



# Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach

Cécile Proust-Lima, Pierre Joly, Jean-François Dartigues, Hélène Jacqmin-Gadda

## ► To cite this version:

Cécile Proust-Lima, Pierre Joly, Jean-François Dartigues, Hélène Jacqmin-Gadda. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. Computational Statistics and Data Analysis, 2009, 53 (4), pp.1142-1154. inserm-00328649

**HAL Id: inserm-00328649**

**<https://inserm.hal.science/inserm-00328649>**

Submitted on 17 May 2010

**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.

# Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach

Cécile Proust-Lima<sup>a,b,\*</sup> Pierre Joly<sup>a,b</sup>

Jean-François Dartigues<sup>b,c</sup> Hélène Jacqmin-Gadda<sup>a,b</sup>

<sup>a</sup>*INSERM, U897, Biostatistics department, Bordeaux, F-33076, France.*

<sup>b</sup>*Université Victor Segalen, Bordeaux, F-33076, France.*

<sup>c</sup>*INSERM, U897, Aging department, Bordeaux, F-33076, France.*

---

## Abstract

A joint model based on a latent class approach is proposed to explore the association between correlated longitudinal quantitative markers and a time-to-event. A longitudinal latent class model describes latent profiles of evolution of the latent process underlying the correlated markers. The latent process is linked to the markers by nonlinear transformations including parameters to be estimated. A proportional hazard model describes the joint risk of event according to the latent classes and two specifications of the risk function are considered: a parametric function and a semi-parametric function based on splines. Depending on the chosen risk function, estimation is performed by a maximum likelihood or a maximum penalized likelihood approach. A simulation study validates the estimation procedure. As a latent class model relies on the strong assumption that the markers and the time-to-event are independent conditionally on the latent classes, a test of conditional independence is proposed using the residuals conditional on time-to-event. The procedure

does not require any posterior classification and can be conducted using standard statistical softwares. The methodology is applied to describe profiles of cognitive decline in the elderly and their associated risk of dementia.

*Key words:* Cognitive aging; conditional independence; joint analysis; latent class model; mixed model; multiple outcomes.

---

## 1 Introduction

When collecting longitudinal markers of a chronic disease and time to a clinical event, the joint modelling of the two quantities represents the best way to study their association. A joint model for longitudinal markers and a time-to-event consists in describing the marker evolution using a model for longitudinal data as the linear mixed model (Laird and Ware (1982)) and the risk of event using a survival model. Based on this principle, two kinds of joint models were proposed: the shared random effect model (Henderson et al. (2000)) and the latent class model (Lin et al. (2000)). A shared random effect model consists in including the random effects of the mixed model as covariates in the model for the event. Thus, it offers a flexible way to model the correlation between the marker evolution and the risk of event. However, the estimation procedure can be numerically intensive since a numerical integration is needed in the computation of the joint log-likelihood. Second, the shared random effects make the interpretation of the dependency difficult in terms of correlation between the two quantities. Finally the mixed model used to model the marker evolution is assumed homogeneous with a Gaussian distribution for the random-effects

---

\* INSERM U897, ISPED, Université Bordeaux II, Case 11, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France. Ph: (33) 5 57 57 45 79. Fax: (33) 5 56 24 00 81  
*Email address:* Cecile.Proust@isped.u-bordeaux2.fr (Cécile Proust-Lima ).

whereas one could assume that the marker evolution is heterogeneous according to the event occurrence. To avoid these limitations, a second way to jointly model the longitudinal marker and the time-to-event is the latent class model. Based on the assumption that the population is constituted of several latent sub-populations in which the evolution of the marker and the risk of the event are different, this approach makes direct use of the heterogeneous nature of the data. It offers an intuitive way of describing the dependency between the marker and the event by distinguishing profiles of marker evolution associated with the risk of event. Another interesting property is that the correlation between the repeated measures of the markers and the link between the markers and the time-to-event are modelled separately using quantitative random-effects for the former and latent classes for the latter. At last, in contrast with the shared random-effect model, as the association between the marker and the event is managed by a discrete latent variable, the joint log-likelihood does not involve any numerical integration on the shared parameters. In spite of these assets, the latent class approach has only been proposed in few studies for jointly studying a marker evolution and an event. Muthén and Shedden (1999) explored the association between the shape of heavy drinking trajectory in the 18-30-year age range and the probability of alcohol dependency at age 30 while Lin et al. (2000) and Lin et al. (2002) developed latent class models for investigating the association between the evolution of the PSA, a biomarker of the Prostate Cancer, and the incidence of a Prostate cancer.

When studying the evolution of a chronic disease, an additional problem is that several markers of the same underlying quantity are frequently collected. For example, when studying the profiles of cognitive ageing in the elderly, cognition is not directly observed but is measured by several non Gaussian markers, the psychometric tests. Roy and Lin (2000) proposed a mixed model

with a latent process to account for the multiplicity of the markers by describing their common factor evolution but they restricted their work to Gaussian markers and linear dependencies between markers and latent process. Proust et al. (2006) introduced flexible nonlinear links between the markers and the latent process in the previous model to account for possibly non Gaussian continuous markers. Based on this nonlinear model for multivariate longitudinal data, Proust-Lima et al. (2007) developed a joint latent class model for multivariate longitudinal outcomes and an event by including a logistic model for the risk of event in order to propose a diagnostic tool of dementia based on repeated measures of several psychometric tests. However, by using a logistic regression for the event, the estimates of the joint model and the diagnostic tool could be biased as they were obtained on a selected sample of subjects with the event status known at a given time.

In this context, the aim of this paper is to propose a joint latent class model for investigating the association between the common factor of several longitudinal markers and a time-to-event using a proportional hazard model. We propose two estimation procedures, a parametric maximum likelihood approach and a semi-parametric maximum penalized likelihood approach. A central hypothesis of the latent class approach is the conditional independence between the markers evolution and the risk of event given the latent classes. In the literature, it was evaluated using methods based on the Bandeen-Roche et al. (1997) strategy of posterior individual affectation to the latent classes (Lin et al. (2002) ; Lin et al. (2004)). We propose an alternative approach using conditional residuals inspired by Dobson and Henderson (2003) that does not require any posterior classification. Next section presents the latent class model specification. Estimation procedures based on the log-likelihood or the penalized log-likelihood are detailed in section three. Section four describes a

simulation study which validated the estimation process. We focus in section five on the test of conditional independence. This work being mainly motivated by the study of cognitive ageing and occurrence of dementia, section six is dedicated to an application on data from the prospective cohort of aging PAQUID (Letenneur et al. (1994)). Finally we conclude in section seven.

## 2 Statistical model

### 2.1 Heterogeneous population

We consider a population of  $N$  subjects that can be divided in  $G$  unobserved sub-populations represented by latent classes. For each subject  $i$ ,  $i = 1, \dots, N$ , the latent class membership is managed by a latent variable  $c_i$  that equals  $g$  if  $i$  belongs to class  $g$  ( $g = 1, \dots, G$ ). The individual probability of belonging to class  $g$  is explained using covariates  $X_{1i}$  in a multinomial logistic regression:

$$\pi_{ig} = P(c_i = g | X_{1i}) = \frac{e^{\xi_{0g} + X_{1i}^T \xi_{1g}}}{\sum_{l=1}^G e^{\xi_{0l} + X_{1i}^T \xi_{1l}}}, \quad (2.1)$$

where  $\xi_{0g}$  is the intercept for class  $g$  and  $\xi_{1g}$  is the  $q_1$ -vector of class-specific parameters associated with the  $q_1$ -vector of time-independent covariates  $X_{i1}$ . For identifiability,  $\xi_{01} = 0$  and  $\xi_{11} = 0$ . Latent class  $g$  is then characterized by a specific evolution of the markers and a specific risk of event.

### 2.2 Markers evolution

In each latent class, the markers evolution follows the latent process model for multivariate longitudinal data proposed by Proust et al. (2006). Each longi-

tudinal marker is assumed to be a specific transformation of a noisy measure of the same latent process which is defined in continuous time  $(\Lambda_i(t))_{t \geq 0}$  and represents their common factor. The latent process evolution is described by a linear mixed model (Laird and Ware (1982)) specific to each class  $g$ :

$$\Lambda_i(t) \mid_{c_i=g} = Z(t)^T u_{ig} + X_{2i}(t)^T \beta, \quad t \geq 0, \quad (2.2)$$

where  $Z(t)$  is the  $q$ -vector of time-dependent covariates associated with the  $q$ -vector of random-effects  $u_{ig}$ ;  $Z(t)$  can typically include a polynomial function of time. The vector of random-effects  $u_{ig}$  is normally distributed according to  $\mathcal{N}(\mu_g, \omega_g^2 B)$  where  $\omega_1 = 1$  and  $B$  is an unstructured  $q \times q$ -matrix. Thus, the mean vector and the variance of the random-effects may be different for each latent class. The  $q_2$ -vector of covariates  $X_{2i}(t)$  is associated with the  $q_2$ -vector of fixed effects  $\beta$  assumed common across the classes to simplify notations. However, a class-specific vector of regression coefficients could also be considered without additional difficulty.

