Joint model with latent state for longitudinal and multistate data

E. Dantan, P. Joly, J.-F. Dartigues, H. Jacqmin-Gadda

Biostatistics (2011), **12**, 4, *pp.* 723–736
doi:10.1093/biostatistics/kxr003
Advance Access publication on March 17, 2011

D

Joint model with latent state for longitudinal and multistate data

E. DANTAN, P. JOLY, J.-F. DARTIGUES, H. JACQMIN-GADDA*

Institut National de la Santé et de la Recherche Médicale, U897, Bordeaux, F-33000, France and Université Victor Segalen Bordeaux 2, Institut de Santé Publique, d'Epidémiologie et de Developpement, Bordeaux, F-33000, France
helene.jacqmin-gadda@isped.u-bordeaux2.fr

SUMMARY

In many chronic diseases, the patient's health status is followed up by quantitative markers. The evolution is often characterized by a 2-phase degradation process, that is, a normal phase followed by a pathological degradation phase preceding the disease diagnosis. We propose a joint multistate model with latent state for the joint modeling of repeated measures of a quantitative marker, time-to-illness and time-to-death. Using data from the PAQUID cohort on cognitive aging, we jointly studied cognitive decline, dementia risk, and death risk. We estimated the mean evolution of cognitive scores given age at dementia for subjects alive and demented, the mean evolution of cognitive scores for subjects alive and nondemented, in addition to age at acceleration of cognitive decline and duration of the pre-dementia phase.

Keywords: Cognitive aging; Informative censoring; Latent state; Longitudinal data analysis; Survival analysis.

1. INTRODUCTION

Many chronic diseases such as cancer, HIV, and Alzheimer's disease involve a long-term process often beginning before the disease diagnosis. The subjects go successively through several states from normal to severe disease, and the disease diagnosis can be considered as one of these states. Moreover, the patient's health status may be followed up by one or several quantitative markers such as prostate-specific antigen for prostate cancer, CD4 T-cell counts for HIV, or cognitive tests for Alzheimer's disease. Multistate models (Hougaard, 1999) may be used to describe the development of such diseases, but they do not describe the time course of the biomarkers. During the last 10 years, joint models have been developed to study the relationship between a longitudinal response process and time-to-event (Tsiatis and Davidian, 2004). Some joint models have been proposed for a longitudinal biomarker and several time-to-events, either for recurrent events (Han *and others*, 2007) or for competing risks (Elashoff *and others*, 2008). However, to our knowledge, there has been no attempt at combining multistate models and mixed models for longitudinal data in order to study the links between biomarker evolution and the risk of transition between the different states of the disease.

Among the chronic diseases, Alzheimer's disease is characterized by a very long pre-diagnosis decline with a 2-phase degradation process of cognitive functions (Amieva *and others*, 2008) and large interindividual variability. Knowledge of the pre-diagnosis phase is a real public health challenge both

*To whom correspondence should be addressed.

for the understanding of the pathological process (successive emergence of clinical symptoms) and for the early detection of subjects at high risk of Alzheimer's disease. Indeed, these subjects could be the target population for new treatments since postdiagnosis treatments have shown only a modest impact on the evolution of clinical symptoms. Random changepoint mixed models have been used to describe this long-term pre-diagnosis cognitive decline (Hall *and others*, 2003; Dominicus *and others*, 2008). Jacqmin-Gadda *and others* (2006) proposed a joint model with random changepoint for time-to-dementia and cognitive decline that distinguishes normal and pathological cognitive aging dealing with the right censoring of dementia. It avoids the selection biases that arise when comparing 2 groups of subjects, normal and demented, at a given time point. However, this model has 2 main limitations.

First, it does not take death into account. In most cohorts on aging, dementia diagnosis is assessed at the follow-up visits, while death times are observed exactly. Subjects who are dementia free at their last visit and then die are considered as censored at their last visit. Estimates can be biased as risk of death is higher among subjects with poor cognitive decline (Wilson *and others*, 2003; Small *and others*, 2003) and among demented subjects (Joly *and others*, 2002). Very recently, Yu and Ghosh (2010) extended the Jacqmin-Gadda *and others* model by jointly modeling dementia-free death considered as a competing event with a cure model approach. However, this model does not address informative censoring due to death since it assumes independence between time-to-death and cognitive level, given covariates. Indeed, a multistate model is required to take into account the dependence of death risk on cognitive status.

Second, in both of these 2 models, the risk of dementia depends on the age at cognitive decline acceleration, but it does not increase when the subject enters the phase of accelerated decline. It is more realistic to assume that the risk of dementia is null before the acceleration of the decline and then increases gradually. The phase of accelerated decline before dementia may be considered as a latent state of the disease, called the pre-dementia or pre-diagnosis phase, and is only indirectly observed through measures of cognition. This transitional state is strongly linked with the concept of mild cognitive impairment (MCI) which is defined as a memory impairment without dementia (Petersen *and others*, 1999) and has been widely discussed in the literature (Dubois *and others*, 2007).

This paper proposes a joint model with latent state for longitudinal and multistate data. We introduce this joint model in Section 2. The maximum likelihood estimation procedure is presented in Section 3. In Section 4, the model is applied to the PAQUID cohort, a French prospective cohort including 3777 elderly subjects followed up for 15 years. This model makes it possible to estimate (a) the mean trajectories of cognitive scores for subjects alive given their age at dementia or their age at death, (b) the age at acceleration of the cognitive decline, and (c) the duration of the pre-dementia phase.

2. JOINT MODEL

We consider a joint model for the repeated measures of a marker, the time-to-illness and the time-to-death. The marker change over time is assumed to be in 2 phases: a phase of slight linear decline corresponding to normal aging, and a phase of accelerated decline corresponding to the pre-diagnosis phase of dementia. This is described by a segmented mixed model with a random changepoint, denoted τ_i which is the age of subject i at entry into the pre-diagnosis phase. We assume that the risk of illness is null before entering in the pre-diagnosis phase (this is a mandatory phase before the disease) while subjects may decease from the 3 other states (healthy, pre-diagnosis, and illness). The joint model is described graphically in Figure 1.

