Random change point model for joint modeling of cognitive decline and dementia.

Hélène Jacqmin-Gadda, Daniel Commenges, Jean-François Dartigues

Random changepoint model for joint modeling of cognitive decline and dementia

Hélène Jacqmin-Gadda,[1,2] **Daniel Commenges**[1,2] **and**

Jean-François Dartigues[2,3]

[1]INSERM, Equipe de biostatistique E0338, Bordeaux, F-33076 France

[2] Université Victor Segalen Bordeaux 2, Bordeaux, F-33076 France

[3]INSERM, U593, Bordeaux, F-33076 France

Corresponding author :

Hélène Jacqmin-Gadda, INSERM E0338,

Université Victor Segalen Bordeaux 2, case 11,

146 rue Léo Saignat, 33076 Bordeaux cedex, France.

Tel : (33) 5 57 57 45 18 ; Fax (33) 5 56 24 00 81 ;

helene.jacqmin-gadda@bordeaux.inserm.fr

1

Summary. We propose a joint model for cognitive decline and risk of dementia to describe the pre-diagnosis phase of dementia. We aim to estimate the time when the cognitive evolution of subjects in the pre-dementia phase becomes distinguishable from normal evolution and to study whether the shape of cognitive decline depends on educational level. The model combines a piecewise polynomial mixed model with a random changepoint for the evolution of the cognitive test and a log-normal model depending on the random changepoint for the time to dementia. Parameters are estimated by maximum likelihood using a Newton-Raphson-like algorithm. The expected cognitive evolution given age to dementia is then derived and the marginal distribution of dementia is estimated to check the log-normal assumption.

Key words: Joint model ; Mixed model ; Longitudinal data ; Random effects

2

1. Introduction

Dementia is a progressive disease defined by cognitive impairment in memory and at least one other cognitive function with consequences for the activities of daily living. Several studies have shown that cognitive impairment is present long before all criteria for dementia diagnosis are fulfilled (Dartigues et al., 1997; Masur et al., 1994). However, little is known about the shape of this decline and about the time at which cognitive evolution of subjects who develop dementia becomes distinguishable from that of normal elderly subjects.

Using a piecewise linear mixed model, Hall et al. (2000) have compared the evolution of a cognitive test for incident cases of dementia in the years preceding the diagnosis and for subjects free of dementia at their last follow-up. However, these analyses may be biased by right-censoring of dementia (subjects without dementia at their last visit may be in preclinical phase of dementia), and by loss of follow-up which may be associated with poor cognitive functioning and high risk of dementia (Jacqmin-Gadda et al., 1997). More recently, Hall *et al.* (2003) have proposed a random changepoint model to describe the cognitive decline of demented subjects but, as time-to-dementia was not jointly modelled, parameters were estimated using only data from subjects diagnosed as demented during the follow-up. This reduces the power of the study, does not allow the comparison of normal and pathological aging and does not avoid selection biases (subjects must be seen as demented before the end of the follow-up to be included in the sample).

To deal with the above problems, the aim of this paper is to propose a joint model for time-to-event and repeated measures of a marker (Wulfsohn

3

and Tsiatis, 1997, Henderson, Diggle et Dobson, 2000) to describe the cognitive decline in the pre-diagnosis stage of dementia and, especially, to estimate the time between the acceleration of the cognitive decline and the diagnosis of dementia. Another important point is to study whether the shape of the cognitive decline before dementia depends on the educational level of the subject. We have previously proposed another joint model for cognitive decline and dementia using a latent stochastic process which represents the cognitive ability and defining dementia as the crossing of a barrier by the latent process (Hashemi, Jacqmin-Gadda and Commenges, 2003). However, in this model, the mean evolution was assumed to be linear and common for future demented subjects and future non demented subjects ; this is not suitable for the study of the accelerated cognitive decline in the pre-diagnosis stage of dementia. In the present work, to take account of non-linearity of the cognitive decline, we combined a piecewise polynomial mixed model with random changepoint for the evolution of the cognitive test and a log-normal model depending on the random changepoint for the time to dementia diagnosis.

The model we propose can be viewed as an extension of the joint model proposed by DeGruttola and Tu (1994) which combined a linear mixed model for the marker and a log-normal survival model. In another context, Faucett et al. (2002) and Pauler and Finkelstein (2002) have proposed combining a piecewise linear mixed model with a random changepoint and a Cox proportional hazards model in a joint model estimated using Markov Chain Monte-Carlo (MCMC). Both papers focussed on survival analysis because the main objective was to predict (or impute) failure time using information given by the marker trajectory. Here we are mainly interested in the longi-

4

tudinal trajectory given the failure time which would be difficult to derive using a semi-parametric survival model. Moreover, we estimate parameters by a direct likelihood approach rather than MCMC and take left-truncation of the data into account.

In Section 2 we introduce the joint model, in section 3 the likelihood and the estimation algorithm and section 4 is devoted to post-fit estimations. In section 5 we apply the methodology to the Paquid cohort to estimate the evolution of the Benton Visual Retention Test (BVRT) in the pre-diagnosis phase of dementia and section 6 concludes with a discussion.