For each subject  $i$  ( $i = 1, \dots, N$ ), and each marker  $k$  ( $k = 1, \dots, K$ ),  $n_{ik}$  repeated measures of the marker are collected. At each occasion  $j$  ( $j = 1, \dots, n_{ik}$ ), the measure of the marker  $y_{ijk}$  at time  $t_{ijk}$  is linked to the latent process value  $\Lambda_i(t_{ijk})$  through a flexible nonlinear measurement model:

$$\tilde{y}_{ijk} = h_k(y_{ijk}; \eta_k) = \Lambda_i(t_{ijk}) + \alpha_{ik} + \epsilon_{ijk}, \quad (2.3)$$

where  $h_k$  is a Beta cumulative density function with parameters  $\eta_k = (\eta_{1k}, \eta_{2k})$ :

$$h_k(y; \eta_k) = \frac{\int_0^y x^{\eta_{1k}-1} (1-x)^{\eta_{2k}-1} dx}{\int_0^1 u^{\eta_{1k}-1} (1-u)^{\eta_{2k}-1} du}. \quad (2.4)$$

The Beta transformations are chosen because they are parsimonious and offer very flexible shapes (Proust et al. (2006)). These transformations required a preliminary rescaling of the markers in  $[0, 1]$ . The subject-and-marker specific random intercept  $\alpha_{ik}$  is defined according to the Gaussian distribution  $\mathcal{N}(0, \sigma_{\alpha_k}^2)$  and the independent errors  $\epsilon_{ijk}$  are defined according to the Gaussian distribution  $\mathcal{N}(0, \sigma_{\epsilon_k}^2)$ . The subject-and-marker specific random intercept represents the inter-individual variability of the ability to perform each test among individuals with the same latent cognitive level.

### 2.3 Survival model

Let define  $(T_i, D_i)$  where  $T_i$  is the minimum between  $T_i^*$  the time of event and  $C_i$  the time of censoring. The indicator of event  $D_i$  equals 1 if  $T_i^* \leq C_i$  and 0 if  $C_i < T_i^*$ . The risk of event in latent class  $g$  is described by a proportional hazard model:

$$\lambda(t \mid c_i = g, X_{3i}; \zeta_g, \delta_g) = \lambda_{0g}(t; \zeta_g) e^{X_{3i} \delta_g}, \quad (2.5)$$

where  $X_{3i}$  is the  $q_3$ -vector of covariates associated with the  $q_3$ -vector of parameters  $\delta_g$  which can be specific to the latent classes, and  $\lambda_{0g}(t; \zeta_g)$  is the baseline risk of event in latent class  $g$  parameterized by  $\zeta_g$ . A simplified model may also be used assuming proportionality of the risks between classes,  $\lambda_{0g}(t) = \lambda_0(t) e^{\delta_{0g}}$ .

We considered two specifications of the baseline risk function: a parametric model using for instance a Weibull or a piecewise constant baseline hazard, or a semi-parametric model through a splines basis survival model. Splines are piecewise polynomial functions that are linearly combined to define a func-

tion on an interval. For the semi-parametric survival model, we used cubic  $M$ -splines and their integrated version,  $I$ -splines (Ramsay (1988) ; Joly et al. (1999)). Thus, in latent class  $g$ , for a sequence of  $m$  knots, baseline risk function  $\lambda_{0g}(\cdot)$  was defined by a linear combination of  $m + 2$  cubic  $M$ -splines, and survival function  $S_{0g}(\cdot)$  was defined according to  $m + 2$   $I$ -splines:

$$\lambda_{0g}(\cdot) = \sum_{l=1}^{m+2} \zeta_{lg} M_l \quad \text{and} \quad S_{0g}(\cdot) = e^{-\sum_{l=1}^{m+2} \zeta_{lg} I_l} \quad \text{where} \quad \zeta_{lg} \geq 0. \quad (2.6)$$

where  $\zeta_g$  is the vector of splines coefficients to estimate.

### 3 Estimation

#### 3.1 Parametric log-likelihood

Let note  $\theta$  the entire vector of parameters. Parameters estimation is achieved by a maximum likelihood method for a known number of latent classes  $G$ . Using the conditional independence assumption between the markers evolution and the risk of event given the latent classes, the individual contribution of subject  $i$  to the likelihood is:

$$L_i(\theta) = \sum_{g=1}^G P(c_i = g; \theta) f(y_i | c_i = g; \theta) \lambda(T_i | c_i = g; \theta)^{D_i} S(T_i | c_i = g; \theta), \quad (3.1)$$

where  $P(c_i = g; \theta) = \pi_{ig}$  is defined in (2.1) according to the multinomial logistic regression and  $y_i = (y_{i11}, \dots, y_{ijk}, \dots, y_{in_K K})^T$ . The density of the longitudinal markers in latent class  $g$ ,  $f(y_i | c_i = g; \theta)$  is the product of the multivariate normal density  $\phi_g(\tilde{y}_i; \theta)$  of the transformed variable  $\tilde{y}_i$  and the Jacobian of the

nonlinear transformations of the markers  $J(y_i; \theta)$  (Proust et al. (2006)):

$$\begin{aligned} f(y_i \mid c_{ig} = 1; \theta) &= f(\tilde{y}_i \mid c_{ig=1}; \theta) J(y_i; \theta) \\ &= \phi_g(\tilde{y}_i; \theta) J(y_i; \theta). \end{aligned} \quad (3.2)$$

Indeed, for each latent class  $g$ , the transformed variable vector  $\tilde{y}_i$  follows a multivariate normal distribution with mean vector  $E_{ig} = (E_{i1g}^T, \dots, E_{ikg}^T, \dots, E_{iKg}^T)^T$  and variance-covariance matrix  $V_{ig}$  defined as follow:

$$E_{ikg} = Z_i^k \mu_g + X_{2i}^k \beta \quad (3.3)$$

$$\text{and } V_{ig} = \begin{pmatrix} Z_i^1 \\ \vdots \\ Z_i^K \end{pmatrix} \omega_g^2 B \begin{pmatrix} Z_i^{1T} & \dots & Z_i^{KT} \end{pmatrix} + \begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_K \end{pmatrix}, \quad (3.4)$$

where  $\Sigma_k = \sigma_{\alpha_k}^2 \mathbf{1}_{n_{ik}} \mathbf{1}_{n_{ik}}^T + \sigma_{\epsilon_k}^2 I_{n_{ik}}$ ;  $Z_i^k = (Z(t_{i1k}), \dots, Z(t_{in_{ik}k}))^T$  is the  $n_{ik} \times (p+1)$ -matrix of time polynomials for subject  $i$  and test  $k$ ;  $X_{2i}^k = (X_{2i}(t_{i1k}), \dots, X_{2i}(t_{in_{ik}k}))^T$  is the  $n_{ik} \times q_2$ -matrix of time-dependent covariates with a common effect across classes.  $I_n$  and  $\mathbf{1}_n$  are respectively the identity matrix of size  $n$  and the  $n$ -vector of 1s. From these results, the log-likelihood  $\mathcal{L}(\theta)$  of the joint model is:

$$\mathcal{L}(\theta) = \sum_{i=1}^N \ln \left( \sum_{g=1}^G \pi_{ig} \phi_g(\tilde{y}_i; \theta) \lambda_g(T_i)^{D_i} S_i(T_i) \right) - \sum_{i=1}^N \ln(J(y_i; \theta)). \quad (3.5)$$

Maximum likelihood estimates of the joint latent class model are obtained from that expression using an algorithm of maximization detailed in section 3.4. When necessary, some parameters were reparameterized to control for constraints. For the Beta transformations, the mean and variance parameters were preferred to the standard parameters.

Estimates are obtained for a given number of latent classes. In practice several models are estimated with different numbers of classes, and the best model is selected with the Bayes Information Criterion (BIC) (Schwartz (1978)). Indeed, several studies in standard mixture models emphasized that the BIC gave satisfactory results when determining the number of components compared to other criteria or tests (Hawkins et al. (2001); Miloslavsky and van der Laan (2003); Bauer and Curran (2003)).