2.1 *Marker model*

Let $Y_i(t)$ be the marker value of subject i at age t for $i = 1, \ldots, N$. We note $Y_{ij} = Y_i(t_{ij})$ for $j = 1, \ldots, n_i$. So, Y_i is the vector of the n_i measurements for subject i. The marker change over time is

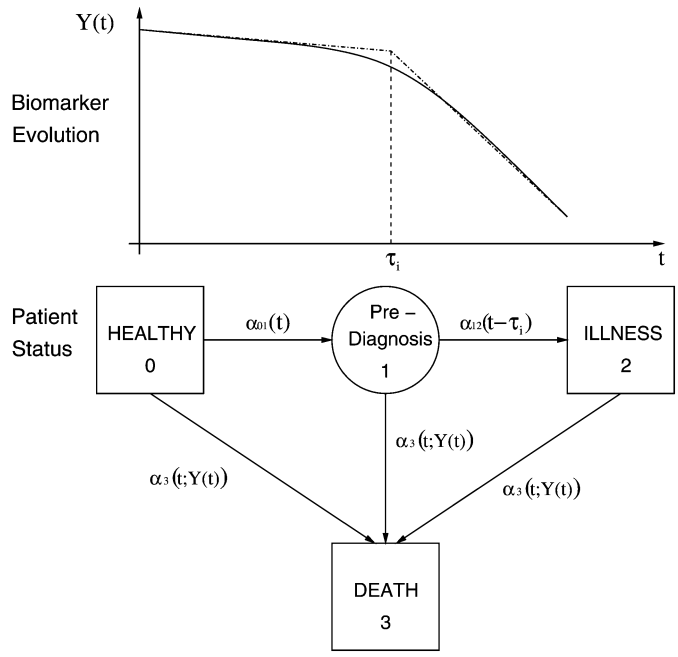

Fig. 1. Joint model for the pre-diagnosis phase of disease.

described by a segmented mixed model with a linear trend before and after the changepoint:

$$Y_{ij} = \begin{cases} b_{01i} + b_{11i} t_{ij} + \epsilon_{ij} & \text{if } t_{ij} \leqslant \tau_i, \\ b_{02i} + b_{12i} t_{ij} + \epsilon_{ij} & \text{if } t_{ij} > \tau_i \end{cases}$$

with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and b_{11i} the slope of cognitive decline in the healthy phase and b_{12i} the slope of cognitive decline in the pre-diagnosis phase. To ensure the continuity at changepoint τ_i, we impose the constraint $b_{02i} = b_{01i} + (b_{11i} - b_{12i})\tau_i$. Using the parameter transformation recommended by Bacon and Watts (1971) for numerical stability,

$$\beta_{0i} = b_{01i} + b_{11i}\tau_i, \quad \beta_{1i} = \frac{b_{11i} + b_{12i}}{2}, \quad \beta_{2i} = \frac{b_{12i} - b_{11i}}{2},$$

the model is rewritten as $Y_{ij} = \beta_{0i} + \beta_{1i}(t_{ij} - \tau_i) + \beta_{2i}(t_{ij} - \tau_i)\text{sgn}(t_{ij} - \tau_i) + \epsilon_{ij}$ with the function $\text{sgn}(t) = -1$ if $t < 0$ and $\text{sgn}(t) = 1$ if $t \geqslant 0$. Thus, β_{0i} corresponds to the mean marker value at the random changepoint, β_{1i} is the mean of the 2 slopes, and β_{2i} is half the difference between the slopes. These coefficients are modeled as the sum of a mean effect that may depend on fixed covariates and an individual random effect:

$$\beta_{ki} = \phi_k + \alpha_k' X_{ki} + u_{ki} \tag{2.1}$$

for $k = 0, 1, 2$. The vector $u_i = (u_{0i}, u_{1i}, u_{2i})'$ of random effects is $N(0, G)$. In (2.1), u_{1i} is a random effect on the mean slopes, whereas u_{2i} is a random effect on half the difference of the 2 slopes. The following alternative formulation of the model allows to introduce separately random slopes in the first

phase ($u_{s_1 i}$) and in the second phase ($u_{s_2 i}$):

$$\begin{aligned}\beta_{1i} &= \phi_1 + \alpha_1 X_{1i} + \frac{u_{s_1 i}}{2} + \frac{u_{s_2 i}}{2},\\ \beta_{2i} &= \phi_2 + \alpha_2 X_{2i} - \frac{u_{s_1 i}}{2} + \frac{u_{s_2 i}}{2}.\end{aligned} \tag{2.2}$$

Thus, for instance, this model formulation makes possible to keep in the model a random slope for the pre-diagnosis phase while excluding the random slope in the healthy phase as we have previously observed that the variability in the pre-diagnosis phase is higher than in the healthy phase (Jacqmin-Gadda *and others*, 2006). In this formulation, the random-effect vector is $u_i = (u_{0i}, u_{s_1 i}, u_{s_2 i})'$.

Finally, clinical observations do not suggest an abrupt change in the rate of decline between the 2 phases but rather a smooth transition. To insure this smooth transition, the function $\mathrm{sgn}(t)$ is replaced by the function $\mathrm{trn}(t;\gamma) = \frac{1}{t}\sqrt{t^2+\gamma}$ (Seber and Wild, 2003):

$$Y_{ij} = \beta_{0i} + \beta_{1i}(t_{ij} - \tau_i) + \beta_{2i}(t_{ij} - \tau_i)\mathrm{trn}(t_{ij} - \tau_i;\gamma) + \epsilon_{ij}. \tag{2.3}$$

This trn function also insures the continuity of the first and second derivatives at τ_i with respect to all parameters in the mixed model that is required for using Newton–Raphson–like algorithm for the parameter estimation. As $\lim_{\gamma\to 0}\mathrm{trn}(t;\gamma) = \mathrm{sgn}(t)$ for small γ, the expectation of Y may be approximated by

$$E(Y_{ij}|\tau_i, u_i) \approx \tilde{Y}_i(t_{ij}|\tau_i, u_i) = \begin{cases} \beta_{0i} + (\beta_{1i} - \beta_{2i})(t_{ij} - \tau_i) & \text{if } t_{ij} \leqslant \tau_i,\\ \beta_{0i} + (\beta_{1i} + \beta_{2i})(t_{ij} - \tau_i) & \text{if } t_{ij} > \tau_i.\end{cases} \tag{2.4}$$

As we assume a linear–linear trend for $Y(t)$, we choose a very small value for γ.