2.   Joint Model

Let $Y_i(t)$ be the cognitive test score of subject i at age t. This is the primary time scale in this analysis because this is the greatest risk factor for dementia. For brevity, we denote $Y_{ij} = Y_i(t_{ij})$ for $j = 1, ..., n_i$, and $i = 1, ..., N$ where N is the number of subjects, t_{ij} is the age of subject i at measurement j and Y_i is the vector of the n_i measurements for subject i. We assume a segmented mixed model for Y_{ij} with a linear trend before the changepoint and a polynomial trend thereafter. This model has a smooth transition at the changepoint which agrees with the clinical belief of a progressive decline in the pre-diagnosis phase of dementia. The linear assumption for normal aging and the non-linear assumption for pathological aging is supported by previous analyses from the Paquid cohort (Jacqmin-Gadda et al.,1997, Amieva *et al.*, 2005). Moreover, the proposed model has the desirable property that the derivative with respect to each parameter is a continuous function of t (Seber and Wild, 1989, chap 9) :

$$Y_{ij} = (\mu_0 + u_{0i}) + (\mu_1 + u_{1i}) \times t_{ij} + \Sigma_{k=2}^{K}(\mu_k + u_{ki}) \times \{(t_{ij} - \tau_i)^+\}^k + \epsilon_{ij} \quad (1)$$

5

where $z^+ = 0$ if $z \leq 0$ and $z^+ = z$ if $z > 0$. The independent random error ϵ_{ij} is $N(0, \sigma^2)$ and the vector $U_i = (u_{0i}, u_{1i}, ..., u_{Ki})'$ of random effects is $N(0, G)$. The individual random changepoint τ_i is the age at which cognitive decline of subject i accelerates : the individual slope before τ_i is $\mu_1 + u_{1i}$ and the individual coefficients of the polynomial curve after τ_i are $\mu_k + u_{ki}$. The mean intercept, slope and polynomial coefficients may depend on fixed covariates : $\mu_k = \beta_k + Z_{ki}\alpha_k$, where the Z_{ki} are row vectors of size p_k. We assume that the random changepoint τ_i is independent from U_i and has a lognormal distribution with mean depending on a row vector of covariates $Z_{\tau i}$ of size p_τ :

$$\log(\tau_i) \sim N\left(Z_{\tau i}\alpha_\tau, \sigma_\tau^2\right) \tag{2}$$

Set $X_i = \log(T_i)$ the logarithm of age at dementia for subject i and Z_{xi} a row vector of covariates. The model for age at dementia is :

$$X_i = \log(T_i) = Z_{xi}\gamma + \eta\log(\tau_i) + \varepsilon_i \tag{3}$$

with $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Thus, given the age at acceleration of cognitive decline (τ_i), the median age at dementia diagnosis is :

$$\mathrm{med}(T_i | \tau_i) = \tau_i^\eta \exp(Z_{xi}\gamma)$$

We make the following remarks about this model. First, it does not impose a pre-diagnosis acceleration of cognitive decline : some subjects may reach the criteria for dementia diagnosis after a steady linear decline of their cognitive functions rather than a two-phase decline. Second, the desirable property that the median age at dementia increases with the age at acceleration of cognitive decline is satisfied only if $\eta > 0$. When $\eta = 1$, $\mathrm{med}(T_i | \tau_i)$ is proportional to

6

τ_i; when $\eta < 1$ (respectively $\eta > 1$), the ratio $\text{med}(T_i|\tau_i)/\tau_i$ is a decreasing (respectively increasing) function of τ_i.

3. Likelihood

Set $\tilde{X}_i = min(X_i, C_i)$ with C_i the logarithm of age at the end of follow-up for subject i and $\delta_i = 1_{\{X_i \leq C_i\}}$, the failure indicator for dementia. We assume that the censoring process for time to dementia is non informative and that missing responses for the cognitive test are ignorable when they are not due to dementia. More specifically, for a subject in the study at age t, the probability of dropout from the study and the probability of obtaining a cognitive test score at this time are both conditionally independent on the time to dementia, the current value of the score and the random effects given the past observed values of the marker. Denoting by θ the vector of all the parameters in (1), (2) and (3), and taking advantage of the conditional independence of Y_i and $\tilde{X}_i$ given τ_i, the likelihood for the observed data $(Y, \tilde{X}, \delta)$ may be written :

$$L(Y, \tilde{X}, \delta; \theta) = \prod_{i=1}^{N} \int f_{Y_i|\tau_i}(y_i|\tau) f_{X_i|\tau_i}(\tilde{x}_i|\tau)^{\delta_i} (1 - F_{X_i|\tau_i}(\tilde{x}_i|\tau))^{1-\delta_i} f_{\log(\tau_i)}(\log(\tau)) d\log(\tau)$$

$$(4)$$

where $f_{Y_i|\tau_i}$ is a multivariate Gaussian density with mean and variance given by :

$$E(Y_{ij}) = \mu_0 + \mu_1 \times t_{ij} + \Sigma_{k>1}^{K} \mu_k \times \{(t_{ij} - \tau_i)^+\}^k$$

and

$$Var(Y_i) = V_i = A_i G A_i' + \sigma^2 I_{n_i}$$

and A_i is the $n_i \times (K+1)$ design matrix with rows $(1, t_{ij}, \{(t_{ij} - \tau_i)^+\}^2, ..., \{(t_{ij} - \tau_i)^+\}^K)$; $f_{X_i|\tau_i}$ and $F_{X_i|\tau_i}$ are the univariate Gaussian density and cumulative

7

distribution function defined by (3) and $f_{\log(\tau_i)}$ is the univariate Gaussian density defined by (2).