### 3.2 Penalized likelihood for the semi-parametric survival model

When estimating the risk function using splines, a penalized likelihood can be used in order to limit local variations of the estimated function. The penalized log-likelihood is the sum of the log-likelihood given in (3.5) and a penalty which increases as the function to estimate has increasing local variations. We chose a penalty by the  $L^2$  norm of the second derivative of the risk function. The penalized log-likelihood was:

$$\mathcal{L}_p(\theta; \kappa) = \mathcal{L}(\theta) - \kappa \int_t \lambda_0''(u; \theta)^2 du, \quad (3.6)$$

where  $\kappa$  is the smoothing parameter which controls the trade-off between smoothness of the baseline function and fit of the data. The log-likelihood  $\mathcal{L}(\theta)$  has the same expression as in (3.5) except that baseline risk and survival functions are defined according to equations given in (2.6).

In order to decrease the number of parameters when using a semi-parametric survival model, we assumed a proportional hazard across latent classes. As a result, we only had one smoothing parameter to estimate. Methods for estimating  $\kappa$  were proposed (O'Sullivan (1988) ; Joly et al. (1999)) but they were

numerically too intensive with this model so  $\kappa$  was chosen heuristically.

### 3.3 Left-truncated data

When subjects are included in a cohort only if they have not experimented the event before their entry at  $T_{0i}$ , data are left-truncated. In that case, parameters have to be estimated by the maximization of the conditional likelihood obtained by dividing individual contribution to the likelihood  $L_i$  by the probability of being free of event at entry:

$$L_i^{\text{tr}}(\theta) = \frac{L_i(\theta)}{S(T_{0i}; \theta)}, \quad (3.7)$$

where  $S(T_{0i}; \theta)$  is the marginal survival function across the classes defined as follows:

$$S(T_{0i}; \theta) = \sum_{g=1}^G P(c_i = g) f(T_i \geq T_{0i} \mid c_i = g; \theta) = \sum_{g=1}^G \pi_{ig} S_g(T_{0i}; \theta). \quad (3.8)$$

### 3.4 Algorithm

Whatever the model, we used a modified Marquardt optimization algorithm (Marquardt (1963)), a Newton-Raphson-like algorithm (Fletcher (2000)) to obtain the maximum likelihood estimators. In this algorithm, the vector of parameters  $\theta$  is updated until convergence using:

$$\theta^{(l+1)} = \theta^{(l)} - \delta \tilde{H} \left( \theta^{(l-1)} \right)^{-1} U \left( \theta^{(l)} \right), \quad (3.9)$$

where  $U(\theta^{(l)})$  and  $H(\theta^{(l)})$  are respectively the gradient and the Hessian matrix at iteration  $l$ , the first and second derivatives being computed by finite differences. The matrix  $\tilde{H}$  is a diagonal-inflated Hessian to ensure positive-definiteness. The step  $\delta$  equals 1 by default but can be modified to ensure that the log-likelihood is improved at each iteration, and convergence is reached when the stopping criterion based on the second derivatives  $U(\theta^{(l)})' H(\theta^{(l)})^{-1} U(\theta^{(l)}) \leq 10^{-5}$ . After convergence, standard-error estimates of the parameter estimates are directly obtained using the inverse of the Hessian.

### 3.5 Posterior probabilities of latent classes and classification

From the maximum likelihood estimates  $\hat{\theta}$ , posterior probabilities of belonging to each latent class  $g$  given the repeated measures of the markers  $y_i$ , the covariates  $x_i$  and the survival data  $(T_i, D_i)$  can be computed:

$$\begin{aligned} \hat{\pi}_{ig}^{y,T} &= P(c_i = g \mid y_i, T_i, D_i, x_i; \hat{\theta}) \\ &= \frac{P(c_i = g \mid x_i; \hat{\theta}) f(y_i, T_i, D_i \mid c_i = g, x_i; \hat{\theta})}{\sum_{l=1}^G P(c_i = l \mid x_i; \hat{\theta}) f(y_i, T_i, D_i \mid c_i = l, x_i; \hat{\theta})}. \end{aligned} \quad (3.10)$$

A posterior classification is derived from these posterior probabilities, each subject being classified in the class in which he has the highest probability to belong. This classification may be used to evaluate goodness-of-fit of the model while posterior probabilities conditional only on the marker measures  $P(c_i = g \mid y_i, x_i; \hat{\theta})$  may also be computed and used to propose diagnostic or prognostic tools for the event.

## 4 Simulation study

We conducted a simulation study to investigate the quality of the maximum likelihood estimators in the parametric model using a Weibull hazard function. We considered the following model for three longitudinal markers and 2 latent classes:

$$\begin{aligned}
\Lambda_i(t) |_{c_i=g} &= (\mu_{0g} + u_{0i}) + (\mu_{1g} + u_{1i}) \times t, \\
\text{with } u_i &= \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 = 0.0095 & \sigma_{01} = -0.0013 \\ \sigma_{01} = -0.0013 & \sigma_1^2 = 0.00074 \end{pmatrix} \right) \\
\text{and } \mu_{01} &= 0.644, \quad \mu_{02} = 0.0613, \quad \mu_{11} = -0.062, \quad \mu_{12} = -0.124. \\
\lambda(t) |_{c_i=g} &= \zeta_{1g}^2 \zeta_{2g}^2 (\zeta_{1g}^2 t)^{\zeta_{2g}^2 - 1}, \\
\text{with } \zeta_{11} &= 0.102, \quad \zeta_{21} = 4.719, \quad \zeta_{12} = 0.109, \quad \zeta_{22} = 4.432. \\
h_k(Y_{ijk}; (\eta_{1k}, \eta_{2k})) &= \Lambda_i(t_{ijk}) + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim_{iid} \mathcal{N}(0, \sigma_\epsilon^{2(k)}), \quad k \in \{1, 2, 3\} \\
\text{with } \eta_{11} &= 1.000, \quad \eta_{21} = 0.719, \quad \eta_{12} = 0.989, \quad \eta_{22} = 0.645, \quad \eta_{13} = 0.428, \\
\eta_{23} &= 0.767, \quad \sigma_\epsilon^{(1)} = 0.048, \quad \sigma_\epsilon^{(2)} = 0.051, \text{ and } \sigma_\epsilon^{(3)} = 0.040.
\end{aligned}$$

Parameter values were chosen to mimic the application. Five hundred samples of 500 subjects were simulated. For each subject, we first simulated the class membership variable  $c_i$  according to a Bernoulli distribution with probability  $\pi_1 = 0.75$ , entry time in the cohort  $T_{0i}$  according to a uniform distribution on  $[65, 75]$ , time of event  $T_i^*$  according to the Weibull function in the latent class of subject  $i$ , and time of censoring  $C_i$  according to a uniform distribution on  $[T_{0i}, 95]$ . Then, we defined times of measurement for each marker:  $t_{i1k} = T_{0i}$  and  $t_{ijk} = \min(t_{i(j-1)k} + 3; T_i)$  while  $t_{i(j-1)k} < T_i$ . The number and times of measurement were the same for the two markers of a same subject but

differed among subjects. Finally we simulated random-effects  $(u_{ig})_{g=1,\dots,G}$ , test-specific intercepts  $(\alpha_{ik})_{k=1,\dots,3}$  and Gaussian errors  $(\epsilon_{ijk})_{(j=1,n_{ik};k=1,\dots,3)}$  in order to compute the vector of transformed variables  $\tilde{y}_i$  that we transformed in the markers scales using the inverse of the nonlinear functions  $(h_k(\cdot; \eta_k)^{-1})_{k=1,\dots,3}$ . Among the 500 replications, subjects had in mean 5.55 repeated measures in the first class and 4.53 in the second class with a higher risk of event. The mean time at entry in the study was 70.01 in the two classes, and the mean time at censoring or event was 82.14 in the first class and 79.06 in the second class. Finally 9.3% had the event in the first class versus 49.9% in the second class. The overall proportion of events was a little higher than in the application dataset (19.5% versus 14.9%). This was due to the simulation of left truncation and right censoring that slightly differed. The true parameter values were used as initial values to shorten the computational time but similar results were found on the first runs when using perturbed initial values. The algorithm did not converge for only 1 over the 500 replicates. Table 1 gives the results showing that the estimators are fairly good in terms of mean, bias and coverage rate. We note that the same conclusions arised when considering samples of 1000 subjects.

