

**Estimating life expectancy of demented and institutionalized subjects from interval-censored observations of a Multi-state model**

Pierre Joly, Cécile Durand, Catherine Helmer, Daniel Commenges

► **To cite this version:**

Pierre Joly, Cécile Durand, Catherine Helmer, Daniel Commenges. Estimating life expectancy of demented and institutionalized subjects from interval-censored observations of a Multi-state model. *Statistical Modelling*, SAGE Publications, 2009, 9 (4), pp.345-360. <10.1177/1471082X0900900405>. <inserm-00677164>

**HAL Id: inserm-00677164**

**<http://www.hal.inserm.fr/inserm-00677164>**

Submitted on 7 Mar 2012

**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.

# Estimating life expectancy of demented and institutionalized subjects from interval-censored observations of a Multi-state model

P. Joly<sup>1,2\*</sup>, C. Durand<sup>3</sup>, C. Helmer<sup>1,2</sup> & D. Commenges<sup>1,2</sup>

February 20, 2009

<sup>1</sup> INSERM, U897, Epidémiologie et Biostatistique, Bordeaux, 33076, France.

<sup>2</sup> Université Victor Ségalen Bordeaux 2, Isped, Bordeaux, 33076, France.

<sup>3</sup> Institut de veille sanitaire, Saint-Maurice, 94415, France.

## ABSTRACT

We consider the problem of estimating life expectancy of demented and institutionalized subjects from interval-censored observations. A mixed discrete-continuous scheme of observation is a classical pattern in epidemiology because very often clinical status is assessed at discrete visit times while times of death or other events can be exactly observed. In this work we jointly modelled dementia, institutionalization and death from data of a cohort study. Due to discrete time observations, it may happen that a subject developed dementia or was institutionalized between the last visit and death. Consequently, there is an uncertainty about the precise number of diseased or institutionalized subjects. Moreover the time of onset of dementia is interval-censored. We use a penalized likelihood approach for estimating the transition intensities of the multi-state model. With these estimators, incidence and life expectancy can be computed easily. This approach deals with incomplete data due to the presence of left-truncation and interval-censoring. It can be generalized to take explanatory variables into account. We illustrate this approach by applying this model to the analysis of a large cohort study on cerebral aging.

Key Words: Dementia, Institutionalization, Interval-censoring, Life expectancy, Multi-state model, Penalized likelihood.

**\*Corresponding author:** Pierre Joly, U897, ISPED, Université de Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux cedex, France. Tel: (33) 5 57 57 45 16; Fax: (33) 5 56 24 00 81; e-mail: pierre.joly@isped.u-bordeaux2.fr

# 1 Introduction

In longitudinal follow-up of cohort studies, subjects are only intermittently observed. Such discrete observations times produce interval-censored observations. Thus, we lack information about the accurate time of transition between states and we may miss some transitions. Transition intensities may be interpreted, in some cases, as incidence or mortality rates (Keiding, 1991). In our application, we were primarily interested in estimating age-specific incidence of dementia and life expectancy for demented and institutionalized subjects. Knowledge of life expectancy for specific population such as demented or institutionalized subjects is useful to improve management of elderly subjects. Our application is based on the Paquid study (Letenneur et al., 1999), a prospective population-based cohort focused on the epidemiology of dementia in the elderly population. The study initially included a community-based sample of 3,777 elderly subjects, aged 65 and older. The subjects were selected conditionally on being non-demented and non-institutionalized at baseline. These subjects were followed-up over a 13-year period with 7 intermittent times of follow-up. Over the follow-up, information on onset of dementia and entry in institution was assessed only at planned visits. Thus, we have to deal with several problems. Firstly, due to intermittently observed data, we must jointly model dementia and death, in order not to under-estimate the incidence of dementia (Joly et al., 2002). Secondly, due to baseline selection conditions of the cohort, we have to take into account dementia and institution in the truncation condition. Thus we have to jointly model entry in institution and onset of dementia. Finally, we have to take into account that data were not observed in continuous time. We propose a non-parametric method for estimating the transition intensities in an irreversible multi-state model for interval-censored data. This work is the continuation of Commenges and Joly (2004); however, in this preliminary work, not applied to real data, the institutionalization status was assumed to be observed in continuous time. We treat the problem on institutionalization assessed only at planned visits in the present work. In addition, we estimate life expectancy for demented and institutionalized subjects. This approach also allows analysing risk factors acting on the transition between the different states.