Prevalent cases of dementia (subjects already demented at the first visit) must be excluded from the analysis because of selection problems discussed below and because their age at diagnosis is unknown and they bring little information on the period before diagnosis. When excluding prevalent cases, the data are left-truncated and the likelihood to be maximised is the likelihood conditional on being free of dementia at entry in the study. If we denote te_i the logarithm of age of subject i at the first visit, the contribution to the likelihood for subject i is :

$$L(Y_i, \tilde{X}_i, \delta_i | X_i > te_i) = L(Y_i, \tilde{X}_i, \delta_i; \theta)/S_{X_i}(te_i) \tag{5}$$

where $S_{X_i}(.) = 1 - F_{X_i}(.)$ is the marginal survival function of the logarithm of time to dementia. Using (2) and (3), we find that the marginal distribution of X_i is Gaussian with mean $E(X_i) = Z_{xi}\gamma + \eta Z_{\tau i}\alpha_\tau$ and variance $\mathrm{Var}(X_i) = \eta^2 \sigma_\tau^2 + \sigma_\varepsilon^2$.

We implemented a Fortran program to estimate the parameters by maximising the logarithm of the likelihood (5) using the Maquardt optimisation algorithm (Maquardt, 1963). When $n_i = 0$, the contribution to the likelihood for subject i is either $f_{X_i}(\tilde{x}_i)$ or $1 - F_{X_i}(\tilde{x}_i)$ and thus has a closed form. In the other cases, the integral in (4) is computed by Gauss-Hermite quadrature with 50 points. Derivatives are computed by finite difference. Variance of parameter estimates are estimated by the inverse of the Hessian matrix computed at the optimum. A Cholesky decomposition of G is used to satisfy the constraint that G is positive definite.

4. Post-fit estimations

When parameters are estimated, we may compute easily the mean curve for the evolution of the cognitive test given the changepoint $(E(Y_i|\tau_i))$ or the survival function (or risk function) for the age at dementia given τ_i, and, using results of the previous section, the marginal survival function for dementia. To describe the evolution of cognitive function in the pre-diagnosis phase of dementia, it is interesting to estimate the mean evolution of the score given age to dementia :

$$E(Y_i(t)|X_i) = \mu_0 + \mu_1 \times t + \Sigma_{k=2}^K \mu_k \times E\{[(t - \tau_i)^+]^k | X_i\}$$

The last term may be decomposed as :

$$E\{[(t - \tau_i)^+]^k | X_i\} = \Sigma_{l=0}^k \binom{k}{l} t^{k-l} (-1)^l E\{\tau_i^l I_{\{\tau_i < t\}} | X_i\}$$

where I is the indicator function. With some additional calculations, we find that :

$$E\{\tau_i^l I_{\{\tau_i < t\}} | X_i\} = E\{\tau_i^l | X_i\} \Phi\left(\frac{\log(t) - \mu_c - l\sigma_c^2}{\sigma_c}\right)$$

where Φ is the standard normal distribution function and μ_c and σ_c^2 are the mean and variance of the Gaussian conditional distribution of $\log(\tau_i)$ given X_i :

$$\mu_c = E(\log(\tau_i)|X_i) = Z_{\tau i} \alpha_\tau + \eta \sigma_\tau^2 (\text{var}(X_i))^{-1} (X_i - E(X_i))$$

and

$$\sigma_c^2 = Var(\log(\tau_i)|X_i) = \sigma_\tau^2 \left(1 - \eta^2 \sigma_\tau^2 (\text{var}(X_i))^{-1}\right)$$

and $E\{\tau_i^l | X_i\} = \exp(l\mu_c + l^2 \sigma_c^2 / 2)$.

Formulas for estimating the random effects using $E(U_i|Y_i, \tilde{X}_i, \delta_i)$ and $E(\log(\tau_i)|Y_i, \tilde{X}_i, \delta_i)$ are given in the appendix.

9

5. Application

5.1 *Data*

The Paquid project is an epidemiological study on normal and pathological aging initiated in 1988 (Letenneur et al., 1994). The cohort consists of 3777 subjects aged 65 years and older and living at home at the beginning of the study. Subjects were interviewed by a psychologist at 1, 3, 5, 8, and 10 years after the baseline visit. Interviews were performed at home or at the institution if the subject had been institutionalized after the baseline visit. At each visit, a battery of neuropsychological tests was completed and dementia was diagnosed according to a two stage procedure. Subjects who met the DSM IIIR criteria for dementia (American Psychiatric Association, 1987) A, B and C (impairment of memory and at least one other cognitive function and interference with daily living) or those presenting a decline of 3 points or more on the MMSE scale since the previous visit were seen by a senior neurologist who made the final diagnosis. The Benton Visual Retention Test (BVRT) (Benton, 1965) is a visual memory test which consists in the presentation of geometric figures that subjects are asked to recognize from an array of four possibilities. Fifteen geometric figures are successively presented and the possible scores range from 0 to 15.

Subjects demented at baseline were excluded from the analysis because they are a selected sample of demented subjects : they must be still alive and non institutionalized at the time of the study and they have accepted to participate in the study while being already demented. Moreover, they bring little information on the cognitive course before the diagnosis. Measures of the BVRT collected at the baseline visit were not used because of the first-

10

passing effect previously described for the cognitive tests (Jacqmin-Gadda et al., 1997) and thus all subjects not followed-up after the baseline visit where excluded leading to a sample of 2960 subjects. In this sample, 437 subjects were diagnosed as incident cases of dementia during the 10 year period (333 Alzheimer cases, 51 vascular dementias, 36 Parkinson dementias and 17 other dementias). Measures of BVRT collected after the diagnosis of dementia were ignored because our objective was to study the pre-diagnosis phase of dementia. The age at dementia was estimated as the mean age between the last visit before diagnosis and the visit of diagnosis because it has previously been shown that this led to results similar to that obtained using method for interval censored survival data (Joly et al., 1998). Among the 2960 subjects, 344 subjects have never completed the BVRT at the 5 follow-up visits, and 649, 493, 403, 528 and 543 subjects have completed the BVRT at 1, 2, 3, 4 or 5 visits respectively. Among the 437 incident cases, 314 had at least one measure of the BVRT (136 without primary school diploma and 178 with it).