[Table 1 about here.]

## 5 Conditional independence assumption: a residual analysis

The conditional independence of the markers and the event can not be strictly controlled because the latent classes are not observed. A few methods were proposed based on the Bandeen-Roche et al. (1997) posterior individual affection. Lin et al. (2002) and (2004) proposed to evaluate the dependency

between the event and the marker after stratification on the posterior latent classes by using a weighted survival analysis in which a function of the markers is included as covariate. To avoid any posterior classification and the choice of a particular estimated function of the markers as covariate, we propose instead to evaluate the conditional independence assumption using a simple residual analysis. This approach was inspired from Dobson and Henderson (2003) who investigated the first and second moments of the longitudinal residuals conditionally on the time-to-event data in order to graphically assess the link between dropout and the longitudinal residuals in a shared-random-effect model (Henderson et al. (2000)).

### 5.1 *Time-to-event conditional residuals for testing the conditional independence*

The residuals of the transformed markers obtained from the joint model conditionally on the time-to-event data are:

$$\begin{aligned} R_{ijk} &= h_k(y_{ijk}; \hat{\eta}_k) - E\left(h_k(y_{ijk}; \hat{\eta}_k) | T_i, D_i; \hat{\theta}\right) \\ &= Z(t_{ijk}) \left(u_i - E\left(u_i | T_i, D_i; \hat{\theta}\right)\right) + \alpha_{ik} + \epsilon_{ijk}, \end{aligned} \tag{5.1}$$

where the random-effects  $u_i \sim \sum_{g=1}^G \pi_{ig} \mathcal{N}(\mu_g, \omega_g B)$  and

$$E\left(u_i | T_i, D_i; \hat{\theta}\right) = \sum_{g=1}^G E(u_i | c_i = g; \hat{\theta}) P(c_i = g | T_i, D_i; \hat{\theta}).$$

Using the Bayes theorem, we found

$$E\left(u_i | T_i, D_i; \hat{\theta}\right) = \frac{\sum_{g=1}^G \mu_g \pi_{ig} \lambda(T_i | c_i = g; \hat{\theta})^{D_i} S(T_i | c_i = g; \hat{\theta})}{\sum_{g=1}^G \pi_{ig} \lambda(T_i | c_i = g; \hat{\theta})^{D_i} S(T_i | c_i = g; \hat{\theta})}.$$

The standardized vector of residuals  $\tilde{R}_i = (\tilde{R}_{i11}, \dots, \tilde{R}_{ijk}, \dots, \tilde{R}_{in_{iKK}})^T$  is obtained from  $R_i = (R_{i11}, \dots, R_{ijk}, \dots, R_{in_{iKK}})^T$  using the Cholesky transformation of  $V_{R_i}^{-1}$  where

$$\begin{aligned}
V_{R_i} = V(R_i|T_i, D_i; \hat{\theta}) &= \begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_K \end{pmatrix} \\
&+ \begin{pmatrix} Z_i^1 \\ \vdots \\ Z_i^K \end{pmatrix} V(u_i|T_i, D_i; \hat{\theta}) \begin{pmatrix} Z_i^{1T} & \dots & Z_i^{KT} \end{pmatrix} \\
\text{and } V(u_i|T_i, D_i; \hat{\theta}) &= \frac{1}{f(T_i, D_i; \hat{\theta})} \sum_{g=1}^G \pi_{ig} f(T_i, D_i|c_i = g; \hat{\theta}) (\omega_g^2 B + \mu_g^2) \\
&- \left( \frac{1}{f(T_i, D_i; \hat{\theta})} \sum_{g=1}^G \pi_{ig} f(T_i, D_i|c_i = g; \hat{\theta}) \mu_g \right)^2. \tag{5.2}
\end{aligned}$$

Under the hypothesis of conditional independence, the standardized residuals  $\tilde{R}_{ijk}$  are independently distributed with zero mean and unit variance given the time-to-event. Thus, to evaluate this assumption, the means of the standardized residuals between censored and uncensored subjects can be simply compared.

## 5.2 Evaluation of the test for conditional independence

To evaluate the type I error and the statistical power of this test, we conducted a simulation study in a simple case of linear latent class model with a single

longitudinal marker. Samples of 500 and 1000 subjects were simulated using the same design as described in section 4 except that left truncation was uniform in  $[0, 10]$  and right censoring was uniform from entry to 30. The simulated model was

$$\begin{aligned}\Lambda_i(t) |_{c_i=g} &= (\mu_{0g} + u_{0i}) + (\mu_{1g} + u_{1i}) \times t + \epsilon_{ijk}, \\ \text{with } u_i &= \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 = 5 & \sigma_{01} = -2 \\ \sigma_{01} = -2 & \sigma_1^2 = 1 \end{pmatrix} \right) \\ \text{and } \epsilon_{ij} &\sim_{iid} \mathcal{N}(0, \sigma_\epsilon^2 = 1), \\ \text{and } \lambda(t) |_{c_i=g} &= \zeta_{1g} \zeta_{2g} (\zeta_{1g} t)^{\zeta_{2g}-1} \exp(-\delta \frac{u_{0i}}{\sigma_0}).\end{aligned}$$

Under the null hypothesis  $H_0$ , a 2 latent class model was simulated with  $\pi_1 = \pi_2 = 0.5$ ,  $(\mu_{01}, \mu_{11}) = (15, 2)$  and  $(\mu_{02}, \mu_{12}) = (12, 0)$  and the Weibull parameters chosen to obtain an event-free-survival at time 15 of 0.96 in class 1 and 0.65 in class 2 (with  $\delta = 0$ ). As alternative hypotheses, we considered two shared random-effect models with 2 latent classes ( $H_{a1} : \delta = 0.5$  and  $H_{a2} : \delta = 1$ ) and two models with 3 latent classes (and  $\delta = 0$ ) with  $\pi_1 = \pi_2 = \pi_3 = 0.33$ ,  $(\mu_{01}, \mu_{11}) = (15, 2)$ ,  $(\mu_{02}, \mu_{12}) = (12, 0)$  and  $(\mu_{03}, \mu_{13}) = (9, -1)$ . The event-free survivals at time 15 were 0.96 in class 1, 0.84 in class 2 and 0.65 in class 3 for the  $H_{a3}$  model and 0.99, 0.84 and 0.35 for the  $H_{a4}$  model. Whatever the simulated sample, a two latent class model was estimated and the means of the residuals among the subjects who experienced the event and those who did not were compared over 500 replicates. The estimated type I errors under  $H_0$  and the statistical power under alternative hypotheses are displayed in Table 2 for two sample sizes (N=500 and N=1000). The estimated type I errors were close to the nominal values for the two sample sizes, with

8.0%, 3.8% and 0.8% (for  $N=500$ ), and 11.2%, 6.0% and 1.4% (for  $N=1000$ ) for a nominal value of 10%, 5% and 1%. Under the shared random-effect models  $H_{a1}$  and  $H_{a2}$ , and using the threshold of 5%, the power was respectively 34.0% and 94.6% with  $N=500$  and 63.6% and 100% with  $N=1000$ . To give an idea of the impact of these misspecifications on the estimated parameters, the coverage rates for the fixed effects in the mixed model with  $N=500$  ranged from 92.8% to 94.4% for  $H_{a1}$  and from 88% to 92.4% for  $H_{a2}$ , and the relative bias were less than 1% and less than 2%. The relative biases for the 4 parameters from the survival model with  $N=500$  were 3.3%, 7.6%, 0.7% and 3.1% for  $H_{a1}$  and 2%, 16%, 2% and 11% for  $H_{a2}$ . Thus these misspecifications had only a modest impact on parameter estimates (especially for  $H_{a1}$ ). When the misspecification was due to a larger number of classes, the power was 34.0% for  $H_{a3}$  and 78.8% for  $H_{a4}$  with  $N=500$ , and 67.0% for  $H_{a3}$  and 97.8% for  $H_{a4}$  with  $N=1000$ , indicating a higher power with clearer separations of classes. This simulation study showed that a simple comparison test of means of the residuals conditional on the event was a satisfying test to evaluate the conditional independence in a latent class model.

[Table 2 about here.]

