the vector of fixed effects β which is common for all the classes while $X_{2i}(t)$ is associated with the class-specific vector of fixed effects γ_g. For ensuring the model identifiability, a covariate cannot be included both in $X_{1i}(t)$ and $X_{2i}(t)$.

We included a class-specific variance-covariance matrix for the random effects $\omega_g^2 B$ with $\omega_1 = 1$. Thus, even if the structure of the variance-covariance matrix in each class g, $g>1$, is the same as in the first class where the variance-covariance matrix is B, the global variability in class g, $g>1$, can differ from the variability in the first class thanks to the proportional term ω_g^2.

Note that $\Lambda_i(t)$ defined in equation (2) could also include a stochastic process such as a Brownian motion shared over the latent classes as proposed in Reference [13]. However models including both a mixture distribution and a Brownian motion can lead to numerical problems. We thus not included the Brownian motion in the mixture model formulation.

2.3. Model for the probability of belonging to class g

Each subject i has a probability π_{ig} of belonging to class g with $\sum_{g=1}^{G} \pi_{ig} = 1$. This class membership probability is modelled using a multinomial logit regression [6] including covariates:

$$\pi_{ig} = P(c_{ig} = 1|X_{3i}) = \frac{\mathrm{e}^{\xi_{0g} + X_{3i}^{\mathrm{T}}\xi_{1g}}}{1 + \sum_{j=2}^{G} \mathrm{e}^{\xi_{0j} + X_{3i}^{\mathrm{T}}\xi_{1j}}} \quad \forall g = 1, G \tag{3}$$

where ξ_{0g} is the intercept for class g and ξ_{1g} is the q_3-vector of class-specific parameters associated with the q_3-vector of time-independent covariates X_{3i}. For identifiability, $\xi_{01} = 0$ and $\xi_{11} = 0$. Thus, each element of ξ_{1g} is the log odds-ratio for the probability of belonging to class g compared to class 1 for a unit increase of the corresponding covariate.

2.4. Model for the probability of the clinical event

Consider the binary variable D_i (D_i equals 1 if the event occurs for i and 0 otherwise). We model the probability of D_i conditionally on the latent class variables $(c_{ig})_{g=1,G}$. It means that D_i has a different probability to occur in each latent class g. The probability of D_i conditionally on g can also differ according to a q_4-vector of covariates X_{4i} as follows:

$$P(D_i = 1|X_{4i}, c_{ig} = 1) = \frac{\mathrm{e}^{\delta_{0g} + X_{4i}^{\mathrm{T}}\delta_{1g}}}{1 + \mathrm{e}^{\delta_{0g} + X_{4i}^{\mathrm{T}}\delta_{1g}}} \tag{4}$$

where δ_{0g} is the log-odds for the occurrence of D_i in the latent class g and δ_{1gl}, the lth element of δ_{1g}, is the log odds-ratio for the occurrence of D_i in latent class g for a unit increase of X_{4il}, the lth covariate of X_{4i}.

DOI: 10.1002/sim

2.5. Covariates interpretation

A richness of the model is that it distinguishes three possible ways of including covariates with possible overlap between time-independent covariates included in X_1 or X_2 and in X_3 and X_4. However, as each way has a specific interpretation, inclusion of covariates must be done in accordance with the clinical hypothesis and the objective of the analysis. Covariates included in the latent process model (2) are assumed to impact only the mean evolution of the latent process either through a common effect over the classes (X_1) or through a class-specific effect (X_2). Covariates are included in the logistic model for the event (X_4) for evaluating their impact on the event risk after adjustment on the shape of the latent process evolution. In contrast, when a covariate is associated with the class membership probability (X_3), it means that the covariate has an effect on both the evolution of the latent process (on the mean and the random-effects variance) and on the event risk. Indeed, the two marginal distributions $f(\Lambda_i(t))$ and $P(D_i = 1)$ are given by

$$f(\Lambda_i(t)) = \sum_{g=1}^{G} \pi_{ig} f(\Lambda_i(t)|c_{ig} = 1) \tag{5}$$

$$P(D_i = 1) = \sum_{g=1}^{G} \pi_{ig} P(D_i = 1|c_{ig} = 1) \tag{6}$$

3. ESTIMATION

3.1. Maximum likelihood estimators

For a given number of classes G, parameter estimation is achieved using maximum likelihood techniques assuming that missing data are missing at random. The vector of parameters θ contains the transformation parameters (η_{1k}, η_{2k}) for $k = 1, \ldots, K$, the fixed effects $\beta^{\mathrm{T}}, \mu_1^{\mathrm{T}}, \ldots, \mu_G^{\mathrm{T}}, \gamma_1^{\mathrm{T}}, \ldots, \gamma_G^{\mathrm{T}}$ in the mixture model, $\xi_{02}, \ldots, \xi_{0G}, \xi_{12}^{\mathrm{T}}, \ldots, \xi_{1G}^{\mathrm{T}}$ in the probability of belonging to the latent classes, $\delta_{01}, \ldots, \delta_{0G}, \delta_{11}^{\mathrm{T}}, \ldots, \delta_{1G}^{\mathrm{T}}$ in the probability of occurrence of the event and the variance–covariance parameters $\mathrm{vec}(U), \omega_2, \ldots, \omega_G, \sigma_{\alpha_1}, \ldots, \sigma_{\alpha_K}, \sigma_{\varepsilon_1}, \ldots, \sigma_{\varepsilon_K}$ where U is the Cholesky transformation of B, the variance–covariance matrix of the random-effects for the first class and ensures that B is a positive-definite matrix. In the following, we will note $y_i = (y_{i1}^{\mathrm{T}}, \ldots, y_{iK}^{\mathrm{T}})^{\mathrm{T}}$ and $n_i = \sum_{k=1}^{K} n_{ik}$.