The model is presented in the next section. Estimation of age-specific incidence of dementia and life expectancy is described in section 3. The application and the results are provided in section 4. Section 5 discusses the proposed approach.

## 2 The Model

### 2.1 The Multi-state Model

We consider a five-state model with irreversible transitions depicted in Figure 1, where the  $\alpha_{hj}$  are the transition intensities of a Markov process  $X$  (Commenges and Joly, 2004). Subjects are in state 0 if they are “healthy” (that is non demented and living at home), in state 1 if they are demented and living at home, in state 3 if they are demented and living in institution and in state 2 if they are non demented and living in institution. State 4, death, is an absorbing state and it is possible to go to this state from all the four other states.

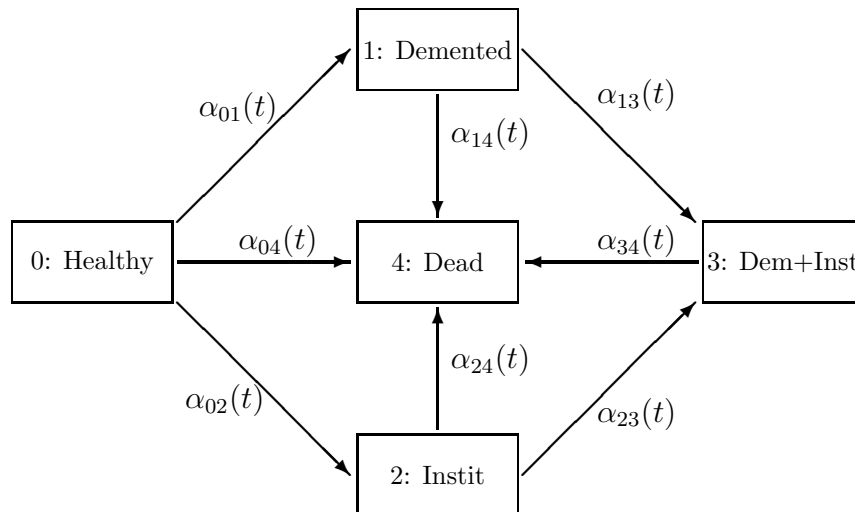


Figure 1: The five-state model for dementia, institutionalization and death.

Age is the major risk factor of dementia (see Launer et al., 1999), institutionalization and death and we are mostly interested in age-specific incidence and age-specific mortality. We thus consider a non-homogeneous Markov model where the intensities only depend on age and not on the time since the onset of the disease or the institutionalization.

### 2.2 Patterns of observation and interval censoring

In longitudinal follow-up studies, subjects are only intermittently observed. Thus, dementia and institutionalization statuses are assessed only at discrete time points, corresponding to the planned visits. If a transition toward dementia is observed, the age at the time of transition is interval-censored. On the contrary, once we observed a subject in an institution we can collect the date of institutionalization. Thus, as we observe institutionalization status in

continuous time, the age at the time of transition is not interval-censored for these subjects. If a subject who dies was non demented at the last visit, it is not known whether the subject has made the transition towards dementia or not. If a subject who dies was seen at home at the last visit, we try to collect the information about institutionalization; but for some subjects it is not possible to know whether the subject has been institutionalized before death or not. Thus, discrete observations times not only produce interval-censored observations but also an uncertainty about the followed path.

In Figure 2 we propose three examples of follow-up to illustrate the pattern of observation and its relation with interval censoring.