## 6 Application

### 6.1 The PAQUID dataset

The aim of the application was to study the association between the profiles of cognitive decline with age in the elderly and the occurrence of dementia. Data came from the French prospective cohort study PAQUID initiated in 1988 to

study normal and pathological aging (Letenneur et al. (1994)). Subjects included in the cohort were 65 and older at initial visit and were followed 6 times with a visit at 1, 3, 5, 8, 10 and 13 years after the initial visit. At each visit, a battery of psychometric tests was completed, and a two-phases screening procedure was carried out for the diagnosis of dementia. Three psychometric tests were considered in the present study: the Isaacs Set Test (IST) (Isaacs and Kennie (1973)) shortened at 15 seconds which evaluates verbal fluency, the Benton Visual Retention Test (BVRT) (Benton (1965)) which evaluates visual memory and the Digit Symbol Substitution Test of Wechsler (DSSTW) (Wechsler (1981)) which evaluates attention and logical reasoning. For the three tests, low values indicate a more severe impairment. The sample was constituted of 2383 subjects living in Gironde and free of dementia at inclusion: 355 subjects had a positive diagnosis of dementia during the follow up with a mean age at dementia of 85.04 (SE=5.77) and 2028 subjects were free of dementia at their last visit with a mean age at censoring of 80.25 (SE=6.83). Dementia was considered as a terminating event so that measures collected after diagnosis were not included in the analyses. The median number of measures per subject was 4 (interquartile range IQR=2-6) for the IST, 3 (IQR=2-6) for the BVRT and 2 (IQR=1-4) for the DSSTW. Two covariates were considered, gender and educational level in two classes (subjects who graduated from primary school e.g. obtained the CEP, the first French diploma, and those who did not). Among the 2383 subjects, 1382 were women (58%) and 692 (29%) did not graduate from primary school.

The common factor is defined between 0 and 1, 0 indicating the minimal cognitive level and 1 the maximal cognitive level. We assumed a quadratic function of age for the common factor in order to account for nonlinear cognitive declines with age. Gender and education (CEP versus not) were included in the class membership probability and in the risk of dementia after adjustment for the latent classes. To take into account the first passing effect described in Jacqmin-Gadda et al. (1997) (worse scores at baseline compared to other visits), we also included a binary indicator for the initial visit as a covariate. Finally, left-truncation was handled as explained in section 3.3 since subjects demented at age of entry in the cohort were excluded, and age was the time basis in the model. The joint latent class model we applied was

$$\begin{aligned}\Lambda_i(t) |_{c_i=g} &= u_{0ig} + u_{1ig} \times \text{AGE}_i + u_{2ig} \times \text{AGE}_i^2 + \beta \mathbf{I}_{t=T_{0i}}, \\ \lambda(t) |_{c_i=g} &= \lambda_{0g}(t) e^{\text{GENDER}_i \delta_1 + \text{CEP}_i \delta_2}, \\ h_k(Y_{ijk}; (\eta_{1k}, \eta_{2k})) &= \Lambda_i(t_{ijk}) + \alpha_{ik} + \epsilon_{ijk}, \\ \text{and } P(c_i = g) &= \frac{e^{\xi_{0g} + \text{GENDER}_i \xi_{1g} + \text{CEP}_i \xi_{2g}}}{\sum_{l=1}^G e^{\xi_{0l} + \text{GENDER}_i \xi_{1l} + \text{CEP}_i \xi_{2l}}},\end{aligned}$$

where  $u_{ig} = (u_{0ig}, u_{1ig}, u_{2ig})^T \sim \mathcal{N}((\mu_{0g}, \mu_{1g}, \mu_{2g})^T, \omega_g^2 B)$ . We performed preliminary analyses without covariates and with a common random-effect variance across the classes to compare 3 formulations of the survival model: a parametric Weibull model and a semi-parametric model both with the assumption of proportional hazard between classes, and a Weibull model stratified on the latent classes. Using the BIC selection criterion (Schwartz (1978)), a model with 4 classes was chosen. Based on this preliminary analysis, we retained the parametric model with proportional hazards over classes because it had

a better AIC than the stratified parametric model (AIC=135,269 vs 135,274) and the four estimated class-specific risk functions were very similar to those estimated by the semi-parametric approach using 5 knots (Figure 1).

[Fig. 1 about here.]

We then estimated this latent class model including covariates for a number of latent classes varying from 2 to 6. According to the BIC selection criterion, the model with 5 latent classes was retained (see table 3). The five latent classes differed by their cognitive evolution with age and their risk of dementia represented in figure 2.

[Table 3 about here.]

[Fig. 2 about here.]

[Table 4 about here.]

Estimates of the 5-class model are given in table 4. The probability of belonging to the latent classes was significantly associated with educational level ( $p < 0.001$  using a 4df Likelihood ratio test (LRT)) but not with gender ( $p = 0.084$  using a 4df LRT). Classes 2 and 4 consisted almost exclusively of subjects without CEP since the ratio of the probability to be in class 2 (respectively class 4) over the probability to be in class 1 was  $e^{\xi_{22}} = 3.9.10^{-3}$  (respectively  $e^{\xi_{24}} = 7.7.10^{-4}$ ) for subjects with CEP compared to subjects without CEP. In contrast, the odds-ratio to be in class 5 vs. class 1 was 0.012 for subjects with CEP compared to subjects without CEP. Conversely, class 1 and 3 included mostly subjects with CEP (explaining the large variance for the parameters in class 3). Consequently, latent classes 1 and 3 had a higher mean initial level than classes 2 and 4. Then, latent classes 1 and 2 differed

respectively from latent classes 3 and 4 by the shape of the cognitive decline with age. While latent classes 1 and 2 were characterized by a slight cognitive decline with age and a low probability of dementia, classes 3 and 4 were characterized by a sharper cognitive decline with age and a higher probability of dementia. At last, latent class 5 was characterized by a dramatic cognitive decline with age and a probability of dementia which reached 1 at 85 years old. Adjusted for the latent classes, graduated subjects had a higher risk of dementia ( $p = 0.0055$  using a Wald test). This means that for a same profile of cognitive evolution, highly educated subjects had a higher risk of being diagnosed as demented than lower educated subjects.

### 6.3 Adequation of the model

Specification of the survival model was evaluated in the preliminary analysis which showed both goodness-of-fit of the Weibull hazard function and correct assumption of proportional hazards over classes (figure 1). In this section, we present other analyses that evaluate goodness-of-fit of the model.

#### 6.3.1 The posterior classification

For each latent class, table 5 presents the means of the posterior probabilities of belonging to this latent class and to the four others.

[Table 5 about here.]

Whatever the latent class, the mean of the probabilities of belonging to the class in which the subjects were classified was between 0.68 and 0.78 while the means of the probabilities in the other classes was less than 0.20. This

suggests an unambiguous classification. Nevertheless, the non-diagonal terms indicated that subjects classified in class 1 had a non-negligible probability of belonging to class 3 (mean of 0.126) and conversely (mean of 0.154). In the same way, subjects classified in class 2 had a non-negligible probability of belonging to class 4 (mean of 0.140) and conversely (mean of 0.204).

### 6.3.2 Longitudinal multivariate model

We evaluated the adequation of the model to the observed repeated measures by comparing the predicted mean evolution and the observed mean evolution for each posterior latent class and each marker. The prediction of an observation in each latent class  $g$ ,  $\hat{y}_{ijk}|_{c_i=g} = E(h_k^{-1}(\tilde{y}_{ijk}; \hat{\eta}_k)|\hat{\theta}; c_i = g)$  was computed using a numerical integration of  $h_k^{-1}(\tilde{y}_{ijk}; \hat{\eta}_k)$  over the distribution of  $\tilde{y}_{ik}|_{c_i=g}$  that is a multivariate Gaussian distribution with mean vector and variance covariance matrix defined in (3.3) and (3.4) (details are given in Proust et al. (2006)). In a second step, the predicted mean evolution for each marker and each latent class was obtained by computing the mean of the predicted values for subjects observed in a window of time weighted by the individual posterior probability  $\hat{\pi}_{ig}^{y,T}$  of belonging to the class  $g$ . This predicted evolution was finally compared to the mean observed evolution obtained by computing the mean of the observations in the same window of time weighted by  $\hat{\pi}_{ig}^{y,T}$ . Figure 3 displays the predicted subject-specific mean evolution which includes the predicted individual random coefficients. It shows a very good adequation to the observed repeated measures.

[Fig. 3 about here.]

We evaluated the conditional independence assumption by comparing the means of the standardized conditional residuals given by (5.1) and (5.2) according to the occurrence of the event in the models including from 2 to 6 latent classes; p-values are given in table 3. The means of the residuals were no longer significantly different between demented and non-demented subjects in the model with 5 latent classes, suggesting that the conditional independence assumption in the selected 5-class model was satisfying.