In a mixture model, the individual contribution to the likelihood can be decomposed according to the latent classes [15]. Then, using the conditional independence assumption between y_i and D_i given the latent classes, we obtain:

$$\begin{aligned} f(y_i, D_i) &= \sum_{g=1}^{G} P(c_{ig} = 1) f(y_i, D_i|c_{ig} = 1) \\ &= \sum_{g=1}^{G} P(c_{ig} = 1) f(y_i|c_{ig} = 1) P(D_i|c_{ig} = 1) \end{aligned} \tag{7}$$

DOI: 10.1002/sim

The probability of belonging to latent class g, $P(c_{ig}=1)$ is given by the multinomial logit regression (3); the probability of the event given the latent class g, $P(D_i|c_{ig}=1)$ is given by the logistic model (4). The density $f(y_i|c_{ig}=1)$ of the longitudinal outcomes in their natural scale can be computed from the density of the transformed variable $\tilde{y}_i=(\tilde{y}_{i1}^{\mathrm{T}},\ldots,\tilde{y}_{iK}^{\mathrm{T}})^{\mathrm{T}}$ with $\tilde{y}_{ijk}=h_k(y_{ijk})$. Indeed, as $\tilde{y}_i$ follows a linear mixed model given the latent class g, the density of $f(y_i|c_{ig}=1)$ can be written as the product of the multivariate Gaussian density of $\tilde{y}_i|c_{ig}=1$ and the Jacobian of the transformations h_k as in Reference [13]:

$$\begin{aligned} f(y_i|c_{ig}=1;\theta) &= f(\tilde{y}_i|c_{ig}=1;\theta)J(y_i;\theta) \\ &= \phi_g(\tilde{y}_i;\theta)J(y_i;\theta) \end{aligned} \tag{8}$$

where ϕ_g is a multivariate Gaussian density with class-specific mean vector $E_{ig}=(E_{i1g}^{\mathrm{T}},\ldots,E_{ikg}^{\mathrm{T}},\ldots,E_{iKg}^{\mathrm{T}})^{\mathrm{T}}$ and class-specific variance–covariance matrix V_{ig} given by

$$E_{ikg}=Z_i^k\mu_g+X_{1i}^k\beta+X_{2i}^k\gamma_k \tag{9}$$

$$V_{ig}=\begin{pmatrix} Z_i^1 \\ \vdots \\ Z_i^K \end{pmatrix}\omega_g^2 B(Z_i^{1\mathrm{T}} \ \ldots \ Z_i^{K\mathrm{T}})+\begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_K \end{pmatrix} \tag{10}$$

$$\text{with } \Sigma_k=\sigma_{\alpha_k}^2\mathbf{1}_{n_{ik}}\mathbf{1}_{n_{ik}}^{\mathrm{T}}+\sigma_{\varepsilon_k}^2 I_{n_{ik}}$$

where $Z_i^k=(Z(t_{i1k}),\ldots,Z(t_{in_{ik}k}))^{\mathrm{T}}$ is the $n_{ik}\times(p+1)$-matrix of time polynomials for subject i and test k; $X_{1i}^k=(X_{1i}(t_{i1k}),\ldots,X_{1i}(t_{in_{ik}k}))^{\mathrm{T}}$ and $X_{2i}^k=(X_{2i}(t_{i1k}),\ldots,X_{2i}(t_{in_{ik}k}))^{\mathrm{T}}$ are, respectively, the $n_{ik}\times q_1$-matrix and $n_{ik}\times q_2$-matrix of time-dependent covariates with a common effect across the classes or a class-specific effect, and I_n and $\mathbf{1}_n$ are, respectively, the identity matrix of size n and the n-vector of 1's.

The term $J(y;\theta)$ is the Jacobian of the Beta CDF of the longitudinal outcomes y at the parameter vector value θ:

$$J(y_i;\theta)=\prod_{k=1}^{K}\prod_{j=1}^{n_{ik}}\frac{y_{ijk}^{\eta_{1k}-1}(1-y_{ijk})^{\eta_{2k}-1}}{\mathcal{B}(\eta_{1k},\eta_{2k})} \tag{11}$$

where $\mathcal{B}(\eta_{1k},\eta_{2k})$ is the complete Beta function with parameters η_{1k} and η_{2k}.

At last, we obtain the following closed form for the log-likelihood of the joint model:

$$\begin{aligned} L(y;\theta)=\sum_{i=1}^{N}\ln\left(\sum_{g=1}^{G}\frac{e^{\xi_{0g}+X_{3i}^{\mathrm{T}}\xi_{1g}}}{\sum_{j=1}^{G}e^{\xi_{0j}+X_{3i}^{\mathrm{T}}\xi_{1j}}}\times\phi_g(\tilde{y}_i;\theta)\times\frac{[e^{\delta_{0g}+X_{4i}^{\mathrm{T}}\delta_{1g}}]^{D_i}}{1+e^{\delta_{0g}+X_{4i}^{\mathrm{T}}\delta_{1g}}}\right) \\ +\sum_{i=1}^{N}\ln(J(y_i;\theta)) \end{aligned} \tag{12}$$

DOI: 10.1002/sim

In spite of the nonlinear structure of the model, the log-likelihood has a closed-form and can be maximized using standard maximization algorithms as described in next section.