2.2 *Multistate model*

Latent state transition intensity. The distribution of the random changepoint τ_i which is the age at entry in the latent pre-diagnosis state (state 1) is defined by a proportional transition intensity model for the transition from healthy state (state 0) to pre-diagnosis state (state 1):

$$\alpha_{01}(t) = \alpha_{01}^0(t)\mathrm{e}^{\theta_L' X_L}, \tag{2.5}$$

where $\alpha_{01}^0(t)$ is the baseline transition intensity and X_L a q_L-vector of covariates associated with a q_L-vector of parameters θ_L and the age t as timescale. The cumulative transition intensity to the latent state at time t is

$$\Lambda_{01}(t) = \int_0^t \alpha_{01}(\upsilon)\mathrm{d}\upsilon.$$

We assume independance between τ_i and the random effects u_i because identifiability problems were observed in the Jacqmin-Gadda *and others* (2006) model when the random changepoint depended on the random slopes. The main clinical consequence behind this constraint is that the slope of cognitive decline in the pre-diagnosis phase does not depend on the age at entry in this phase after adjustment on covariates (educational level for instance). Given that this pre-diagnosis slope may depend on the slope before τ_i ($u_{s_1 i}$) and on the cognitive level at τ_i (u_{0i}), we think the assumption of independence with τ_i is reasonnable.

Illness transition intensity. The illness risk is assumed to be null before entering in the pre-diagnosis state and then increase with the time spent in the pre-diagnosis phase. The transition intensity from the pre-diagnosis state (state 1) to the illness state (state 2) follows a proportional transition intensity model that depends on both the time since entry in state 1 ($t - \tau_i$) as timescale and possibly on random effects.

$$\alpha_{12}(t - \tau_i|u_i) = \alpha_{12}^0(t - \tau_i)\mathrm{e}^{\theta_e' X_{ei} + \nu' u_i}, \tag{2.6}$$

where $\alpha_{12}^0(t-\tau_i)$ is the baseline transition intensity, X_{ei} a q_e-vector of covariates associated with a q_e-vector of parameters θ_e, and ν a k-vector of parameters associated with the random effects u_i. Thus, the cumulative transition intensity from state 1 to state 2 is

$$\Lambda_{12}(t-\tau_i|u_i) = \int_0^{t-\tau_i} \alpha_{12}(\upsilon|u_i)\mathrm{d}\upsilon.$$

As we assume that direct transition from state 0 to state 2 is impossible, the illness intensity function is defined as

$$\alpha_e(t|\tau_i,u_i) = \mathbb{1}_{\{t>\tau_i\}}\alpha_{12}(t-\tau_i|u_i)$$

and the corresponding cumulative transition intensity as

$$\Lambda_e(t|\tau_i,u_i) = \int_0^t \alpha_e(\upsilon|\tau_i,u_i)\mathrm{d}\upsilon = \mathbb{1}_{\{t>\tau_i\}}\Lambda_{12}(t-\tau_i|u_i).$$

Death transition intensity. The transition intensity to death state (state 3) is assumed to be independent of the current patient state (healthy, pre-diagnosis, or illness) conditionally on the current expected biomarker value $\tilde{Y}(t|\tau_i,u_i)$. This assumption is discussed in Section 5. We assume that a proportional transition intensity model describes the death transition intensity as a function of age, $\tilde{Y}(t|\tau_i,u_i)$ and a q_d-vector of covariates X_{di}:

$$\alpha_3(t|\tau_i,u_i) = \alpha_3^0(t)\mathrm{e}^{\theta_d' X_{di} + \eta\tilde{Y}_i(t|\tau_i,u_i)}, \tag{2.7}$$

where $\alpha_3^0(t)$ is the baseline transition intensity for death and $\tilde{Y}_i(t|\tau_i,u_i)$ is defined in (2.4). The cumulative transition intensity to death is

$$\Lambda_3(t|\tau_i,u_i) = \int_0^t \alpha_3(\upsilon|\tau_i,u_i)\mathrm{d}\upsilon.$$

3. Estimation

Let us define $Y_i = (Y_{i1},\ldots,Y_{in_i})^T$ the vector of repeated measurement for subject i and (T_{di},δ_{di}) where $T_{di} = \min(T_{di}^*, C_{di})$ and $\delta_{di} = \mathbb{1}_{\{T_{di}^*<C_{di}\}}$ with T_{di}^* the death age and C_{di} the censoring age for death. Similarly, we define (T_{ei},δ_{ei}) where $T_{ei} = \min(T_{ei}^*, C_{ei}, T_{di})$ and $\delta_{ei} = \mathbb{1}_{\{\{T_{ei}^*<C_{ei}\}\&\{T_{ei}^*<T_{di}\}\}}$ with T_{ei}^* the age at illness diagnosis and C_{ei} the censoring age for this event. Note that the censoring time for illness diagnosis may be different from the censoring time for death since a face-to-face interview is required for the disease diagnosis while age at death is exactly observed. Parameters are estimated using a maximum likelihood approach under the following additional assumptions:

1. Missing values of the marker not due to the disease or the death are missing at random.
2. Censoring for death is not informative (in practice, this is often an administrative censoring corresponding to the end of the study).
3. For subjects alive, censoring for illness diagnosis is not informative. Given the model definition, the competitive risk due to death is taken into account.

We assume parametric baseline transition intensities $\alpha_{01}^0(t)$, $\alpha_{12}^0(t-\tau_i)$ and $\alpha_3^0(t)$ and denote θ the vector of all regression parameters in (2.3), (2.5–2.7) and variance–covariance parameters (G and σ_ϵ^2). Parameter γ of the function trn is fixed by the user to a small value. Since the 3 outcomes (the marker Y_i, the illness age T_{ei}, and the death age T_{di}) are independent given τ_i and u_i, the individual contribution to the likelihood is developed as follows:

$$
\begin{aligned}
L_i(\theta) &= L_i(Y_i, T_{ei}, \delta_{ei}, T_{di}, \delta_{di}; \theta) \\
&= \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \alpha_e(T_{ei}|\tau_i, u_i)^{\delta_{ei}} \, \mathrm{e}^{-\Lambda_e(T_{ei}|\tau_i, u_i)} \alpha_3(T_{di}|\tau_i, u_i)^{\delta_{di}} \, \mathrm{e}^{-\Lambda_3(T_{di}|\tau_i, u_i)} \\
&\quad \times f_Y(Y_i|\tau_i, u_i) \, f(\tau_i, u_i) \mathrm{d}\tau_i \, \mathrm{d}u_i \qquad (3.1) \\
&= \int_{-\infty}^{+\infty} \int_{0}^{+\infty} [\mathbb{1}_{\{T_{ei} > \tau_i\}} \alpha_{12}(T_{ei} - \tau_i|u_i)]^{\delta_{ei}} \, \mathrm{e}^{-\mathbb{1}_{\{T_{ei} > \tau_i\}} \Lambda_{12}(T_{ei} - \tau_i|u_i)} \\
&\quad \times \alpha_3(T_{di}|\tau_i, u_i)^{\delta_{di}} \, \mathrm{e}^{-\Lambda_3(T_{di}|\tau_i, u_i)} \, f_Y(Y_i|\tau_i, u_i) \, f_\tau(\tau_i) \, f_u(u_i) \mathrm{d}\tau_i \, \mathrm{d}u_i,
\end{aligned}
$$