## 7 **Concluding remarks**

We developed a latent class approach to jointly model the multivariate quantitative outcomes and the time-to-event. Joint latent class models had been yet restricted to the study of a single and Gaussian longitudinal outcome (Lin et al. (2002); Lin et al. (2004)) while it is not rare in applications to deal with possibly non Gaussian quantitative outcomes and/or several outcomes of the same latent quantity (Proust et al. (2006)).

The application emphasized that the model is particularly well designed for the study of cognitive aging and risk of dementia. Indeed, first it allows to explore heterogeneity of cognitive aging by distinguishing several profiles of normal cognitive change over age from the pathological profile of decline associated with high risk of dementia. Thanks to the latent classes, covariate effects on the probability of belonging to each cognitive profile can also be separated from effects on the risk of dementia given the cognitive profiles. Second as the longitudinal part of the model is a multivariate model with a latent process (Proust et al. (2006)), the analysis can focus on the evolution of the latent cog-

nitive level rather than on a specific psychometric test. Thus, several tests can be modeled at the same time, and biases due to misspecification of a Gaussian distribution for the tests are avoided by estimating nonlinear transformations between tests and latent process. To account for variability over markers, we chose to include marker-specific random-intercepts that were highly significant. Additional marker-specific random-slopes could also be added in the model in order to incorporate time-varying variability over markers. However, in our application, such additional inclusion did not change the results in terms of optimal number of classes and parameter estimates. Finally as the time to dementia is modeled instead of dementia status at a given time (Proust-Lima et al. (2007)), selection biases are avoided and cognitive decline is described according to age instead of according to the time-to-diagnosis.

In addition to a parametric specification of the risk function, we proposed a semi-parametric approach based on splines functions and a penalized likelihood approach. Until now, semi-parametric specifications had only been developed for simpler cases without joint modelling of longitudinal markers nor latent classes, and evaluation of the quality of the estimates had been assessed (Joly et al. (1998), Joly and Commenges (1999)). In joint modelling context, semi-parametric models are a lot more computationally intensive. However, we would recommend to use them for assessing the goodness-of-fit of parametric risk functions and as an alternative model when parametric functions are not adequate.

It has been shown recently that a deviation from the assumption of conditional independence in latent class models induced biased estimates (Pepe and Janes (2007)). For assessing this pivotal assumption, we proposed a simple test that compares the mean of the longitudinal residuals conditional on the event.

This test that avoids posterior classification and assumptions regarding the conditional dependency may also be used as a criterion to select the number of classes as an alternative to the BIC. Another way of evaluating the conditional independence assumption could consist in testing the need of a shared random-effect in addition to the latent class structure. However, estimation of such models would cumulate the numerical problems from the latent class and the shared random-effect approaches.

## References

- Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., Rathouz, P. J., 1997. latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association* 92 (440), 1375–1386.
- Bauer, D. J., Curran, P. J., 2003. Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods* 8 (3), 338–63.
- Benton, A., 1965. Manuel pour l’application du Test de Rétention Visuelle. Applications cliniques et expérimentales, 2nd Edition. Centre de Psychologie appliquée, Paris.
- Dobson, A., Henderson, R., 2003. Diagnostics for joint longitudinal and dropout time modeling. *Biometrics* 59, 741–51.
- Fletcher, R., 2000. Practical methods of optimization, second edition, chap. 3 Edition. John Wiley & Sons, New-York.
- Hawkins, D. S., Allen, D. M., Stromberg, A. J., 2001. Determining the number of components in mixtures of linear models. *Computational Statistics and Data Analysis* 38 (1), 15–48.
- Henderson, R., Diggle, P., Dobson, A., 2000. Joint modelling of longitudinal

- measurements and event time data. *Biostatistics* 1 (4), 465–80.
- Isaacs, B., Kennie, A. T., 1973. The Set test as an aid to the detection of dementia in old people. *British Journal of Psychiatry* 123 (575), 467–70.
- Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D., Dartigues, J. F., 1997. A 5-year longitudinal study of the Mini-Mental State Examination in normal aging. *American Journal of Epidemiology* 145 (6), 498–506.
- Joly, P., Commenges, D., 1999. A penalized likelihood approach for a progressive three-state model with censored and truncated data: application to AIDS. *Biometrics* 55, 887–890.
- Joly, P., Commenges, D., Letenneur, L., 1998. A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics* 54, 185–194.
- Joly, P., Letenneur, L., Alioum, A., Commenges, D., 1999. PHMPL: a computer program for hazard estimation using a penalized likelihood method with interval-censored and left-truncated data. *Computer Methods and Programs in Biomedicine* 60 (3), 225–231.
- Laird, N., Ware, J., 1982. Random-effects models for longitudinal data. *Biometrics* 38, 963–74.
- Letenneur, L., Commenges, D., Dartigues, J. F., Barberger-Gateau, P., 1994. Incidence of dementia and Alzheimer’s disease in elderly community residents of south-western France. *International Journal of Epidemiology* 23 (6), 1256–61.
- Lin, H., McCulloch, C. E., Rosenheck, R. A., 2004. Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* 60 (2), 295–305.
- Lin, H., McCulloch, C. E., Turnbull, B. W., Slate, E. H., Clark, L. C., 2000. A latent class mixed model for analysing biomarker trajectories with irreg-

- ularly scheduled observations. *Stat Med* 19 (10), 1303–18.
- Lin, H., Turnbull, B. W., McCulloch, C. E., Slate, E. H., 2002. 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* 97, 53–65.
- Marquardt, D., 1963. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* 11, 431–41.
- Miloslavsky, M., van der Laan, M. J., 2003. Fitting of mixtures with unspecified number of components using cross validation distance estimate. *Computational Statistics and Data Analysis* 41 (3-4), 413–428.
- Muthén, B., Shedden, K., 1999. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 55 (2), 463–9.
- O’Sullivan, F., 1988. Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and Statistical Computing* 9, 363–379.
- Pepe, M. S., Janes, H., 2007. Insights into Latent Class Analysis of Diagnostic Test Performance. *Biostatistics* 8(2), 474–484.
- Proust, C., Jacqmin-Gadda, H., Taylor, J. M., Ganiayre, J., Commenges, D., 2006. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics* 62 (4), 1014–24.
- Proust-Lima, C., Letenneur, L., Jacqmin-Gadda, H., 2007. A nonlinear latent class model for joint analysis of multivariate longitudinal data and a binary outcome. *Statistics in Medicine* 26, 2229–45.
- Ramsay, J., 1988. Monotone regression splines in action. *Statistical science* 3 (4), 425–461.
- Roy, J., Lin, X., 2000. Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* 56, 1047–1054.

Schwartz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics* 6, 461–464.

Wechsler, D., 1981. WAIS-R manual. Psychological Corporation, New York.

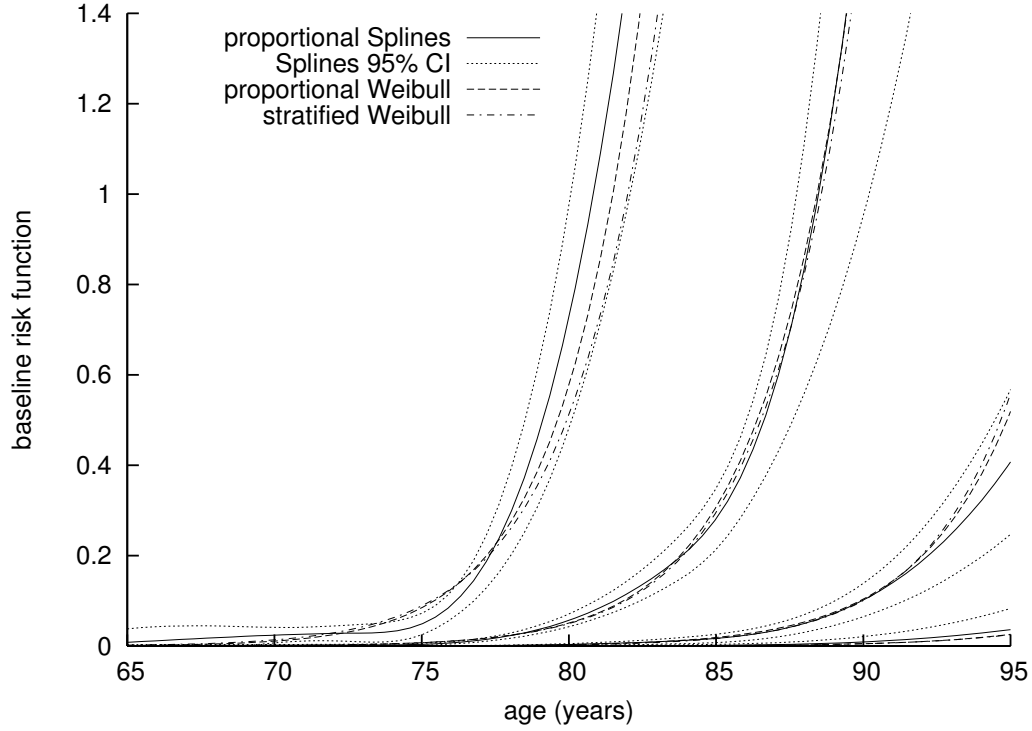


Fig. 1. Predicted baseline risk functions using proportional splines (plain line and dotted line for the 95% confidence bands computed using the Delta-method) or using either a Weibull function with proportional hazard between classes (dashed line) or stratified on the latent classes (dashed and dotted line) in the 4-class model without adjustment for covariates.