3.2. Optimization algorithm

The log-likelihood (12) is maximized using a modified Marquardt algorithm [16]. This is a Newton–Raphson like algorithm where the diagonal of the Hessian matrix at iteration l, $H^{(l)}$, is inflated to obtain a positive definite matrix: $H^{*(l)} = (H_{ij}^{*(l)})$ with $H_{ii}^{*(l)} = H_{ii}^{(l)} + \lambda[(1-v)|H_{ii}^{(l)}| + v\mathrm{tr}(H)]$ and $H_{ij}^{*(l)} = H_{ij}^{(l)}$ if $i \neq j$. Initial values for λ and v are $\lambda = 0.01$ and $v = 0.01$. They are reduced when H^* is positive definite and increased if not. The estimates $\theta^{(l)}$ are then updated to $\theta^{(l+1)}$ using the current modified Hessian $H^{*(l)}$ and the current gradient of the parameters $g(\theta^{(l)})$ according to the formula:

$$\theta^{(l+1)} = \theta^{(l)} - \kappa H^{*(l)-1} g(\theta^{(l)}) \tag{13}$$

where, if necessary, a linesearch for κ ensures that the log-likelihood is improved at each iteration. The convergence is reached when the three following convergence criteria are satisfied: $\sum_{j=1}^{m} (\theta_j^{(l)} - \theta_j^{(l-1)})^2 \leqslant 10^{-4}$, $|L^{(l)} - L^{(l-1)}| \leqslant 10^{-4}$ and $g(\theta^{(l)})^{\mathrm{T}} H^{(l)-1} g(\theta^{(l)}) \leqslant 10^{-5}$. First and second derivatives are computed by finite differences. Standard errors of the elements of B are computed by the Δ-method while standard errors of the other parameters are directly computed from the inverse of the observed Hessian matrix.

A mixture model is estimated with a fixed number of components G. To choose the optimal number of components, estimated models with different values of G are compared using the Bayesian information criterion (BIC) [17]. Due to the possible multimodality of the likelihood, each model is estimated using several sets of initial values to ensure convergence to the global maximum. Moreover, due to the complexity of the model, simple models are first estimated in order to provide adequate initial values for more complicated models.

3.3. Posterior classification

Posterior classification of the subjects in the different latent classes can be achieved in a mixture model using posterior conditional probabilities. These posterior conditional probabilities $\hat{\pi}_{ig}^{y,D}$ to belong to class g given the observations (y_i, D_i) are computed using the Bayes theorem as follows:

$$\hat{\pi}_{ig}^{y,D} = P(c_{ig} = 1 | y_i, D_i; \hat{\theta}) = \frac{P(c_{ig} = 1) f((y_i, D_i) | c_{ig} = 1; \hat{\theta})}{\sum_{l=1}^{G} P(c_{il} = 1) f((y_i, D_i) | c_{il} = 1; \hat{\theta})} \tag{14}$$

where the denominator is the contribution to the likelihood of subject i at the optimum. Each subject is then classified in the profile or latent class in which he has the largest probability $\hat{\pi}_{ig}^{y,D}$ of belonging.

In our model, information is of two kinds: information from repeated measures of the markers y_i and information from the event occurence D_i. In a prognostic or diagnostic perspective, it could be of interest to consider the posterior probabilities $\hat{\pi}_{ig}^{y}$ of belonging to class g given only the longitudinal observations y_i. These probabilities are computed as follows:

$$\hat{\pi}_{ig}^{y} = P(c_{ig} = 1 | y_i; \hat{\theta}) = \frac{P(c_{ig} = 1) f(y_i | c_{ig} = 1; \hat{\theta})}{\sum_{l=1}^{G} P(c_{il} = 1) f(y_i | c_{il} = 1; \hat{\theta})} \tag{15}$$

DOI: 10.1002/sim

3.4. Diagnostic and prognostic tools

To propose a diagnostic and a prognostic tool based on repeated marker measures, we compute from the estimated parameter vector $\hat{\theta}$ the probability of occurence of the event conditionally on the longitudinal observations as follows:

$$P(D_i = 1|y_i; \hat{\theta}) = \sum_{g=1}^{G} P(D_i = 1|c_{ig} = 1; \hat{\theta}) P(c_{ig} = 1|y_i; \hat{\theta})$$

$$= \sum_{g=1}^{G} \frac{e^{\delta_{0g} + X_{4i}^{T} \delta_{1g}}}{1 + e^{\delta_{0g} + X_{4i}^{T} \delta_{1g}}} \times \hat{\pi}_{ig}^{y} \tag{16}$$

For each threshold between 0 and 1, we classify as demented the subjects with $P(D_i = 1|y_i; \hat{\theta})$ above the threshold and compute sensitivity and specificity of the procedure. The ROC curve may then be drawn and the area under the ROC curve (AUC) may be computed to evaluate the performances of the model in detecting the event.

4. APPLICATION

4.1. The PAQUID cohort study

The data come from the French prospective cohort study PAQUID, initiated in 1988 to study normal and pathological ageing [14]. Subjects included in the cohort were 65 years and older at the initial visit and were followed six times with a visit at one (T1), 3 (T3), 5 (T5), 8 (T8), 10 (T10) and 13 (T13) years after the initial visit (T0). At each visit, a battery of psychometric tests was completed and a two phase screening procedure was carried out for the diagnosis of dementia. Subjects who met DSM-III-R criteria [18] A, B and C (impairment of memory and at least one other cognitive function and interference with daily living) or those presenting a decline of three points or more on the Mini Mental State Examination [19] scale since the previous visit were seen by a neurologist who made the clinical diagnosis. Measurements at the initial visit were excluded from the analysis because of a first passing effect [20].

The aim of the analysis was both to describe the profiles of cognitive evolution up to T13 in a subpopulation of subjects free of dementia until T10, and to propose a diagnostic and a prognostic tool of dementia using repeated measures of psychometric tests. Three psychometric tests were considered ($K = 3$): the Isaacs Set Test (IST) [21] shortened at 15 s which evaluates verbal fluency, the Benton Visual Retention Test (BVRT) [22] which evaluates visual memory and the Digit Symbol Substitution Test of Wechsler (DSSTW) [23] which evaluates attention and logical reasoning. For the three tests, low values indicate a more severe impairment.