where $f_u(u_i)$ is a multivariate Gaussian density with mean 0 and variance G and $f_\tau(\tau_i)$ the density for τ_i:

$$f_\tau(\tau_i) = \alpha_{01}(\tau_i)\mathrm{e}^{-\Lambda_{01}(\tau_i)}.$$

The function $f_Y(Y_i|\tau_i, u_i)$ is a multivariate Gaussian density with mean and variance given by

$$E(Y_{ij}|\tau_i, u_i) = \beta_{0i} + \beta_{1i}(t_{ij} - \tau_i) + \beta_{2i}(t_{ij} - \tau_i)\mathrm{trn}(t_{ij} - \tau_i; \gamma) \quad \text{and } V_i = \mathrm{Var}(Y_i|\tau_i, u_i) = \sigma_\epsilon^2 I_{n_i}.$$

This likelihood is developed in the Web Supplementary Material A of the supplementary material available at *Biostatistics* online using a multistate model approach that is by distinguishing all the different trajectories between the 4 states. In particular, we account for the fact that, due to interval censoring of the disease, a subject who was free of disease at the last visit and then died may have passed by the state illness between the last visit and the death. Then it is demonstrated that the corresponding terms in the likelihood vanish, and we obtain likelihood (3.1) thanks to the independence between the risk of death and the state conditionally on the current marker value. Thus, the main bias associated with interval censoring of illness diagnosis is addressed without complicating the likelihood.

Most studies include only subjects alive and not already diagnosed at the beginning of the study. This induces left truncation that must be taken into account in the likelihood by dividing the latter by the joint probability to be alive and illness-free at the age of entry into the study T_{0i}:

$$l(\theta) = \log \prod_{i=1}^{N} \frac{L_i(Y_i, T_{ei}, \delta_{ei}, T_{di}, \delta_{di}; \theta)}{\Pr(T_{ei}^* > T_{0i}, T_{di}^* > T_{0i})} \qquad (3.2)$$

with

$$\Pr(T_{ei}^* > T_{0i}, T_{di}^* > T_{0i}) = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} f_\tau(\tau_i) \, f_u(u_i) \mathrm{e}^{-\mathbb{1}_{\{T_{0i} > \tau_i\}} \Lambda_{12}(T_{0i} - \tau_i|u_i)} \, \mathrm{e}^{-\Lambda_3(T_{0i}|\tau_i, u_i)} \mathrm{d}\tau_i \, \mathrm{d}u_i.$$

Parameters were estimated by the maximum likelihood using a Newton–Raphson–like optimization algorithm detailed in the Web Supplementary Material B of the supplementary material available at *Biostatistics* online and Gaussian quadrature for computing the integrals. Variance of the estimates were obtained by the inverse of the observed Hessian matrix at convergence. A simulation study presented in the Web Supplementary Material C of the supplementary material available at *Biostatistics* online showed the good behavior of these estimates. In particular, they were less biased than estimates obtained by considering death as a noninformative censoring.

4. APPLICATION

4.1 *The PAQUID data set*

The PAQUID cohort is a French prospective study on cognitive aging including 3777 subjects aged 65 years and older and living at home at the initial visit. Subjects were initially interviewed at home in 1988

and 1, 3, 5, 8, 10, 13, and 15 years later. Initial measurements were excluded because of the learning effect previously described between the first 2 exams (Jacqmin-Gadda *and others*, 1997). Dementia diagnosis is assessed at each visit by the investigating psychologist and then confirmed by clinical examination by a neurologist if the subjects were screened positive by the psychologist. The time basis used for the analyses is age. Dementia age is computed as the mean between the age at dementia diagnosis and the age at the last visit without dementia. Age of censoring for dementia is the age at the last follow-up visit for the subject. As we aimed at estimating pre-diagnosis cognitive decline, the cognitive measures collected after the diagnosis were excluded. This exclusion was also justified by the high rate of missing cognitive measures after dementia diagnosis in Paquid, and by the fear that the assumption of a long-term linear trend post-diagnosis was not appropriate. We also excluded 102 subjects diagnosed as demented at baseline. Thus, the left truncation has to be taken into account. Vital status was collected on every participant (including refusals and dropouts) by phone contact and, if necessary, age at death was obtained from the general practitioner. Age at death was censored at the end of follow up, that is, age at entry plus 15.

The psychometric test used was the Benton Visual Retention Test (Benton, 1965) which evaluates visual memory with scores ranging from 0 to 15. In this analysis, we compared the characteristics of the pre-diagnosis phase according to the educational level considered in 2 categories: no education or primary school level without diploma (denoted PSD−) versus primary school diploma or higher level (denoted PSD+). The samples included 2396 PSD+ and 1279 PSD− nondemented at the initial visit. The mean age of entry into the study was, respectively, 74.5 years and 76.7 years for PSD+ and PSD−. The mean number of measurements before dementia diagnosis was 2.88 for PSD+ (first quartile: 1, median: 2, third quartile: 5) and 1.77 for PSD− (first quartile: 0, median: 1, third quartile: 3). We included subjects with no measure (23% of PSD+ vs. 38% of PSD−) because they contributed to the death risk assessment. During the follow up, 365 PSD+ and 296 PSD− were diagnosed as demented with a mean dementia age of 83.66 years (standard error [SE] = 5.88) and 83.56 years (SE = 5.82); and 1437 PSD+ and 893 PSD− died with a mean age at death of 84.69 years (SE = 7.07) and 86.03 years (SE = 6.68).