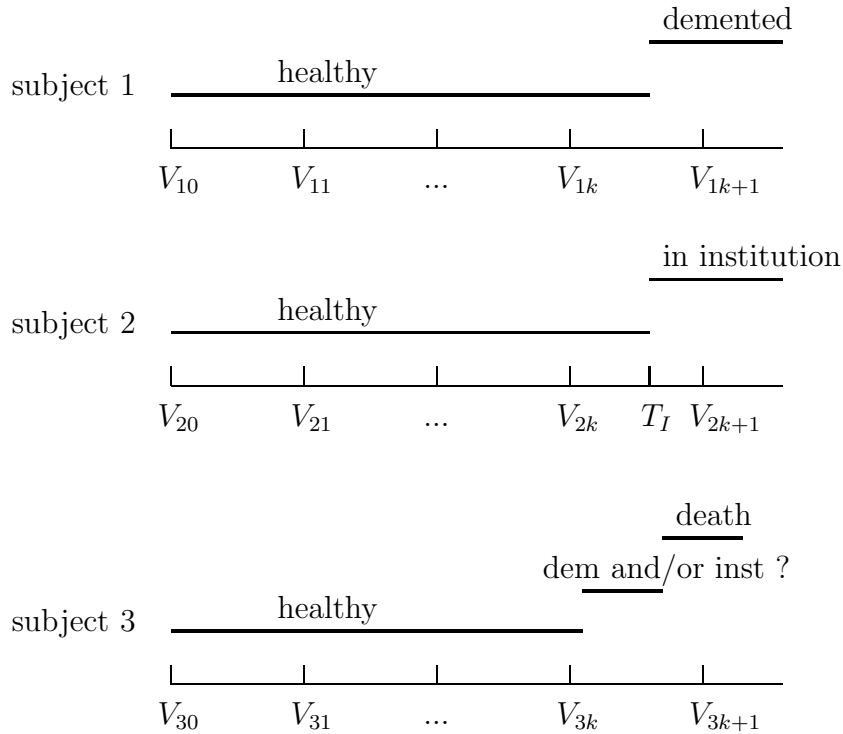


Figure 2: Examples of follow-up of three subjects 1, 2 and 3.

Subject 1 was seen healthy at the planned visit  $V_{1k}$  and was diagnosed demented at the next planned visit  $V_{1k+1}$ . The date of transition toward dementia is not known exactly; it is only known to lie between two visits (here between  $V_{1k}$  and  $V_{1k+1}$ ): there is interval-censoring.

Subject 2 was seen healthy at the planned visit  $V_{2k}$  and was seen in institution at the next planned visit  $V_{2k+1}$ . Once we observe a subject in institution we can collect the exact date of institutionalization (here  $T_I$  is the age at the time of institutionalization). There is no interval-censoring for entry in institution.

Subject 3 was seen healthy at the planned visit  $V_{3k}$  and he died before the next planned visit  $V_{3k+1}$ . It is not known whether he has made the transition

towards dementia. Similarly it is difficult, and some time impossible to know if he was institutionalized before death. This is the most difficult case.

There are other types of possible observation. It is possible to find examples in the sub-sample of the data set used for the application provided in Appendix B. All the possible trajectories are given in the likelihood part.

## 2.3 The Likelihood

The likelihood used in this article is slightly different than the likelihood detailed in Commenges and Joly (2004) because the assumptions on the observation of the institutionalization status differ. Subjects are not followed-up from the origin (their birth) but from their age at entry in the cohort. Demented and institutionalized subjects at the initial visit are excluded from the sample. This produces a left-truncation that has to be taken into account in the likelihood. We assume that the truncating and censoring mechanisms are independent from the multi-state process (for a discussion on independent censoring see Grüger et al., 1991). This happens for instance if the visits are fixed or are random variables independent of the illness-death process (visits times have been prespecified in advance in the Paquid study), and if loss of follow-up obeys an independent censoring mechanism. The last assumption is difficult to verify in our study and may not hold.

The model is completely specified by the eight transition intensities (see figure 1):  $\alpha_{01}(t)$ ,  $\alpha_{02}(t)$ ,  $\alpha_{04}(t)$ ,  $\alpha_{13}(t)$ ,  $\alpha_{14}(t)$ ,  $\alpha_{23}(t)$ ,  $\alpha_{24}(t)$ ,  $\alpha_{34}(t)$ . It is natural to write the likelihood in terms of both transition intensities and transition probabilities  $p_{hj}(s, t)$ , where  $p_{hj}(s, t) = P(X_t = j | X_s = h)$ . Kolmogorov equations allow expressing the latter in terms of the former:

$$\begin{aligned}
 p_{00}(s, t) &= e^{-A_{01}(s,t) - A_{02}(s,t) - A_{04}(s,t)} \\
 p_{11}(s, t) &= e^{-A_{13}(s,t) - A_{14}(s,t)} \\
 p_{22}(s, t) &= e^{-A_{23}(s,t) - A_{24}(s,t)} \\
 p_{33}(s, t) &= e^{-A_{34}(s,t)} \\
 p_{01}(s, t) &= \int_s^t p_{00}(s, u) \alpha_{01}(u) p_{11}(u, t) du \\
 p_{02}(s, t) &= \int_s^t p_{00}(s, u) \alpha_{02}(u) p_{22}(u, t) du \\
 p_{03}(s, t) &= \int_s^t \int_u^t p_{00}(s, u) \left( \alpha_{01}(u) p_{11}(u, v) \alpha_{13}(v) + \alpha_{02}(u) p_{22}(u, v) \alpha_{23}(v) \right) p_{33}(v, t) dv du \\
 p_{13}(s, t) &= \int_s^t p_{11}(s, u) \alpha_{13}(u) p_{33}(u, t) du \\
 p_{23}(s, t) &= \int_s^t p_{22}(s, u) \alpha_{23}(u) p_{33}(u, t) du
 \end{aligned}$$

Here,  $A_{hj}(s, t) = \int_s^t \alpha_{hj}(u) du$ .

Example: likelihood contribution for a subject  $i$  not demented and not seen in institution at the time of the last visit  $V_{Li}$ .  $\tilde{T}_{Vi}$  is the time of death or the time of right-censoring ( $\delta_{Vi} = 1$  if the subject has died). The subject is healthy at  $V_{0i}$ , the beginning of the follow-up; left-truncation is taken into account by using the probability  $p_{00}(V_{0i}, V_{Li})$  rather than  $p_{00}(0, V_{Li})$  (this amounts to divide the unconditional likelihood by  $p_{00}(0, V_{0i})$ , the likelihood of the condition).

$$\begin{aligned} \mathcal{L}_i = p_{00}(V_{0i}, V_{Li}) & \left[ \begin{aligned} & p_{00}(V_{Li}, \tilde{T}_{Vi}) \alpha_{04}(\tilde{T}_{Vi})^{\delta_{Vi}} \\ & + p_{01}(V_{Li}, \tilde{T}_{Vi}) \alpha_{14}(\tilde{T}_{Vi})^{\delta_{Vi}} \\ & + p_{02}(V_{Li}, \tilde{T}_{Vi}) \alpha_{24}(\tilde{T}_{Vi})^{\delta_{Vi}} \\ & + p_{03}(V_{Li}, \tilde{T}_{Vi}) \alpha_{34}(\tilde{T}_{Vi})^{\delta_{Vi}} \end{aligned} \right] \end{aligned}$$

The likelihood contribution for all the different trajectories of the subjects are detailed in Appendix A. This heuristic likelihood can be rigorously justified using the results of Commenges and Gégout-Petit (2007).

Inference is based on maximizing the likelihood. We propose to use a penalized likelihood approach detailed in the next paragraph. Since there are eight transition intensities to estimate in this model we propose a semi-parametric approach in order to reduce the number of functions to estimate non-parametrically. We assume proportionality for all transitions intensities towards death:

$$\begin{aligned} \alpha_{14}(t) &= \alpha_{04}(t) e^{\theta_{14}} \\ \alpha_{24}(t) &= \alpha_{04}(t) e^{\theta_{24}} \\ \alpha_{34}(t) &= \alpha_{04}(t) e^{\theta_{34}} \end{aligned}$$

We also assume proportionality for the transitions towards dementia:

$$\alpha_{23}(t) = \alpha_{01}(t) e^{\theta_{23}}$$

and proportionality for the transitions towards institution:

$$\alpha_{13}(t) = \alpha_{02}(t) e^{\theta_{13}}$$

Keiding *et al.* (2001) and Klein *et al.* (1993), for example, used the same kind of assumption in a multi-state model of bone marrow transplantation. Verifying these assumptions is theoretically possible by considering a more general model, with one more estimation of a transition intensity for example, and using an approximate cross-validation criterion (see Commenges *et al.*, 2007) to compare the two models. We did not do it because of the numerical complexity of this more general model.