The sample included 834 subjects who were visited at T13, were visited and free of dementia at T10, and completed each of the three psychometric tests at least once between T1 and T13. The median number of measures was 4 for the DSSTW and 5 for the IST and the BVRT. Among the 834 subjects, 114 had a positive diagnosis of dementia at T13-visit. The time variable was the exact time in decade between the age at the T13-visit and the age at the current visit. Thus it ranged from about -1.2 to 0. The polynomial function of time chosen for the evolution of the latent process included a random intercept, a random slope and a random quadratic slope. The

DOI: 10.1002/sim

Table I. Likelihood, BIC and AUC of the adjusted model for various numbers of latent classes.

Number of classes (G)	Number of parameters	Log-likelihood	BIC	AUC
1	26	−29337.87	58850.61	
2	33	−29227.74	58677.44	0.855
3	40	−29216.26	58701.57	0.860
4	47	−29206.93	58730.00	0.862

covariates included in the model were educational level (graduated from primary school *versus* not graduated), gender and a dichotomous variable for age with a cut-off at the median (less than 83 years old at T13-visit *versus* more than 83 years old at T13-visit). The three covariates were tested in the three parts of the model: in the longitudinal model either as a common effect across classes (X_1) or as a class-specific effect (X_2), in the model for the probability of class membership (X_3) and in the model for the probability of occurence of the event (X_4).

4.2. The nonlinear latent class model

The model adjusted on the three covariates was estimated for a number of latent classes G varying from 1 to 4. Whatever the number of classes, the best adjustment on covariates included age, educational level and gender in association with the mean latent cognitive level but not with the time variables. These associations with the trajectories were common across the classes. Educational level was found associated with the probability of dementia at T13 after adjustment on the latent classes, and for the model with 2, 3 and 4 classes, age was also associated with the probability of latent class membership.

Using the BIC [17] for selecting the model with the optimal number of latent classes, the model with two classes was retained (see Table I). Fixed-effect estimates are presented in Table II and the two predicted evolutions of the latent process ($Z(t)\hat{\mu}_g$, $g = 1, 2$) are displayed in Figure 1.

The first class trajectory is quite linear with a slight decline until T13-visit. As seen in Table II, the associated probability of dementia is almost null ($\exp(-2.815)/(1 + \exp(-2.815)) = 0.057$). This trajectory seems to represent the mean cognitive evolution in normal ageing. In contrast, the mean cognitive level in the second class is always lower than in the first class and the decline is sharper and nonlinear. The associated probability of dementia is very high ($\exp(2.715)/(1 + \exp(2.715)) = 0.938$). Thus this trajectory seems to represent the mean decline in a pre-diagnostic phase of dementia. In the following, we will call class 1 'slight decline' class and class 2 'marked decline' class.

Whatever the class trajectory, subjects who graduated from primary school had a significantly better mean cognitive level, subjects older than 83 years had a significantly lower cognitive level than younger subjects and men had a significantly higher mean cognitive level. After adjustment on the latent classes, educational level was found associated with the probability of dementia at T13: adjusted on the shape of the cognitive decline, subjects not graduated from primary school had a higher risk of dementia at T13 ($\mathrm{OR} = 5.464$, $\mathrm{CI}_{95\ \text{per cent}} = [1.107; 26.967]$) than subjects graduated from primary school. This effect of education on the risk of dementia was not significantly different according to the classes. Finally, age was found to be associated with the class membership probability. The odds-ratio for being in the marked decline class rather than in

DOI: 10.1002/sim

Table II. Estimations of the fixed effects in the final nonlinear mixture model with two latent classes.

Parameter	Estimate	Standard-error
Class 1		
Intercept	0.511	0.015
Linear slope	−0.0484	0.0063
Quadratic slope	0.0100	0.0048
Class 2		
Intercept	0.376	0.021
Linear slope	−0.259	0.024
Quadratic slope	−0.106	0.017
Covariates		
Gender*	0.0121	0.0055
Educational level†	0.108	0.008
Age‡	−0.0364	0.0057
Probability of dementia at T13		
Intercept class 1 (δ_{01})	−2.815	0.453
Intercept class 2 (δ_{02})	2.715	0.916
Educational level†	−1.698	0.815
Probability of belonging to class 2		
Intercept	−2.188	0.208
Age‡	0.889	0.231

*Reference: female.
†Reference: no diploma.
‡Reference: younger than 83 years old at T13-visit.

the slight decline class was 2.433 ($CI_{95\ per\ cent} = [1.547; 3.826]$) for subjects above 83 years old compared to younger people. Older subjects were more prone to experiment both a sharp cognitive decline and a dementia.

The class-specific scale parameter in the variance–covariance matrix for the random-effects was estimated at $\omega_2 = 1.056$ (SE = 0.105). Thus, the variability was not significantly different for the two classes.

4.3. Posterior classification

According to the posterior class-membership probabilities given all the observed data $\hat{\pi}_{ig}^{y,D}$, 719 subjects were classified in the slight decline class including 708 subjects without dementia at T13 and 115 subjects were classified in the marked decline class including 103 subjects diagnosed demented at T13. This almost perfect discrimination underlines that we were able to discriminate with this model the evolution of pre-demented and normal subjects. Nevertheless, a good discrimination was expected since information about dementia was used to compute the posterior probabilities.