5.2 *Model*

Previous analyses performed separately for demented and non-demented subjects have suggested that subjects without dementia tended to have a slight linear decline of the BVRT score while subjects in the pre-diagnosis phase of dementia have a nonlinear decline. Thus, we used a linear-cubic model that is a model with a linear trend before the changepoint and a cubic curve thereafter. This model had a better Akaike criterion (AIC=Deviance-2#parameters=32395) than the linear-quadratic model (AIC=32457). The educational level ($Z_i=1$ if no education or primary school level without di-

11

ploma versus $Z_i{=}0$ if primary school diploma or higher level) was included as an explanatory variable in the model for the BVRT score with an interaction with all the time components, in the survival model for the log of time to dementia given the changepoint and in the mean of the changepoint. As the estimate of the variance of the random slope before the changepoint tended toward 0, this random slope was excluded from the model. The covariance matrix of the random effects G was left unspecified. The joint model may be written :

$$
\begin{aligned}
Y_{ij} &= (\beta_0 + Z_i\alpha_0 + u_{0i}) + (\beta_1 + Z_i\alpha_1) \times (t_{ij} - 65)/10 \\
&+ (\beta_2 + Z_i\alpha_2 + u_{2i}) \times \{(t_{ij} - \tau_i)^+\}^2 + (\beta_3 + Z_i\alpha_3 + u_{3i}) \times \{(t_{ij} - \tau_i)^+\}^3 + \epsilon_{ij}
\end{aligned}
\tag{6}
$$

with $(u_{0i}, u_{2i}, u_{3i})' \sim N(0, G)$, $\log(\tau_i) \sim N(\alpha_{\tau 0} + \alpha_{\tau 1}Z_i, \sigma_\tau^2)$ and

$$
X_i = \log(T_i) = \gamma_0 + \gamma_1 Z_i + \eta\log(\tau_i) + \varepsilon_i.
\tag{7}
$$

The choice of the lognormal distribution for the age at dementia was supported by preliminary analyses which showed that this distribution fitted the observed data much better than a Weibull distribution.

5.3 *Results*

Estimates for the joint model defined by (6) and (7) are presented in table 1. Low educated subjects have a lower score at 65 years (-1.85 points, p<0.01). Before the changepoint, results show a slight decline of the cognitive score for both educational levels : -0.80 points every 10 years (95% confidence interval : -0.91 ; -0.68) for subjects with high educational level (HEL) and -0.28 points ($\hat{\beta}_1 + \hat{\alpha}_1$, 95% CI : -0.61 ; 0.04) for subjects with low educational level (LEL).

12

After the changepoint, the shape of the decline was significantly different between the two educational levels (Likelihood ratio statistic for α_2 and α_3, $\chi_2^2 = 16.7$ $p < 0.001$) with a much more dramatic decline for subjects with HEL. To illustrate the differential evolution given the educational level, figure 1 displays the expected BVRT course for a subject with HEL and a changepoint at 85 years and for a subject with LEL and a changepoint at 70 years. The median age at the change of cognitive trend was very different between the two educational levels ($\alpha_{\tau 1}$=-0.26, 95% CI : -0.33 ; -0.19, p<0.001). This median was estimated to 90.3 years for HEL (95% CI : 89.3 ; 91.4) and 69.7 years for LEL with a much larger confidence interval (95% CI : 65.0 - 74.6) which underlines the difficulty to distinguish early normal and pathological decline among less educated subjects. As expected, the median age at dementia is also higher for subjects with HEL (94.1 years versus 87.9 years). However, given the age at the acceleration of the cognitive decline, the median age at dementia is higher for less educated subjects (γ_1=0.13, 95% CI : 0.075 - 0.19, p<0.001), that is the time between the changepoint and the diagnosis of dementia is longer for LEL. This is enlightened by table 2 which displays estimates of med$(T|\tau, Z)$. As expected since $\hat{\eta}$ is less than 1 ($\hat{\eta}$=0.77, 95% CI : 0.69 - 0.84), table 2 also shows that the time between the changepoint and the median age at dementia decreases with the age at changepoint.

Figure 2 displays the estimated curves of the mean BVRT score given the age at dementia for both educational levels. This curve highlights the differential evolution of the visual memory according to the educational level in the years preceding the diagnosis. To evaluate how the model fits the

evolution of incident cases of dementia, we compare on figure 3, the observed mean of the BVRT score at each visit among incident cases and the average of the expected BVRT score given the age at dementia computed for cases with available measure at the visit. All the fitted means are in the 95% confidence interval of the observed means.

Finally, to check the log-normal hypothesis for the age at dementia, we present on figure 4, the risk functions for the marginal distribution of age at dementia estimated by the joint model and estimated non-parametrically using a penalized likelihood approach (Joly, Commenges and Letenneur, 1998). This clearly shows that the log-normal model fits the data well.

5.4   *Simulation study*

We performed a simulation study to evaluate the robustness of the maximum likelihood estimators assuming normal random effects and normal error when the outcome has a discrete distribution with only 16 possible values such as the BVRT score. When the reponses were generated using the estimated model for the 2960 subjects of the Paquid sample, only 0.22% of the simulated data were negative and only 3.10 % were above 15. When the responses were generated using the same distribution truncated at 0 and 15 and then discretized, the histogram of the simulated data was close to the observed one. Comparison of MLE based on two series of 200 data sets of 1000 subjects generated using the Gaussian assumptions or using the truncated discrete distribution showed that the bias and the mean square error for the intercept of both groups and for the first slope of one group increase when the data are discrete and truncated but they remain small. The impact on the other parameter is negligible. Detailed presentation of the simulations

14

study may be found on the Biometrics web site.