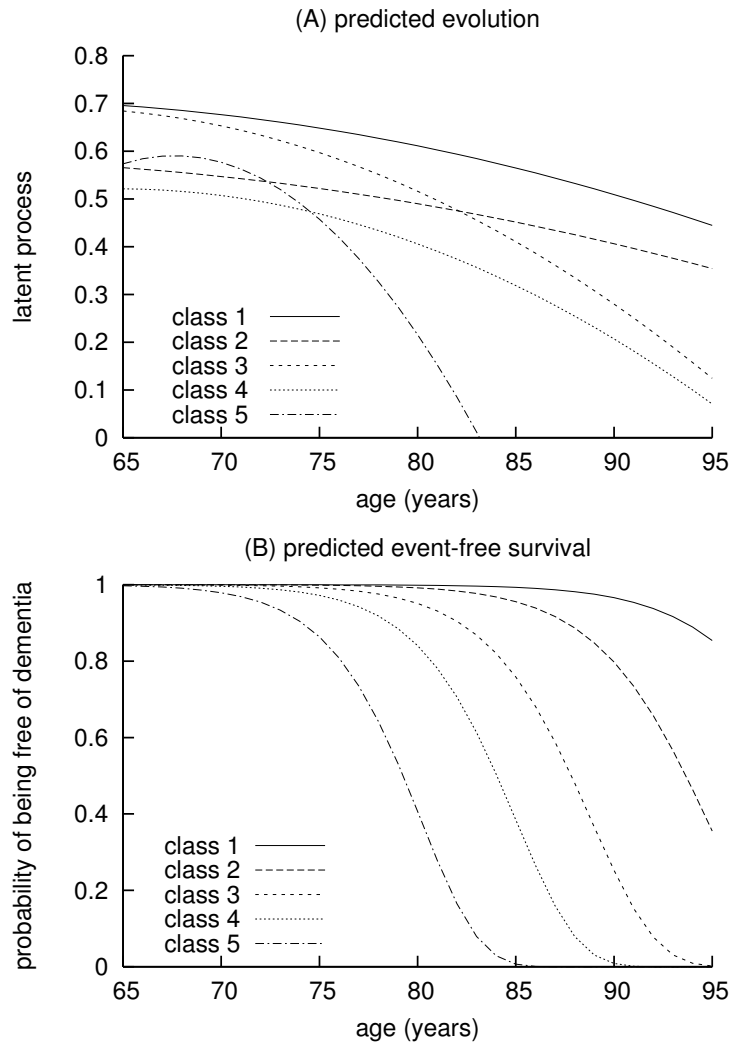


Fig. 2. (A) Predicted cognitive evolution according to age and (B) predicted survival curves without dementia according to age in the five latent class model for a woman without CEP.

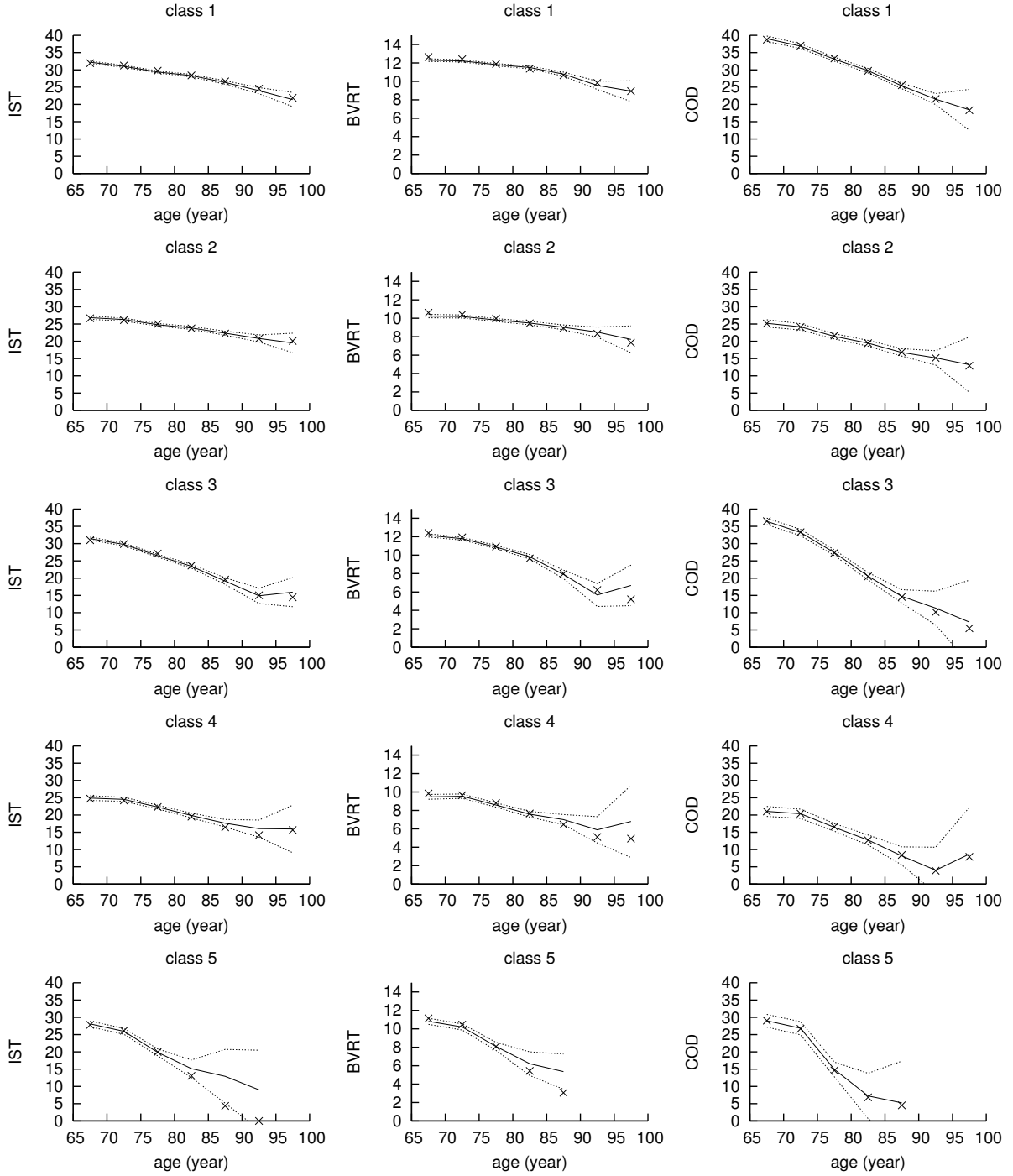


Fig. 3. Subject specific predicted weighted mean evolution ( $\mathbf{x}$ ) and observed weighted mean evolution (—) with 95% confidence bands (- - -) for each marker and each latent class.