Thus, classification obtained using only information from the longitudinal outcomes, that is to say $\hat{\pi}_{ig}^{y}$, is displayed in Table III: 749 subjects were classified in the slight decline class including 685 subjects without dementia at T13 and 85 subjects were classified in the marked decline class

DOI: 10.1002/sim

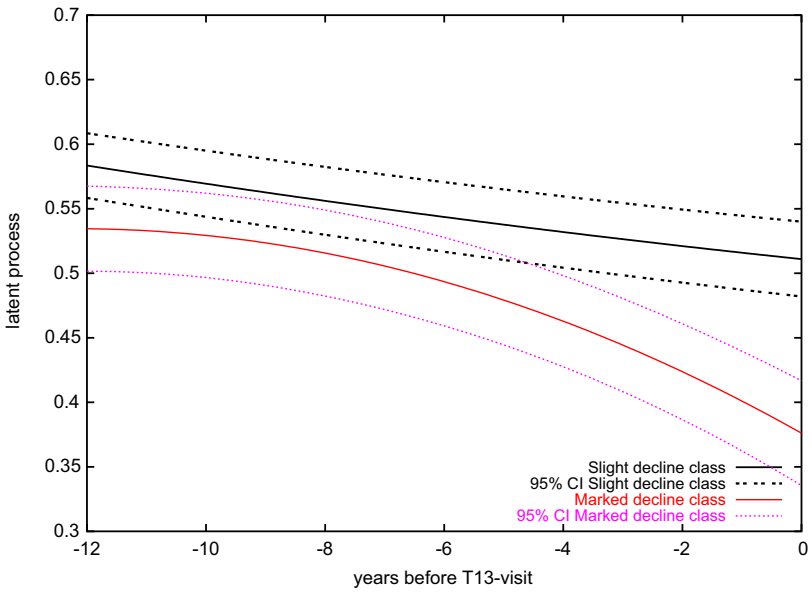

Figure 1. Predicted trajectories for the two classes of the adjusted nonlinear latent class model for a female without diploma and younger than 83 years old at T13 (plain lines = mean predicted evolution; dashed lines = 95 per cent confidence bands).

Table III. Posterior classification from $\hat{\pi}_{ig}^{y}$ and dementia diagnosis at T13.

	Classification		
Dementia diagnosis at T13	Slight decline class	Marked decline class	Total
Positive	64	50	114
Negative	685	35	720
Total	749	85	834

including 50 subjects with a positive diagnosis of dementia at T13. As expected with the significant effect of age on class membership, we found that subjects classified in the marked decline class were older than those classified in the slight decline class (67 per cent were older than 83 years at the end of the follow-up *versus* 44 per cent in the slight decline class). In contrast, the two posterior classes did not differ according to gender (42 per cent of men in the marked decline class *versus* 38 per cent in the slight decline class) and educational level (78 per cent with a high educational level *versus* 81 per cent).

According to this classification, we also computed in Table IV the means of the posterior probabilities of belonging to each class l over the n_g subjects classified in class g $((1/n_g)\sum_{i=1}^{n_g}\hat{\pi}_{il}^{y}$, $(g,l) \in \{1,2\}^2)$, high diagonal terms indicating a good classification quality [24]. The mean of the posterior probabilities to be in the slight decline class was 0.920 for subjects classified in this class showing a very high discrimination and an unambiguous class affectation. In contrast, the mean posterior probability to be in the marked decline class was 0.753 for subjects classified

DOI: 10.1002/sim

Table IV. Posterior classification table: mean of the posterior probabilities of belonging to each class l over the n_g subjects classified in class g ($(1/n_g)\sum_{i=1}^{n_g} \hat{\pi}_{il}^{y}$, $(g, l) \in \{1, 2\}^2$).

Class g\class l	Slight decline class	Marked decline class
Slight decline class	0.920	0.080
Marked decline class	0.247	0.753

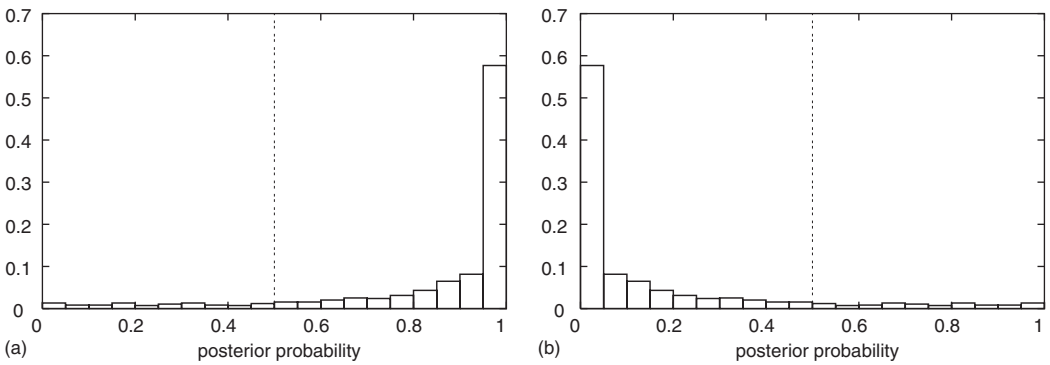

Figure 2. Empirical distribution of the posterior probabilities of being in the slight decline class (a) and in the marked decline class (b).

in this class revealing a more ambiguous affectation. Indeed, subjects classified in the marked decline class had a non-negligible mean posterior probability of being in the slight decline class ($(1/n_2)\sum_{i=1}^{n_2} \hat{\pi}_{i1}^{y} = 0.247$). The histogram of the posterior probabilities displayed in Figure 2 emphasizes the more ambiguous affectation to the marked decline class.