6.   Discussion

We have presented a model to jointly describe the evolution of a marker and a survival time taking into account a change in the trend in the marker evolution. This model allows the pre-dementia shape of decline to differ according to educational level, and the mean cognitive evolution is estimated given the age at dementia while avoiding selection biases which arise when two groups are compared following selection according to their status at a predefined timepoint.

As noted by the Associate Editor, the interpretation of the changepoint as an acceleration of the cognitive decline is not obvious since it is not possible to constrain the individual rate of decline to increase after the changepoint. Thus, we have computed for each subject in the sample the difference between the rate of change before the changepoint and 1 year after the changepoint (details given in the appendix). When this difference is positive, the change is an acceleration of the decline. The difference is negative for only 102 subjects (3.4%), and for only 62 subjects 3 years after the diagnosis. Moreover, the mean difference at one year among the negative values is -0.014 (range : -0.078 ; -0.000020) while the mean difference for subjects with positive value is 0.247 (range : 0.00015 ; 0.426). Thus these negative values, which occur only for subjects with a low educational level, may be considered as very close to zero and do not invalidate our interpretation of the changepoint in this application. These results are in agreement with our conclusion that, among subjects with low educational level, the cognitive decline is smooth with a changepoint that is difficult to detect.

15

The parametric assumptions may be viewed as a limit of our approach. We shall examine successively the log-normal survival model, the piecewise polynomial model and the dependency between the two parts of the model. Firstly, we have shown that the log-normal assumption for the survival time may be easily checked and is suitable in our application. A possible extension of this model would be to assume that a fraction of the population is not liable to become demented such as in a cure model (Law, Taylor and Sandler, 2002). Secondly, a piecewise polynomial model with a random changepoint has been chosen to locate the time at which the cognitive decline of subjects in pre-diagnosis phase of dementia begins to be distinguishable from that of normal subjects. However, this time may be dependent from the parametric form of the model. For instance, a linear-linear model (Hall *et al*, 2000) assuming a non-smooth transition would probably lead to a later estimate of the age of acceleration of the cognitive decline but the interpretation of the changepoint would be different. Moreover, the stronger estimated decline before the changepoint for HEL compared to LEL suggests that the evolution of highly educated subjects might be better fitted with a three-phase model : normal evolution, slightly accelerated decline beginning long before dementia followed by a sharper decline just before diagnosis. However, such a three-phase model would raise numerical problems. Thirdly, the failure time depends on the evolution of the marker, only through the random changepoint. A model including dependency between the failure time and every random effects from the mixed model has been fitted, but it has led to unstable results due to numerical problems (increased size of the integral) and probably to more crucial identifiability problems.

16

Our analyses are valid under the assumption that the censoring process for time to dementia is non informative and the missing responses for the cognitive test is ignorable when they are not due to dementia. A useful improvement would be to jointly model the survival time using a multi-state model. This would allow estimation of cognitive evolution conditional on being alive and this would take account of potential informative censoring due to death. Indeed, in the present work, subjects who died are treated as dropouts after their last visit while the risk of dementia seems to be higher in the years preceding the death (Joly et al., 2002). Thus, the assumption that dropout time and time at dementia are conditionally independent given the observed values of the cognitive test may be too strong for these subjects. In other respects, if missing responses to the BVRT score are informative among demented subjects it is difficult to predict the direction of the bias. Thus, it could be interesting to extend this model to take informative missingness into account with different assumptions and perform a sensitivity analysis.

Despite limitations discussed above, joint modeling of cognitive decline and risk of dementia is appropriate for investigating hypotheses regarding the association between educational level and the pattern of decline before dementia. For instance, our results reinforce the hypothesis of a greater "reserve capacity" (Stern et al., 1994) in highly educated subjects which could explain their slight decline in the pre-diagnosis phase of dementia, followed by a faster decline just before the diagnosis when the mechanism of compensation failed.

17

7.   References

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. DSM-III-R. 3rd ed. revised. Washington DC, 1987.

Amieva H., Jacqmin-Gadda, H., Orgogozo, J.M., LeCarret, N., Helmer, C., Letenneur, L., Barberger-Gateau, P., Fabrigoule C., Dartigues, J.F. (2005). The 10-year cognitive decline before dementia of the Alzheimer type : a prospective population-based study, *Brain*, in press.

Benton, A.L. (1965). Manuel pour l'Application du Test de Retention Visuelle. Applications Cliniques et Expérimentales. 2ème édition française. Paris : Centre de Psychologie Appliquée.

Dartigues, J.F., Commenges, D., Letenneur, L., Barberger-Gateau, P., Gilleron, V., Fabrigoule, C., Mazaux, J.M., Orgogozo, J.M. and Salamon, R. (1997). Cognitive predictors of dementia in elderly community residents. *Neuroepidemiology* **16**, 29-39.

De Gruttola, V. and Tu, X.M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003-1014.

Faucett, C.L., Schenker, N. and Taylor, J.M. (2002). Survival analysis using auxiliary variables via multiple imputation with application to AIDS clinical trial data. *Biometrics* **58**, 37-47.

Hall, C.B., Lipton, R.B., Sliwinski, M. and Stewart, W.F. (2000). A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Statistics in Medicine* **19**, 1555-1566.

Hall, C.B., Ying, J., Kuo, L., Lipton, R.B. (2003). Bayesian and profile likelihood change point methods for modeling cognitive function over time.