4.2 *Model*

The main study was conducted on the 2 separated samples defined by educational level. The baseline transition intensity from state 0 to state 1 and from state 1 to state 2 was defined by Weibull distributions. The baseline intensity for death was a stepwise function with 7 steps (every 5 years). The number of steps was chosen by Akaike information criterion (AIC) between severals models with 5, 6, or 7 steps and different lengths of step. Following results of the simulation study, we decided to use 20 quadrature points for the integration to estimate the models. However, due to computation time, it was not possible to estimate a model with 4 random effects using 20 quadrature points on these samples. Thus, we first selected the best model including only 3 random effects: 5 structures with 2 or 3 random effects were compared by AIC, all including the random changepoint τ_i and a random effect on the score level at τ_i, β_{0i} (model 1.a). Models 1.b to 1.e included a third random effect either on the mean slope (model 1.b), or on the difference between the slopes (model 1.c), or on the slope in the second phase (models 1.d and 1.e). In model 1.e, the dementia transition intensity depended on the random slope in the second phase. Finally, the best 3 random-effect model selected by AIC was model 1.d including the random slope for the second phase but no dependence from the dementia transition intensity on this random slope:

$$\begin{aligned} Y_i(t) &= \beta_{0i} + \beta_1(t-\tau_i) + \beta_{2i}\sqrt{(t-\tau_i)^2+\gamma} + \epsilon_i \\ &\quad \text{with } \beta_{0i} = \phi_0 + u_{0i}, \quad \beta_{1i} = \phi_1 + \frac{u_{s_2 i}}{2}, \quad \beta_{2i} = \phi_2 + \frac{u_{s_2 i}}{2} \quad \text{and } \gamma = 0.1, \\ \alpha_{01}(t) &= \lambda_L \kappa_L (\kappa_L t)^{\lambda_L - 1} \text{ with } \kappa_L > 0,\ \lambda_L > 0, \end{aligned}$$

$$\alpha_{12}(t-\tau_i) = \lambda_e \kappa_e (\kappa_e (t-\tau_i))^{\lambda_e - 1} \text{ with } \kappa_e > 0,\ \lambda_e > 0,$$
$$\alpha_3(t|\tau_i, u_i) = a_l \, \mathrm{e}^{\eta \tilde{Y}_i(t|\tau_i, u_i)} \text{ if } T_l \leqslant t < T_{l+1}, \quad l = 1, \ldots, 7$$
$$\text{with } a_l > 0, \quad T_1 = 65, \quad T_2 = 70, \quad T_3 = 75, \quad T_4 = 80, \quad T_5 = 85,$$
$$T_6 = 90 \quad \text{and } T_7 = 95.$$

We reestimated model 1.d with 9 quadrature points and we compared it with the full model including 4 random effects. For PSD−, the 3 random-effect model had a slightly better AIC (AIC = 18 321 for 3 random-effect model vs. AIC = 18 330 for 4 random-effect model), while, for PSD+, the 4 random-effect model had a slightly better AIC (AIC = 41 852 for 3 random-effect model vs. AIC = 41 842 for 4 random-effect model). However, using the D criteria proposed by Commenges *and others* (2008) that estimates the Kullback–Leibler distance between the models, these differences in AIC are classified as small for PSD− ($D = 2.00 \times 10^{-3}$) and negligible for PSD+ ($D = -7.23 \times 10^{-4}$). Moreover, parameters estimates and standard errors from the 2 models were close. Thus, we retained the model 1.d with 3 random effects.

4.3 *Results*

From the stratified analysis, the joint model estimates for the 2 educational level are presented in Table 1. The mean Benton score at entry into the pre-diagnosis state was 10.74 (SE = 0.07) for the PSD+, whereas it was 8.88 (SE = 0.14) for the PSD−. In the healthy phase, cognitive decline was slower for PSD+ ($p_1 = -0.069$ points each year, SE = 0.005) than for PSD− ($p_1 = -0.091$ points each year, SE = 0.013),

Table 1. *Mean estimates (ME), standard errors (SEs) for the joint model for cognition, dementia, and death (model 1.d), and the joint model assuming independence between cognition and death (model 2), stratified on educational level from PAQUID data ($N = 2396$ for PSD+ and $N = 1279$ for PSD−)*

Parameters	Model 1.d				Model 2			
	PSD+		PSD−		PSD+		PSD−	
	ME	SE	ME	SE	ME	SE	ME	SE
ϕ_0	10.741	0.068	8.885	0.140	10.821	0.067	9.162	0.137
ϕ_1	−0.373	0.017	−0.249	0.016	−0.323	0.015	−0.209	0.013
ϕ_2	−0.304	0.018	−0.158	0.020	−0.258	0.016	−0.133	0.018
$\sqrt{\lambda_e}$	1.242	0.044	1.410	0.126	1.207	0.051	1.408	0.092
$\sqrt{\kappa_e}$	0.427	0.012	0.401	0.013	0.414	0.012	0.365	0.012
$\sqrt{a_1}$	0.306	0.034	0.211	0.053	—	—	—	—
$\sqrt{a_2}$	0.299	0.027	0.360	0.046	—	—	—	—
$\sqrt{a_3}$	0.377	0.029	0.428	0.047	—	—	—	—
$\sqrt{a_4}$	0.473	0.033	0.531	0.052	—	—	—	—
$\sqrt{a_5}$	0.605	0.039	0.659	0.061	—	—	—	—
$\sqrt{a_6}$	0.759	0.047	0.803	0.074	—	—	—	—
$\sqrt{a_7}$	0.864	0.063	1.059	0.106	—	—	—	—
η	−0.132	0.013	−0.174	0.024	—	—	—	—
$\sqrt{\lambda_L}$	3.881	0.008	3.396	0.058	3.908	0.043	3.097	0.047
$\sqrt{\kappa_L}$	0.106	0.0001	0.109	0.0001	0.106	0.0001	0.109	0.0002
$E_{01}(t)$	86.13	0.038	81.11	0.229	86.31	0.041	80.06	0.221
$E_{12}(t-\tau_i)$	4.93	0.271	5.51	0.374	5.29	0.308	6.63	0.431
p_1	−0.069	0.005	−0.091	0.013	−0.066	0.005	−0.076	0.013
p_2	−0.678	0.034	−0.407	0.034	−0.581	0.031	−0.342	0.029

whereas the acceleration was greater for PSD+ ($p_2 = -0.68$ points each year, SE $= 0.034$) than for PSD− ($p_2 = -0.41$ points each year, SE $= 0.034$) in the pre-diagnosis phase. The expected age at entry into the pre-diagnosis phase was later for PSD+ ($E_{01}(t) = 86.13$ years, SE $= 0.04$) than for PSD− ($E_{01}(t) = 81.11$ years, SE $= 0.23$), while the expected transition time between the pre-diagnosis state and the dementia state was slightly earlier for PSD+ ($E_{12}(t - \tau_i) = 4.93$ years, SE $= 0.27$) than for PSD− ($E_{12}(t - \tau_i) = 5.51$ years, SE $= 0.37$).