4.4. Adequation of the model

To evaluate the fit of the data, we compared for each test and each class, the weighted mean trajectory of the observed scores with the weighted mean trajectory of the predicted values. For each observation, that is each triplet (i, j, k), the marginal predicted value in the natural scale of the outcome, that is the estimate of $E(h_k^{-1}(\tilde{y}_{ijk})|c_{ig} = 1)$ was obtained using a numerical integration of $h_k^{-1}(\tilde{y}_{ijk})$ using the estimated Gaussian density $\phi_g(\tilde{y}_{ijk}; \hat{\theta})$ defined by (9) and (10). Then, at each visit, the mean of the available marginal predictions weighted by the posterior probabilities was computed and compared with the corresponding weighted mean of the observed values. The results are displayed in Figure 3. For the two classes, the fit is good whatever the psychometric test indicating a good adequacy of the model.

4.5. Diagnosis of dementia

To evaluate the performances of the model to diagnose dementia using only information about the psychometric tests, we computed the posterior probabilities of dementia $P(D_i = 1|y_i; \hat{\theta})$ using (16) and built the ROC curve represented in Figure 4. The area under the curve was 0.855. We

DOI: 10.1002/sim

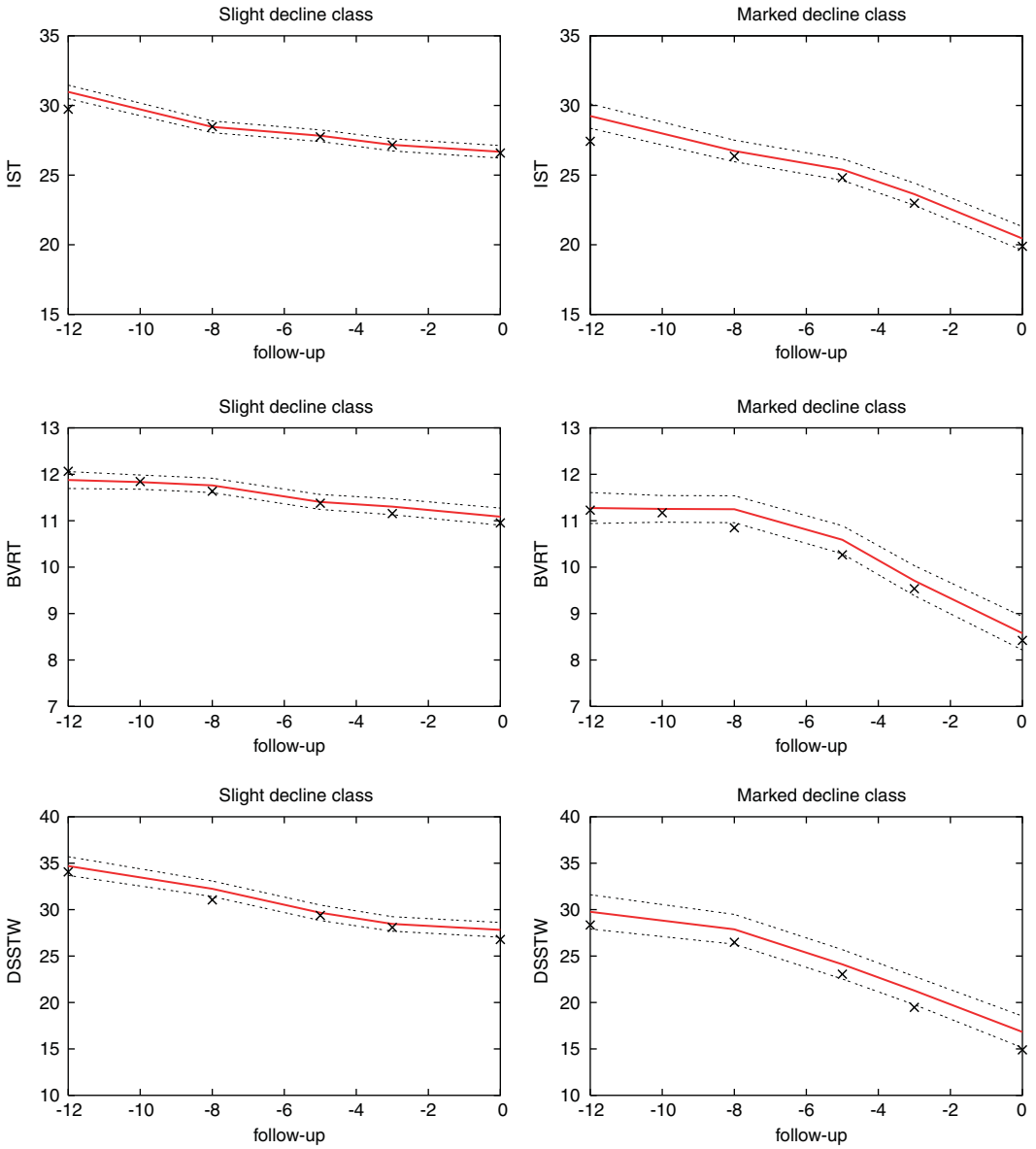

Figure 3. Marginal predicted weighted mean evolution (crosses) *versus* observed weighted mean evolution (plain line) with its 95 per cent confidence bands (dashed lines) for each class and each test in their natural scale according to the visits.

also performed a leave-one-out procedure for computing the area under the curve. It consisted in leaving iteratively each subject i out, estimating the parameters $\hat{\theta}_{(-i)}$ without i and computing $P(D_i = 1|y_i; \hat{\theta}_{(-i)})$. This procedure lead to a very close AUC (0.852).