Computational Statistics & Data Analysis **42**, 91-109.

Hashemi, R., Jacqmin-Gadda, H., Commenges, D. (2003) A Latent process model for joint modeling of events and marker. *Lifetime Data Analysis* **9**, 331-343.

Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465-480.

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. and 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., 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.

Law, N.J., Taylor, J.M.G, and Sandler, H. (2002). The Joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics* **3**, 547-563.

Letenneur, L., Commenges, D., Dartigues, J. F., and Barberger-Gateau, P. (1994). Incidence of dementia and Alzheimer's disease in elderly community residents of southwestern France. *International Journal of Epidemiology* **23**, 1256-1261.

Marquardt, D.W. (1963). An algorithm for least squares estimation of nonlinear parameters. *SIAM Journal* **11**, 431-441.

19

Masur, D.M., Sliwinski, M., Lipton, R.B., Blau, A.D. and Crystal H.A. (1994). Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons. *Neurology* **44**, 1427-1432.

Pauler, D.K. and Finkelstein, D.M. (2002). Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in Mededicine* **21**, 3897-3911.

Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear regression.* John Wiley and Sons.

Stern, Y., Gurland, B., Tatemichi, T.K., Tang, M.X., Wilder, D., Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *Journal of the American Medical Association* **271**, 1004-1010.

Wulfsohn, M.S. and Tsiatis, A.A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330-339.

APPENDIX

Random effects estimators :

To estimate the random effects, we must compute their expectations conditionally on the data :

$$E(\log(\tau_i)|Y_i, \tilde{X}_i, \delta_i) = \frac{1}{L_i} \int \log(\tau) f_{Y_i|\tau_i}(y_i|\tau) f_{X_i|\tau_i}(\tilde{x}_i|\tau)^{\delta_i} S_{X_i|\tau_i}(\tilde{x}_i|\tau)^{1-\delta_i} f_{\log(\tau_i)}(\log(\tau)) d\log(\tau)$$

where $L_i = L(Y_i, \tilde{X}_i, \delta_i; \theta)$ is the contribution to the likelihood of subject i.

$$
\begin{aligned}
& E(U_i|Y_i, \tilde{X}_i, \delta_i) \\
= & \int E(U_i|Y_i, \log(\tau)) f_{\log(\tau_i)|Y_i, \tilde{X}_i, \delta_i}(\log(\tau)) d\log(\tau) \\
= & \frac{1}{L_i} \int G A_i' V_i^{-1} (Y_i - E(Y_i|\tau)) f_{Y_i|\tau_i}(y_i|\tau) f_{X_i|\tau_i}(\tilde{x}_i|\tau)^{\delta_i} S_{X_i|\tau_i}(\tilde{x}_i|\tau)^{1-\delta_i} f_{\log(\tau_i)}(\log(\tau)) d\log(\tau)
\end{aligned}
$$

In both cases, integrals are estimated by Gaussian quadrature and parameters θ are replaced by their MLE. An estimate of τ_i is obtained by taking the exponential of the estimate of $\log(\tau_i)$.

Rates of change estimators :

The difference between the individual rates of change during the first phase and l years after the changepoint is given by $-\Sigma_{k=2}^{K} k(\mu_k + u_{ki}) \times l^{k-1}$ and is estimated by replacing parameters by their MLE and random effects by their empirical Bayes estimates.

21

Table 1 : Estimates of the joint model for dementia and evolution of the

BVRT score for the Paquid cohort.

Parameter	estimates	SE	95% CI
Mixed model for the BVRT			
β_0	12.5	0.08	12.3 ; 12.6
β_1	-0.80	0.06	-0.91 ; -0.68
β_2	-0.19	0.03	-0.24 ; -0.14
β_3	0.012	0.002	0.008 ; 0.017
α_0	-1.85	0.19	-2.21 ; -1.48
α_1	0.51	0.17	0.18 ; 0.84
α_2	0.18	0.03	0.12 ; 0.23
α_3	-0.012	0.002	-0.016 ; -0.0075
Model for the changepoint			
$\alpha_{\tau 0}$	4.50	0.006	4.49 ; 4.51
$\alpha_{\tau 1}$	-0.26	0.036	-0.33 ; -0.19
Survival model for dementia			
γ_0	1.09	0.17	0.76 ; 1.43
γ_1	0.13	0.029	0.075 ; 0.187
η	0.77	0.039	0.69 ; 0.84

Table 2 : Estimated median age at dementia given the age at the acceleration of cognitive decline (and standard error computed by the delta method).

	HEL		LEL	
τ	$\hat{\text{med}}(T\|\tau)$	(SE)	$\hat{\text{med}}(T\|\tau)$	(SE)
65	73.1	(0.78)	83.4	(2.6)
75	81.6	(0.55)	93.0	(2.8)
85	89.9	(0.49)	102.4	(3.1)

Figure 1 : Expected BVRT score given the age at acceleration of cognitive decline (and 95% confidence interval).

plain line : high educational level and changepoint at 85 years

short dashed line : 95% confidence interval

long dashed line : low educational level and changepoint at 70 years

dotted line : 95% confidence interval

Figure 2 : Expected BVRT score in the years before the diagnosis given age at dementia.

plain line : high educational level, dementia at 75 years

long dashed line : high educational level, dementia at 90 years

short dashed line : low educational level, dementia at 75 years

dotted line : low educational level, dementia at 90 years

Figure 3 : Observed mean of the BVRT score at each visit (and 95% confidence interval) for incident cases of dementia and average of the expected BVRT score given the age at dementia computed for cases with available measure at the visit.