Figure 2 displays the estimated transition intensities. Figure 2A shows the transition intensity from the healthy state to the pre-diagnosis state is higher for PSD−. Figure 2B shows nonproportionality of transition intensity to dementia state between the 2 educational level (justifying the stratified analysis): the risk is higher for PSD+ before 5 years and lower after. Figures 2C and D displays the transition intensity to death for each educational level given different values of τ_i: the median ($\tau_i = 87.04$ years for the PSD+ and $\tau_i = 82.12$ years for the PSD−), $\tau_i = 65$ years to represent a subject always in the pre-diagnosis phase during the follow up, and $\tau_i = 100$ years to represent a subject always in the healthy phase. This figure illustrates the association between cognitive state and risk of death. Moreover, the estimated hazard ratio of death for 5 points lost in the Benton score is 1.93 for PSD+ and 2.39 for PSD−.

Figure 3 displays the estimated marginal mean scores $E(Y(t))$ given the age and also the estimated means of $Y(t)$ given that the subject is alive at age t ($E(Y(t)|T_d^* > t)$), and given that the subject is

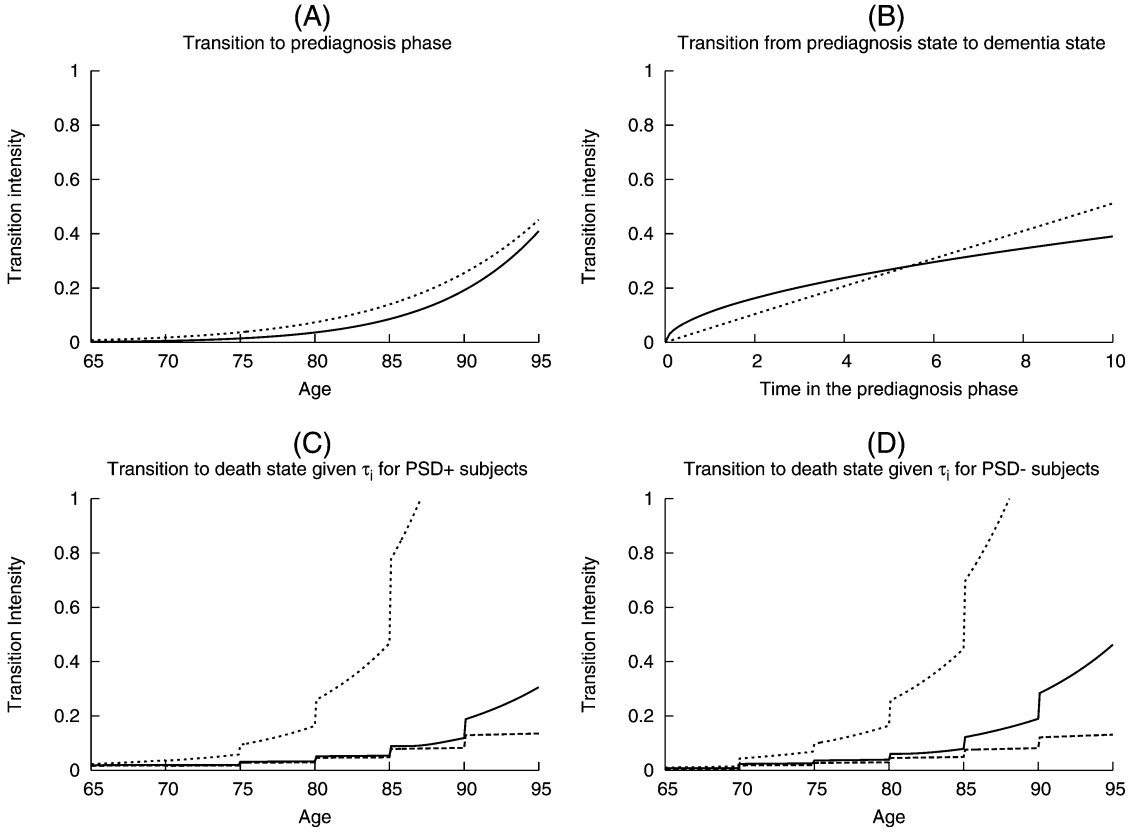

Fig. 2. Transition intensities from state 0 to state 1 (A), from state 1 to state 2 (B) for PSD+ (thick plain line) and PSD− (thick dashed line), for death separately for PSD+ (C) and PSD− (D) given different values of τ_i: τ_i = median ($\tau_i = 87.04$ for PSD+ and $\tau_i = 82.12$ for PSD−) (plain lines), $\tau_i = 65$ (dotted line), $\tau_i = 100$ (dashed line).

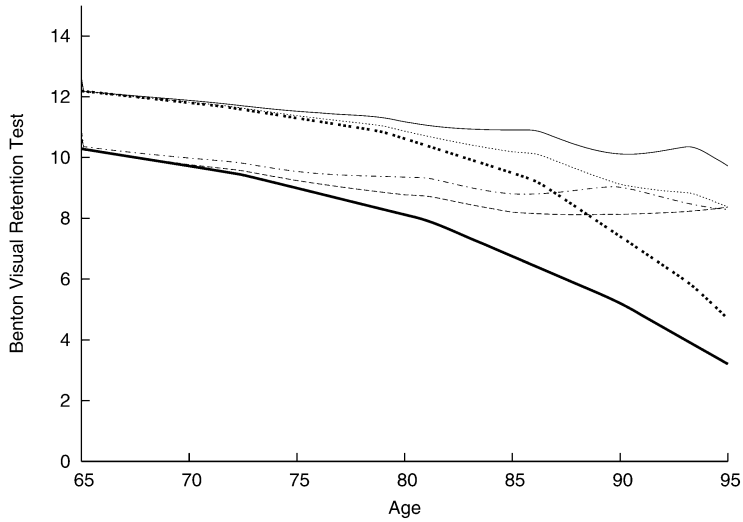

Fig. 3. Marginal mean scores $E(Y(t))$ given age for PSD+ (thick dotted line) and PSD− (thick plain line) and several mean trajectories given information on dementia age and death age: $E(Y(t)|T_d^* > t)$ for PSD+ (thin dotted line) and for PSD− (thin dashed line), $E(Y(t)|T_d^* > t, T_e^* > t)$ for PSD+ (thin plain line) and for PSD− (thin dashed-dotted line).

alive and nondemented at age t ($E(Y(t)|T_d^* > t, T_e^* > t)$). Formulas to estimate these conditional means are given in the Web Supplementary Material D of the supplementary material available at *Biostatistics* online. The estimated mean score is always higher for PSD+ but the decline is more pronounced compared to oldest PSD−. Another interesting result is that among subjects surviving in oldest ages, the difference in cognitive level between PSD+ and PSD− vanishes. This illustrates once again the strong link between death and cognition and thus the selection phenomenon by death especially among subjects with low education.