DOI: 10.1002/sim

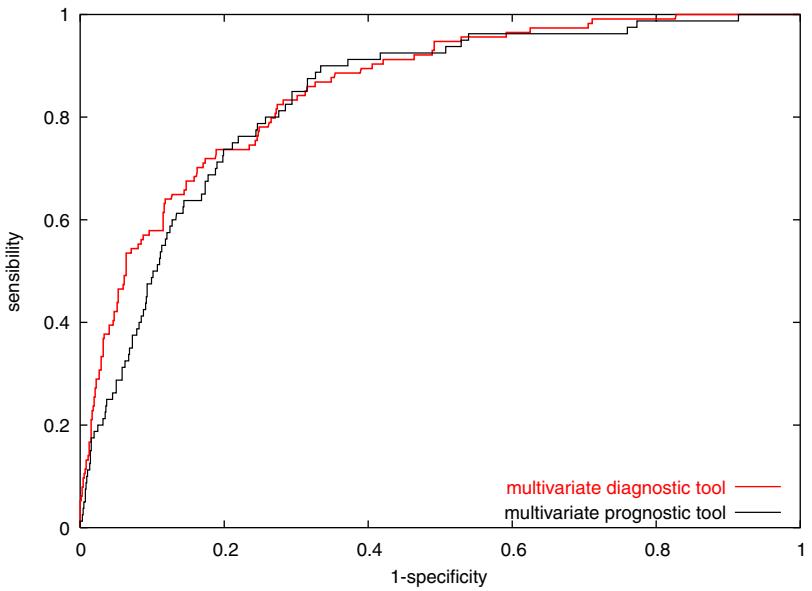

Figure 4. ROC curve of the posterior probabilities of dementia from the multivariate nonlinear latent class model for the diagnosis analysis ($N = 834$) in grey plain line and for the prognosis analysis ($N = 1059$) in black plain line.

One advantage of this diagnostic tool for dementia is that it can be used whatever the number and the times of measurements of the cognitive tests in the past 12 years and even if the subject has measurements for only one test. We compared the diagnostic ability of the estimated multivariate model (Table II) according to the number of psychometric tests used for the computation of the posterior probabilities of dementia.

Table V clearly shows that the AUC increases with the number of psychometric tests used for computing the posterior probability of dementia, underlining the interest of a flexible model which can handle multivariate data.

We compared the performance of our model with a simpler method for diagnosing dementia according to cognitive ability, that is a logistic regression including six explanatory variables: age, educational level, gender and cognitive test scores for the IST, the BVRT and the DSSTW at the last visit T13. As the logistic regression requires no missing data for all the explanatory variables included in the model, the model was estimated on a reduced sample of 647 subjects having a measure of each test at the T13-visit. On this reduced sample, the AUC from the logistic regression was 0.852 which is nearly as good as for our multivariate longitudinal model. However, 187 subjects (22 per cent of the sample) with incomplete measures were excluded from this analysis including a high proportion of demented subjects (58 demented were excluded among the 114 demented of the initial sample). By using longitudinal and multivariate cognitive assessment, our model allows to compute a probability of dementia whatever the available information on the three tests unlike the logistic regression. It is of particular interest in epidemiological studies in which subjects in pre-dementia phase do not often complete all the tests.

DOI: 10.1002/sim

Table V. Areas under the curve (AUC) obtained from the multivariate nonlinear mixture model in a diagnosis analysis (probabilities of dementia at T13 computed using cognitive measures between T1 and T13) or in a prognosis analysis (probabilities of dementia at T10 computed using cognitive measures between T1 and T8).

Psychometric tests used to compute probabilities of dementia	AUC (diagnostic at T13)	AUC (prognostic at T10 given T1–T8)
IST	0.775	0.840
BVRT	0.796	0.788
DSSTW	0.808	0.802
IST + BVRT	0.831	0.841
IST + DSSTW	0.831	0.832
BVRT + DSSTW	0.842	0.814
IST + BVRT + DSSTW	0.855	0.837

4.6. Prognosis of dementia

Even if the model was built in a diagnostic perspective, it can be used to compute probability of dementia at time T knowing at least one measure from IST, BVRT or DSSTW in the last 12 years. Thus, it can be used as a prognostic tool. We explored the ability of our model to predict dementia 2 years after the last cognitive measurements. More specifically, we tried to predict dementia at T10 using cognitive measurements between T1 and T8 in the PAQUID study. In this way, for each of the 1059 subjects with at least 1 measure at each test between T1 and T8 and with a negative diagnosis of dementia at T8, we computed the probability of dementia at T10 knowing measures collected between T1 and T8 and using the vector of parameters estimated on the previous sample (Table II). As for the previous analysis on diagnosis of dementia, we compared the results according to the number of psychometric tests used to compute the probability of dementia. We imposed that each subject had at least one measure at each test to compare the AUC according to the number of tests on samples of the same size. The results are presented in Table V.

Whatever the number of psychometric tests used for computing the probability of dementia at 2 years, the AUC was quite high. The ROC curve when using the three psychometric tests is displayed in Figure 4. For the BVRT and the DSSTW, we found again that the AUC for the prognosis increased when using more information. However, in this sample, the prognostic ability of IST was very good and did not increase when other tests were added.

5. DISCUSSION

We proposed a nonlinear latent class model for multivariate longitudinal outcomes and a binary variable. In ageing context, by modelling jointly cognitive trajectory using three psychometric tests and dementia diagnosis, we distinguished the cognitive decline in a normal cognitive ageing process and the cognitive decline in a pre-diagnosis phase of dementia. From the estimates of the model, we proposed a diagnostic and prognostic tool and compared, according to this tool, the ability of our model to predict dementia when varying the quantity of available information on cognitive ability. Thus, we underlined that using several cognitive tests increased the power of the diagnostic and prognostic tool.