## 2.4 The penalized likelihood approach

Usually, intensity functions are expected to be smooth. A possible means to introduce such *a priori* knowledge, is to penalize the log-likelihood by a term which takes large values for rough functions. The roughness penalty function chosen in our five-state model is the sum of the  $L_2$  norms of the second derivatives of the three intensities  $\alpha_{01}, \alpha_{02}$  and  $\alpha_{04}$ . In functional analysis, the integrated squared second derivative is a widespread choice for the roughness of a function (see Ramsay and Silverman, 1997). The penalized log-likelihood is thus defined as:

$$pl(\alpha_{01}, \alpha_{02}, \alpha_{04}, \boldsymbol{\theta}) = l(\alpha_{01}, \alpha_{02}, \alpha_{04}, \boldsymbol{\theta}) - \sum_{k=1,2,4} \kappa_{0k} \int_0^{+\infty} \alpha_{0k}''^2(u) du \quad (2.1)$$

where  $l$  is the full log-likelihood,  $\boldsymbol{\theta} = (\theta_{14}, \theta_{24}, \theta_{34}, \theta_{13}, \theta_{23})$  and  $\kappa_{01}, \kappa_{02}$  and  $\kappa_{04}$  are the three positive smoothing parameters which control the trade-off between the data fit and the smoothness of the intensity functions. Maximization of (2.1), in the desired class of function, defines the maximum penalized likelihood estimators (MPLE)  $\hat{\alpha}_{01}(\cdot), \hat{\alpha}_{02}(\cdot), \hat{\alpha}_{04}(\cdot)$  and  $\hat{\boldsymbol{\theta}}$ .

## 2.5 Estimation procedure

The MPLE cannot be calculated explicitly. However, for each transition intensity, it can be approximated using splines. Splines are piecewise polynomial functions which are combined linearly to approximate a function on an interval. We use cubic  $M$ -splines, which are normalized  $B$ -splines. For more details, see Ramsay (1988) and Joly *et al.* (1998). There is one base of splines  $\mathbf{M}_{0k}(\cdot) = (M_{10k}(\cdot), \dots, M_{m_{0k}0k}(\cdot))^T$  for each transition. Each transition intensity estimator  $\hat{\alpha}_{0k}(\cdot)$  is approximated by a linear combinations of  $m_{0k}$  cubic  $M$ -splines  $\tilde{\alpha}_{0k}(\cdot) = \tilde{\boldsymbol{\gamma}}_{0k} \mathbf{M}_{0k}(\cdot)$ . The spline function  $\tilde{\alpha}_0$  is defined by a sequence of increasing knots and a vector of coefficients  $\tilde{\boldsymbol{\gamma}}_0$ . The first knot is set at 0 or just before the first follow-up time point (for delayed entry) and the last knot is set just after the last follow-up time point. Other knots are put equidistantly between the first and the last one. Theoretically, the more knots, the better the approximation. Increasing the number of knots does not deteriorate the MPLE: this is because the degree of smoothing in the penalized likelihood approach is tuned by the smoothing parameter  $\boldsymbol{\kappa}_0$  and not by the number of knots. Once a sufficient number of knots is established, there is no advantage in adding more because the more knots, the longer the running time. Usually, we found that a good choice for the number of knots is between 7 and 15. The three vectors of spline's coefficients  $\tilde{\boldsymbol{\gamma}}_{01}, \tilde{\boldsymbol{\gamma}}_{02}$  and  $\tilde{\boldsymbol{\gamma}}_{04}$  are obtained simultaneously by maximizing  $pl$ . When the four vectors  $\tilde{\boldsymbol{\gamma}}_{01}, \tilde{\boldsymbol{\gamma}}_{02}, \tilde{\boldsymbol{\gamma}}_{04}$  and  $\hat{\boldsymbol{\theta}}$  are obtained, all the functions of interest can be computed, as in a parametric model. To choose the three smoothing parameters, we



used an approximate cross-validation criterion (see Commenges *et al.*, 2007). The criterion was maximized using a pseudo-grid method. The model can be generalized to regression models to study explanatory variables using a proportional intensity function model. The regression parameters and the baseline functions are estimated simultaneously. The variance of parameters are computed using the inverse of the matrix of the second derivatives of the penalized likelihood.