Horizontal bar with plain line : observed means for highly educated cases

Square with dashed line : observed means for less educated cases

Cross : predicted means

Figure 4 : Risk function for the marginal distribution of the age at dementia estimated by the joint model and non-parametrically.

plain line : high educational level, joint model

short dashed line : high educational level, non parametric estimation (with confidence bands)

long dashed line : low educational level, joint model

dotted line : low educational level, non parametric estimation (with confidence bands)
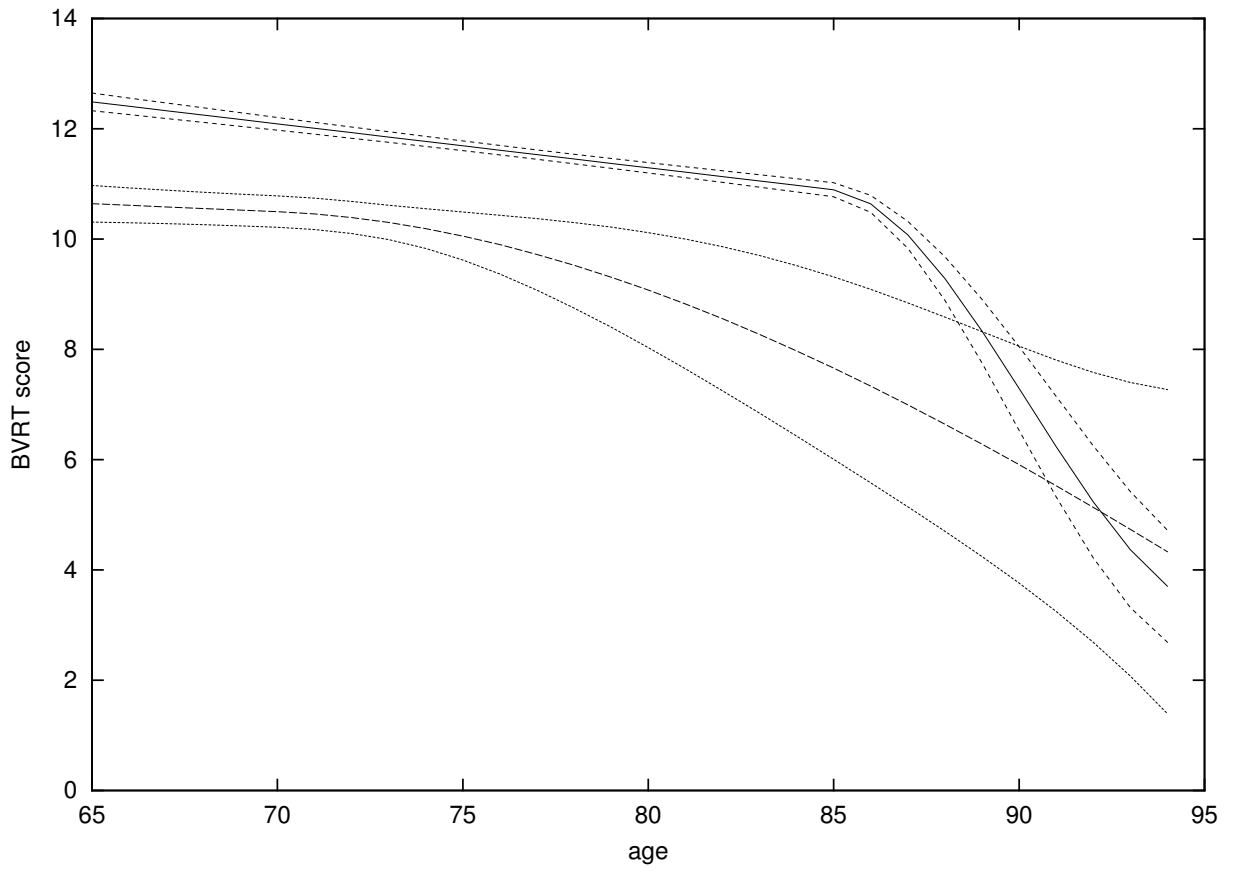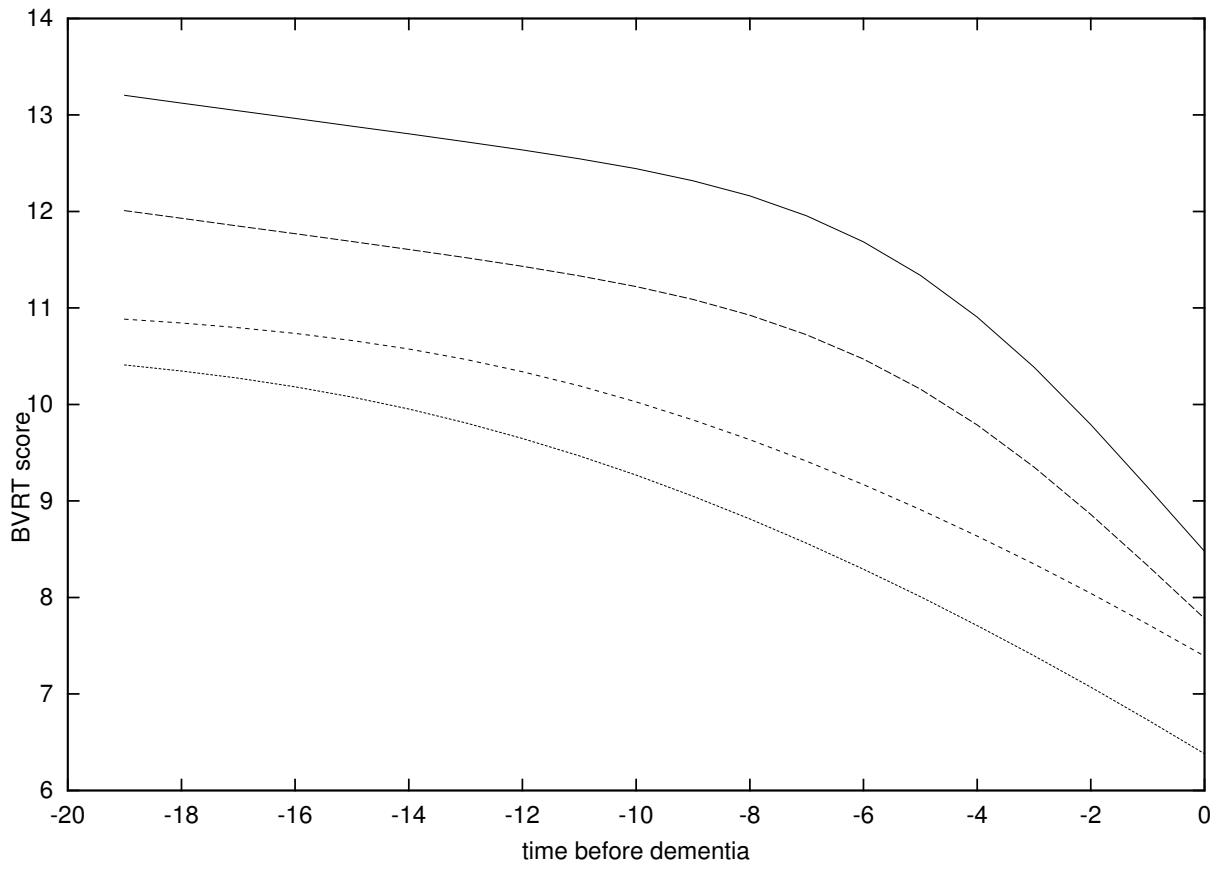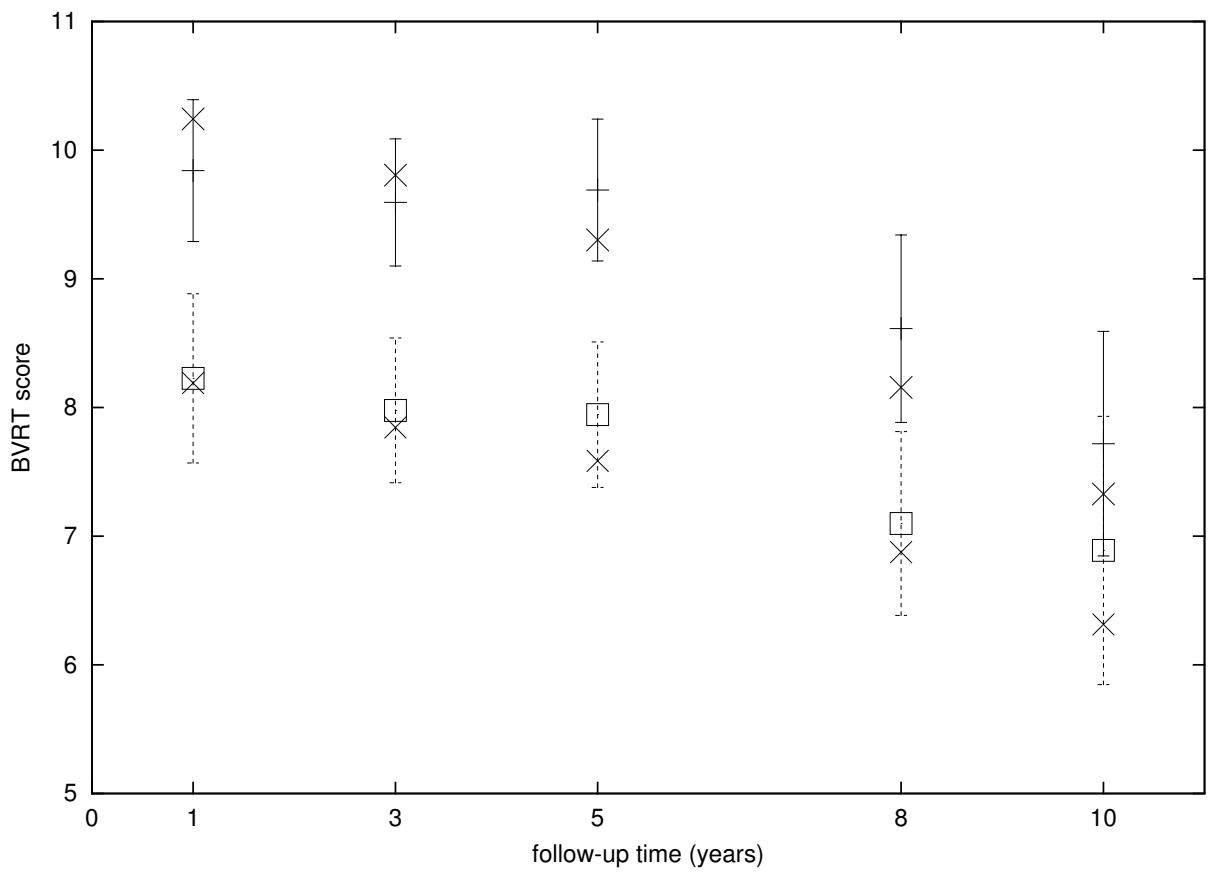
Figure 1

Figure 2

27

Figure 3

Figure 4