Table 1

Mean estimate, relative bias, empirical standard error (SE), asymptotical SE and coverage rate for the 500 replications of the joint model for two classes and two markers and a sample of 500 subjects.

| parameter               | simulated | mean     | relative | empirical | asymptotic | coverage |
|-------------------------|-----------|----------|----------|-----------|------------|----------|
|                         | value     | estimate | bias (%) | SE        | SE         | rate (%) |
| $\pi_1^\dagger$         | 1.099     | 1.078    | -1.89    | 0.197     | 0.183      | 92.4     |
| $\zeta_{11}$            | 0.102     | 0.102    | -0.09    | 0.0007    | 0.0007     | 95.6     |
| $\zeta_{12}$            | 4.719     | 4.732    | 0.27     | 0.502     | 0.486      | 93.2     |
| $\zeta_{21}$            | 0.109     | 0.109    | 0.01     | 0.0005    | 0.0005     | 92.2     |
| $\zeta_{22}$            | 4.432     | 4.513    | 1.82     | 0.321     | 0.312      | 94.2     |
| $\mu_{01}$              | 0.644     | 0.644    | -0.03    | 0.013     | 0.013      | 95.0     |
| $\mu_{02}$              | 0.613     | 0.613    | -0.02    | 0.017     | 0.017      | 95.0     |
| $\mu_{11}$              | -0.062    | -0.061   | -1.62    | 0.0037    | 0.0037     | 93.6     |
| $\mu_{12}$              | -0.124    | -0.122   | -1.19    | 0.0081    | 0.0078     | 93.2     |
| $\eta_{11}^\dagger$     | 0.332     | 0.322    | -3.06    | 0.047     | 0.046      | 94.4     |
| $\eta_{21}^\dagger$     | -2.415    | -2.398   | -0.69    | 0.058     | 0.056      | 93.6     |
| $\eta_{12}^\dagger$     | 0.428     | 0.416    | -2.65    | 0.050     | 0.048      | 94.8     |
| $\eta_{22}^\dagger$     | -2.400    | -2.383   | -0.72    | 0.060     | 0.058      | 92.0     |
| $\eta_{13}^\dagger$     | -0.582    | -0.583   | -0.24    | 0.041     | 0.041      | 95.0     |
| $\eta_{23}^\dagger$     | -2.256    | -2.245   | -0.48    | 0.043     | 0.043      | 94.4     |
| $\sigma_1^\dagger$      | 0.097     | 0.095    | -2.13    | 0.0060    | 0.0057     | 91.6     |
| $\sigma_2^\dagger$      | -0.013    | -0.013   | -1.33    | 0.0023    | 0.0022     | 93.8     |
| $\sigma_3^\dagger$      | 0.024     | 0.024    | -2.11    | 0.0021    | 0.0020     | 93.0     |
| $\sigma_\epsilon^{(1)}$ | 0.048     | 0.047    | -1.47    | 0.0024    | 0.0023     | 93.4     |
| $\sigma_\epsilon^{(2)}$ | 0.051     | 0.050    | -1.67    | 0.0025    | 0.0024     | 92.0     |
| $\sigma_\epsilon^{(2)}$ | 0.040     | 0.039    | -1.38    | 0.0020    | 0.0019     | 93.0     |

<sup>†</sup> Reparametrized parameters. Latent class probability  $\pi_1$  was replaced by its logit correspondence. Canonical parameters  $\eta$  of Beta functions were replaced by transformations of the means and variances of the Beta functions. Variance covariance parameters ( $\sigma_0^2, \sigma_{01}, \sigma_1^2$ ) were replaced by the corresponding Cholesky parameters ( $\sigma_1, \sigma_2, \sigma_3$ ).

Table 2

Estimated type I error under  $H_0$  and statistical power (in %) under the 4 alternative hypotheses ( $H_{a1}$ ,  $H_{a2}$ ,  $H_{a3}$  and  $H_{a4}$ ) for a threshold of 5% over 500 simulations with two sample sizes (N=500 and N=1000).

| Hypothesis | N=500 | N=1000 |
|------------|-------|--------|
| $H_0$      | 3.8   | 6.0    |
| $H_{a1}$   | 34.0  | 63.6   |
| $H_{a2}$   | 94.6  | 100.0  |
| $H_{a3}$   | 34.0  | 67.0   |
| $H_{a4}$   | 67.0  | 97.8   |

Table 3

Adequation of the adjusted latent class model with the proportional risk over the classes on the 2383-subjects sample (number of latent classes  $G$ , number of parameters  $p$ , log-likelihood  $\mathcal{L}$ , BIC criterion, p-value of the test comparing the residuals means according to the event).

| $G$ | $p$ | $\mathcal{L}$ | BIC       | p-value for residuals means |
|-----|-----|---------------|-----------|-----------------------------|
| 2   | 34  | -67390.49     | 135045.38 | $p < 0.001$                 |
| 3   | 42  | -67262.12     | 134850.83 | $p = 0.037$                 |
| 4   | 50  | -67186.32     | 134761.43 | $p = 0.004$                 |
| 5   | 58  | -67141.29     | 134733.60 | $p = 0.486$                 |
| 6   | 66  | -67115.15     | 134743.53 | $p = 0.318$                 |

Table 4

Estimates (standard-errors) of the adjusted joint latent class model including 5 classes.

| Parameter                               | Common    | Estimates specific to the latent classes |          |          |          |         |
|-----------------------------------------|-----------|------------------------------------------|----------|----------|----------|---------|
|                                         | Estimates | 1                                        | 2        | 3        | 4        | 5       |
| <u>class membership probability:</u>    |           |                                          |          |          |          |         |
| Intercept                               |           | 0                                        | 4.338    | -10.944  | 3.714    | 2.104   |
|                                         |           |                                          | (0.942)  | (88.28)  | (0.947)  | (0.975) |
| Gender <sup>†</sup>                     |           | 0                                        | -0.991   | -0.155   | -0.503   | -0.245  |
| ( $p = 0.084$ )                         |           |                                          | (0.307)  | (0.230)  | (0.392)  | (0.315) |
| CEP <sup>‡</sup>                        |           | 0                                        | -5.535   | 9.932    | -7.163   | -4.435  |
| ( $p < 0.0001$ )                        |           |                                          | (0.983)  | (88.29)  | (1.272)  | (0.960) |
| <u>Risk of dementia:</u>                |           |                                          |          |          |          |         |
| $\sqrt{\zeta_1}$                        | 0.099     |                                          |          |          |          |         |
|                                         | (0.001)   |                                          |          |          |          |         |
| $\sqrt{\zeta_2}$                        | 5.302     |                                          |          |          |          |         |
|                                         | (0.284)   |                                          |          |          |          |         |
| $\delta_{0g}$                           |           | 1                                        | 1.879    | 3.686    | 4.927    | 6.572   |
|                                         |           |                                          | (0.337)  | (0.337)  | (0.418)  | (0.401) |
| Gender <sup>†</sup>                     | -0.129    |                                          |          |          |          |         |
|                                         | (0.145)   |                                          |          |          |          |         |
| CEP <sup>‡</sup>                        | 0.787     |                                          |          |          |          |         |
|                                         | (0.269)   |                                          |          |          |          |         |
| <u>Longitudinal multivariate model:</u> |           |                                          |          |          |          |         |
| Intercept                               |           | 0.696                                    | 0.565    | 0.684    | 0.521    | 0.574   |
|                                         |           | (0.0096)                                 | (0.012)  | (0.016)  | (0.018)  | (0.031) |
| Age <sup>§</sup>                        |           | -0.0289                                  | -0.0300  | -0.0374  | -0.0034  | 0.129   |
|                                         |           | (0.0075)                                 | (0.012)  | (0.0223) | (0.0197) | (0.057) |
| Age <sup>2</sup> §                      |           | -0.018                                   | -0.0135  | -0.0497  | -0.0489  | -0.245  |
|                                         |           | (0.0026)                                 | (0.0039) | (0.111)  | (0.0083) | (0.032) |
| First visit T0                          | -0.017    |                                          |          |          |          |         |
|                                         | (0.001)   |                                          |          |          |          |         |

<sup>†</sup> reference: women

<sup>‡</sup> reference: subjects without CEP 36

<sup>§</sup> Age in decades from 65 years old  $\left(\frac{age - 65}{10}\right)$

Table 5

Mean of the posterior probabilities of belonging to each latent class according to the final classification.

| Final<br>classification | Number of<br>subjects | Mean of the probabilities of belonging to each class |       |       |       |       |
|-------------------------|-----------------------|------------------------------------------------------|-------|-------|-------|-------|
|                         |                       | 1                                                    | 2     | 3     | 4     | 5     |
| 1                       | 73                    | 0.779                                                | 0.083 | 0.126 | 0.003 | 0.009 |
| 2                       | 187                   | 0.073                                                | 0.723 | 0.040 | 0.140 | 0.024 |
| 3                       | 197                   | 0.154                                                | 0.106 | 0.681 | 0.038 | 0.021 |
| 4                       | 652                   | 0.001                                                | 0.204 | 0.002 | 0.727 | 0.066 |
| 5                       | 1274                  | 0.012                                                | 0.044 | 0.074 | 0.097 | 0.773 |