### 3 Estimation of age-specific incidence of dementia and life expectancy

Unlike with an “illness-death” model (see Joly *et al.*, 2002) the transition intensity  $\alpha_{01}$  is not enough to estimate the age-specific incidence of dementia. Using the previously defined transition probabilities, the age-specific incidence of dementia for subjects living at home at 65 years old is:

$$Ic(t) = \frac{p_{00}(65, t)\alpha_{01}(t) + p_{02}(65, t)\alpha_{23}(t)}{p_{00}(65, t) + p_{02}(65, t)} \quad (3.1)$$

This is a weighted mean of the incidences of healthy and institutionalized subjects, the weights being the probabilities of these states at time  $t$ . For a similar expression for obtaining death intensity in an illness-death model see, for example, Andersen (1988).

Likewise, obtaining all the transition intensities allows to compute easily life expectancy in institution and life expectancy for demented people. The life expectancy for a subject demented and living at home at age  $t_d$  is:

$$\int_{t_d}^{+\infty} [p_{11}(t_d, s) + p_{13}(t_d, s)]ds \quad (3.2)$$

For a subject demented and living in institution at age  $t_d$  the life expectancy is:

$$\int_{t_d}^{+\infty} p_{33}(t_d, s)ds \quad (3.3)$$

In the same way, the life expectancy for a subject non demented living in institution at age  $t_I$  is:

$$\int_{t_I}^{+\infty} [p_{22}(t_I, s) + p_{23}(t_I, s)]ds \quad (3.4)$$

Finally, the life expectancy for a subject healthy (non demented and living at home) at age  $t$  is:

$$\int_t^{+\infty} [p_{00}(t, s) + p_{01}(t, s) + p_{02}(t, s) + p_{03}(t, s)]ds \quad (3.5)$$

When the estimators  $\tilde{\gamma}_{01}$ ,  $\tilde{\gamma}_{02}$ ,  $\tilde{\gamma}_{04}$  and  $\hat{\theta}$  are obtained, we have the estimators of all the transition intensities and then de probability functions  $p_{11}$ ,  $p_{13}$ ,  $p_{33}$ ,  $p_{22}$  and  $p_{23}$  and then we plug them in 3.1, 3.2, 3.3, 3.4 and 3.5 to obtain estimators of age-specific incidence of dementia and life expectancy. A Bayesian technique for generating confidence bands for each transition intensity was used in Joly et al. (2002). But, even with the delta-method, we cannot use this technique to compute confidence bands for age-specific incidence and for life expectancy because the relation between transition intensities and these functions is too complex. To obtain confidence bands for these functions we propose to use a parametric bootstrap (Efron and Tibshirani, 1993). We generated 2000 realizations from the asymptotic distribution of the parameters. For each replication we computed age-specific incidence and life expectancy and for each point considered, we order the 2000 values obtained; the lower bound and the upper bound of our confidence band are respectively given by the 2.5th and the 97.5th empirical percentiles. However, this estimator does not take into account the variability due to the choice of smoothing parameters.

## 4 Application: estimation of age-specific incidence of dementia and survival time in institution

The application is based on the Paquid research programme, a prospective cohort study of mental and physical aging that evaluates social environment and health status. The target population consists of subjects aged 65 years and older living at home at the baseline visit in southwestern France. The baseline variables registered included socio-demographic factors, medical history and psychometric tests. Subjects were re-evaluated 1, 3, 5, 8, 10 and 13 years after the initial visit. Demented subjects at the baseline visit were removed from the sample because the initial sample may not be representative. Therefore, this produced a left-truncation problem, subjects living at home and not demented at baseline visit. As mentionned before, the inspection processes for dementia, institutionalization and death are all different. Dates of death are always reported unlike dates of onset of dementia and entry in institution which were assessed only at planned visits. But for entry in institution, unlike for onset of dementia, we can most often collect the exact date of institutionalization when the subject is seen in institution.

Numbers of observed women and men diagnosed as demented and in institution over the 13 years of follow-up are detailed in Table 1.