DOI: 10.1002/sim

A key aspect of the flexibility of our model lies in the nonlinear transformations which allow to model continuous outcomes with a distribution far from a Gaussian distribution. Beta CDFs were chosen because firstly, they are flexible functions with only two parameters per marker and secondly, they gave a very good fit of the cognitive test scores in our elderly sample as it was previously shown [13]. The interest of these estimated transformations was extensively discussed in Reference [13] showing in particular that they give interesting results about the metrological properties of each psychometric test.

In a perspective of proposing diagnostic and prognostic tools, one could wonder whether it is sensible to select the optimal number of classes using the BIC criterion which tends to retain a parsimonious model. One could prefer selecting the optimal number of classes using a less conservative criterion like the AIC for example, or stopping the process when the AUC of the model is not significantly improved. In our application on dementia diagnosis, the AUC from the three and four class model were just a little improved so we retained only two latent classes of evolution, the first one representing an almost normal cognitive ageing process and the second one the cognitive decline in a pre-diagnosis phase of dementia. By stratifying on dementia status and estimating the two separated multivariate models, we could have probably obtained quite the same evolutions but we would not have been able to compute a diagnostic tool of dementia.

Our joint model is an extension of the classic mixture models [25, 26]. We no longer study the main heterogeneity in the population, we rather investigate the unobserved subpopulation structure in association with the clinical event. Such joint models including a mixture model for longitudinal data and a logistic regression had already been proposed by Lin *et al.* [5] for studying one longitudinal outcome with a Gaussian assumption. They had then improved their model by replacing the logistic regression by a survival model [1]. In this paper, we focused on a multivariate approach allowing flexible relationships between the observed markers and the latent process. We only considered a logistic regression for the clinical event but we showed that even without a time-to-event model, we were able to propose a prognostic tool in addition to the diagnostic tool. However, as a model for a binary event does not account for time-to-event, parameters must be estimated on a selected sample for which the dementia status is known at a given time. To avoid this selection in the estimation step and include cases where the event arises at any time, the model should be next extended for including a proportional hazard model for the time-to-event instead of the logistic regression.

REFERENCES

1. Lin H, Turnbull BW, Mc Culloch CE, Slate EH. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* 2002; **97**:53–65.
2. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**:465–480.
3. Whulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**:330–339.
4. Jacqmin-Gadda H, Commenges D, Dartigues J-F. Random changepoint model for joint modeling of cognitive decline and dementia. *Biometrics* 2006; **62**:254–260.
5. Lin H, Mc Culloch CE, Turnbull BW, Slate EH, Clark LC. A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine* 2000; **19**:1303–1318.
6. Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 1999; **50**:463–469.
7. Thiébaut R, Jacqmin-Gadda H, Babiker A, Commenges D. The CASCADE collaboration. Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+

DOI: 10.1002/sim

cell count and HIV RNA viral load in response to treatment of HIV infection. *Statistics in Medicine* 2005; **24**:65–82.
8. Xu J, Zeger SL. The evaluation of multiple surrogate endpoints. *Biometrics* 2001; **57**:81–87.
9. Roy J, Lin X. Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* 2000; **56**:1047–1054.
10. Rabe–Hesketh S, Skrondal A, Pickles A. Generalized multilevel structural equation modeling. *Psychometrika* 2004; **69**:167–190.
11. Ganiayre J, Commenges D, Letenneur L. A latent process model for dementia and psychometric tests. *Technical Report*, 2006.
12. Hashemi R, Jacqmin-Gadda H, Commenges D. A latent process model for joint modeling of events and marker. *Lifetime Data Analysis* 2003; **9**:331–343.
13. Proust C, Jacqmin-Gadda H, Taylor JMG, Ganiayre J, Commenges D. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics* 2006, online publication on 26 April 2006. doi: 10.1111/j.1541-0420.2006.00573.x
14. Letenneur L, Commenges D, Dartigues J-F, Barberger-Gateau P. Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. *International Journal of Epidemiology* 1994; **23**:1256–1261.
15. Proust C, Jacqmin-Gadda H. Estimation of linear mixed models with a mixture of distribution for the random-effects. *Computer Methods and Programs in Biomedicine* 2005; **78**:165–173.
16. Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics* 1963; **11**:431–441.
17. Schwartz G. Estimating the dimension of a model. *The Annals of Statistics* 1978; **6**:461–464.
18. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (DSM-III-R.3rd edition revised). Washington, DC, 1987.
19. Folstein MF, Folstein SE, McHugh PR. Mini Mental State. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975; **12**:189–198.
20. Jacqmin-Gadda H, Fabrigoule C, Commenges D, Dartigues J-F. A 5-year longitudinal study of the Mini Mental State Examination in normal aging. *American Journal of Epidemiology* 1997; **145**:498–506.
21. Isaacs B, Kennie AT. The Set Test as an aid to the detection of dementia in old people. *British Journal of Psychiatry* 1973; **123**:467–470.
22. Benton AL. *Manuel pour l'application du Test de Rétention Visuelle. Applications cliniques et expérimentales* (2ème édition française). Centre de Psychologie appliquée, Paris, 1965.
23. Wechsler D. *WAIS–R Manual*. Psychological Corporation: New York, 1981.
24. Muthén B, Brown CH, Masyn K, Jo B, Khoo S-T, Yang C-C, Wang C-P, Kellam SG, Carlin JB, Liao J. General growth mixture modeling for randomized preventive intervention. *Biostatistics* 2002; **3**:459–475.
25. Legler JM, Davis WW, Potosky AL, Hoffman RM. Latent variable modelling of recovery trajectories: sexual function following radical prostatectomy. *Statistics in Medicine* 2004; **23**:2875–2893.
26. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data* (Chapter 12). Springer Series in Statistics, 2000.

DOI: 10.1002/sim