For purposes of comparison, we estimated in each sample another model, denoted model 2, that was identical to model 1.d but assuming independence between cognition and time-to-death ($\eta = 0$). Estimates were close to those of the model 1.d, but the estimated rate of decline in the pre-diagnosis phase was lower ($p_2 = -0.58$ points each year for PSD+, SE $= 0.031$ and $p_2 = -0.34$ points each year for PSD−, SE $=$ 0.029), and the time lag between state 1 and state 2 was longer ($E_{12}(t - \tau_i) = 5.29$ years, SE $= 0.30$ for PSD+ and $E_{12}(t - \tau_i) = 6.63$ years, SE $= 0.43$ for PSD−). This is in agreement with the previously reported steeper cognitive decline in late pre-death period (Wilson *and others*, 2003).

4.4 *Goodness-of-fit*

We evaluated the fit of the model separately for the 3 outcomes (cognitive scores, age at dementia, and death). First, for nondemented and demented subjects, we compared the means of the posterior score expectations $\hat{Y}(t) = E(Y(t)|\hat{\tau}_i, \hat{u}_i)$ and the means of the observed cognitive scores using 5-year age intervals (Figure 4) for the 2 educational levels. The random-effect predictions $\hat{\tau}_i, \hat{u}_i$ were computed by the mode of their posterior distribution given the data $Y_i, T_{ei}, \delta_{ei}, T_{di}, \delta_{di}$. The Web Figure 1 of the supplementary material available at *Biostatistics* online displays the observed data and individual trajectories estimated by $E(Y(t)|\hat{\tau}_i, \hat{u}_i)$ for 16 selected subjects among the 2 groups. These figures show that the model captures well the different evolutions of demented and nondemented subjects and the individual evolution shapes.

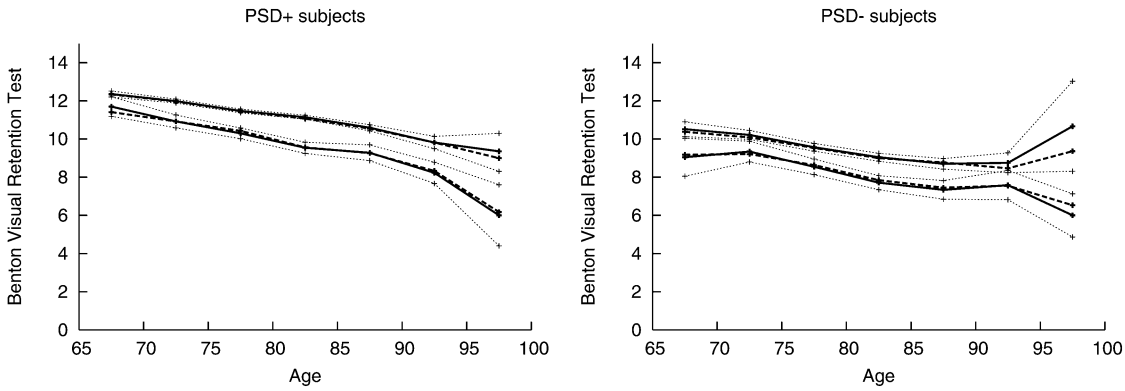

Fig. 4. Posterior evolution (thick dashed lines) compared to empirical mean evolution (thick plain lines) for nondemented subjects (upper curves) and demented subjects (lower curves) and their confidence interval (thin dotted lines) for PSD+ and PSD−.

To evaluate the fit of the death data, we compared the marginal survival function from the joint model estimated by

$$S_d(t) = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \mathrm{e}^{-\Lambda_3(t|\tau_i,u_i)} f_\tau(\tau_i) f_u(u_i) \mathrm{d}\tau_i \, \mathrm{d}u_i \tag{4.1}$$

with the death–survival function estimated using a nonparametric proportional hazard model and a penalized likelihood approach (PHMPL) (Joly *and others*, 1999) (Web Figure 2 of the supplementary material available at *Biostatistics* online).

To check the parametric assumption regarding baseline transition intensities, it was not possible to compare the estimates for the risk of dementia obtained from the complete model with those obtained with PHMPL because PHMPL considers dead subjects as censored at their last visit. Thus, we compared PHMPL estimates of the dementia survival function with the same function estimated by model 2 (Web Figure 2 of the supplementary material available at *Biostatistics* online), where dead subjects were treated as censored, using the following formula:

$$\begin{aligned} S_e(t) &= 1 - P(T_e^* < t) \\ &= 1 - \int_{-\infty}^{+\infty} f_u(u_i) \left\{ \int_0^t \mathrm{e}^{-\Lambda_{01}(\tau_i)} \alpha_{01}(\tau_i) \int_{\tau_i}^t \mathrm{e}^{-\Lambda_{\{\upsilon>\tau_i\}}\Lambda_{12}(\upsilon-\tau_i)} \alpha_{12}(\upsilon-\tau_i) \mathrm{d}\upsilon \, \mathrm{d}\tau_i \right\} \mathrm{d}u_i . \end{aligned} \tag{4.2}$$

This figure shows an appropriate adjustment for the transition to death and the transition to dementia.

5. Discussion

We have developed a joint model with latent state for a longitudinal process and illness–death data. Such data are frequent in the monitoring of chronic diseases. This new model makes it possible to describe the pre-diagnosis phase of disease while limiting the bias in parameter estimation. The application shows that the model is particularly well suited for studying cognitive decline in the pre-diagnosis phase. It may be viewed as an improvement on the random changepoint model proposed by Jacqmin-Gadda *and others* (2006) since it improves the handling of informative right censoring due to death and better accounts for the succession of events in time. In the context of cognitive aging, some authors consider that MCI is a transitional state between healthy and dementia without reversibility (Petersen *and others*, 1999), while others thinks it is a risk factor for dementia and may be reversible (Larrieu *and others*, 2002; Ritchie *and*

others, 2001). In this work, the pre-diagnosis phase corresponds to the former concept of MCI. Another subject of divergence in the literature regards the cognitive tests and the thresholds used to define MCI. We think it is more rational to define it as a change in the slope of decline instead of a threshold on a cognitive score, as the cognitive levels are highly variable in the population irrespective of any pathological process.

The application focused on the role of educational level. Our previous work has shown that pre-diagnosis cognitive evolution exhibits a clear phase of accelerated decline that is well approximated by a segmented linear–linear model (see Figure 1 in Jacqmin-Gadda *and others*, 2006). The estimated age at entry into the accelerated decline phase was later in the previous model (90.3 for PSD+ vs. 87.3 years for PSD−), but the meaning of this phase is different in the 2 models. Here, this is an obligatory transition state before dementia (with a null risk of dementia before this state), while, in Jacqmin-Gadda *and others* (2006) and Yu and Ghosh (2010), dementia may arise before τ_i and the risk of dementia does not increase with the time spent in the second phase. It would be interesting to test the existence of this pre-diagnosis phase of accelerated decline in different populations, for various cognitive tests or according to the type of dementia (Alzheimer vs. vascular dementia for instance). However, the development of such a test would require modeling more flexibly the dependence between risk of dementia and cognitive scores that could be limited by numerical difficulties.

In the proposed joint model, the transition intensities from the 3 transient states to the death state were identical after adjustment on the current cognitive score. With this model, we demonstrate in the Web Supplementary Material A of the supplementary material available at *Biostatistics* online that the likelihood accounts for possible unobserved transitions to dementia between the last visit and death without any complication. This is important because this is the most serious bias due to intermittent observations. An alternative model would include 3 state-specific transition intensities to death. However, this would seriously complicate the likelihood due to interval censoring of dementia, and convergence problems could be observed if few deaths were observed before the changepoint τ_i. Moreover, in our opinion, a model with 3 state-specific transition intensities is less flexible than dependence on the quantitative cognitive score, which changes with time.

While we have checked the parametric assumptions of the model for the PAQUID data, it would be useful to relax some of them. For instance, a cure model allowing a subpopulation to have a null risk of dementia could be investigated as in Yu and Ghosh (2010). Alternatively, flexible distribution, such as the semi-nonparametric distribution of Gallant and Nychka (1987), would be interesting for age at entry into the pre-diagnosis phase since the assumption regarding this distribution cannot be evaluated by comparison with observed data. Once again, such extensions are limited by computation time. Improvement in integration and optimization algorithms would be necessary to allow these extensions of the model.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENT

Conflict of Interest: None declared.

REFERENCES

AMIEVA, H., LE GOFF, M., MILLET, X., ORGOGOZO, J. M., PÉRÈS, K., BARBERGER-GATEAU, P., JACQMIN-GADDA, H. AND DARTIGUES, J.-F. (2008). Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. *Annals of Neurology* **64**, 492–498.

BACON, D. W. AND WATTS, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* **58**, 525–534.

BENTON, A. L. (1965). *Manuel pour l'application du Test de Retention Visuelle. Applications cliniques et expérimentales*, 2ème édition. Paris, française: Centre de Psychologie Appliquée.

COMMENGES, D., SAYYAREH, A., LETENNEUR, L., GUEDJ, J. AND BAR-HEN, A. (2008). Estimating a difference of Kullback-Leibler risks using a normalized difference of AIC. *The Annals of Applied Statistics* **2**, 1123–1142.

DOMINICUS, A., RIPATTI, S., PEDERSEN, N. L. AND PALMGREN, J. (2008). A random change point model for assessing variability in repeated measures of cognitive function. *Statistics in Medicine* **27**, 5786–5798.

DUBOIS, B., FELDMAN, H. H., JACOVA, C., DEKOSKY, S. T., BARBERGER-GATEAU, P., CUMMINGS, J., DELACOURTE, A., GALASKO, D., GAUTHIER, S., JICHA, G. *and others* (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology* **6**, 734–746.

ELASHOFF, R. M., LI, G. AND LI, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64**, 762–771.

GALLANT, A. R. AND NYCHKA, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrika* **55**, 363–390.

HALL, C. B., YING, J., KUO, L. AND LIPTON, R. B. (2003). Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics and Data Analysis* **42**, 91–109.

HAN, J., SLATE, E. H. AND PEÑA, E. A. (2007). Parametric latent class joint model for a longitudinal biomarker an recurrent events. *Statistics in Medicine* **26**, 5285–5302.

HOUGAARD, P. (1999). Multi-state models. A review. *Life Time Data Analysis* **5**, 239–264.

JACQMIN-GADDA, H., COMMENGES, D. AND DARTIGUES, J.-F. (2006). Random changepoint model for joint modeling of cognitive decline and dementia. *Biometrics* **62**, 254–260.

JACQMIN-GADDA, H., FABRIGOULE C., COMMENGES D. AND DARTIGUES J.-F. (1997). A five year longitudinal study of mini-mental state examination in normal aging. *American Journal of Epidemiology* **145**, 498–506.

JOLY, P., COMMENGES, D., HELMER, C. AND LETENNEUR, L. (2002). A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics* **3**, 433–443.

JOLY, P., LETENNEUR, L., ALIOUM, A. AND COMMENGES, D. (1999). PHMPL: a computer program for hazard estimation using a penalised likeihood method with interval-censored and left-truncated data. *Computer Methods and Programs in Biomedicine* **60**, 225–231.

LARRIEU, S., LETENNEUR, L., ORGOGOZO, J. M., FABRIGOULE, C., AMIEVA, H. LE CARRET, N., BARBERGER-GATEAU, P. AND DARTIGUES, J.-F. (2002). Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. *Neurology* **59**, 1594–1599.

PETERSEN, R. C., SMITH, G. E., WARING, S. C., IVNIK, R. J., TANGALOS, E. G. AND KOKMEN, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology* **56**, 303–308.

RITCHIE, K., ARTERO, S. AND TOUCHON, J. (2001). Classification criteria for mild cognitive impairment: a population-based validation study. *Neurology* **56**, 37–42.

SEBER, G. A. F. AND WILD, C. J. (2003). Non linear regression. Wiley series in probability and mathematical statistics. New York: Wiley & Sons.

SMALL, B. J., FRATIGLIONI, L., VON STRAUSS, E. AND BCKMAN, L. (2003). Terminal decline and cognitive performance in very old age: does cause of death matter ? *Psychology and Aging* **18**, 193–202.

TSIATIS, A. A. AND DAVIDIAN, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* **14**, 809–834.

WILSON, R. S., BECKETT, L. A., BIENAS, J. L., EVANS, D. A. AND BENNETT, D. A. (2003). Terminal decline in cognitive function. *Neurology* **60**, 1782–1787.

YU, B. AND GHOSH, P. (2010). Joint modeling for cognitive trajectory and risk of dementia in the presence of death. *Biometrics* **66**, 294–300.

[Received January 11, 2010; revised January 18, 2011; accepted for publication January 20, 2011]