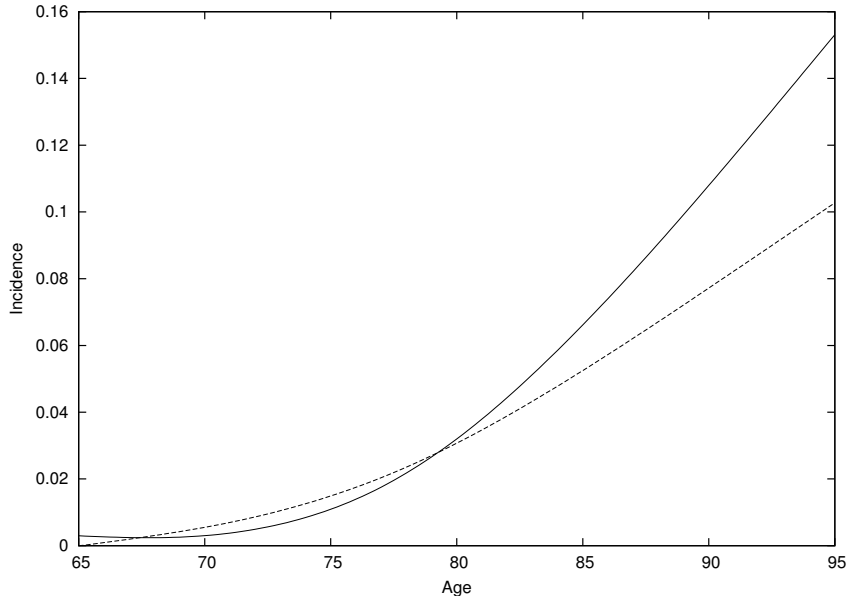


Figure 3: Age-specific incidence of dementia for women (—) and men (- - -).

Table 1: Observed number of women and men demented and in institution over the 13 years of follow-up.

	Women	Men	Total
	2133	1540	3673
diagnosed as demented	396	189	585
in institution	342	113	455
dead*	1038	939	1977

\* dead over the 13 years of follow-up include subjects previously diagnosed as demented or institutionalized.

In a previous work (Joly *et al.*, 2002) we found that the age-specific incidence of dementia was different between women and men. It is also well known that the age-specific risk of death is higher in men than in women, especially in France (Robine *et al.*, 2002). Thus we separately estimated the transition intensities for women and men. We used 7 knots for each intensity function. Figure 3 displays age-specific incidence of dementia for women and men. We observe, as in Joly *et al.* (2002), that the estimate of age-specific incidence of dementia is slightly higher for men than for women under 79 years old and slightly higher for women than for men after this age. We give in Table 2 the estimates of the parameters  $\theta$  and the corresponding standard errors.

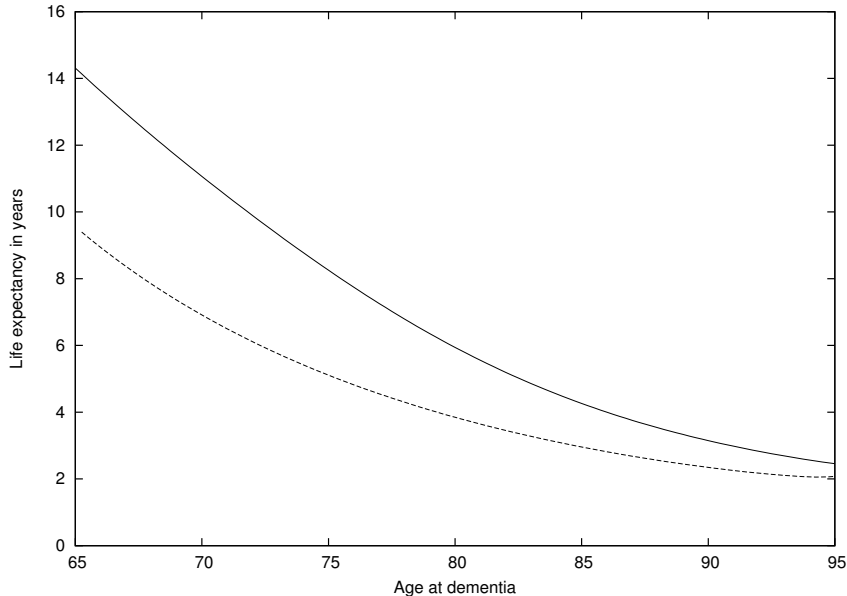


Figure 4: Life expectancy for demented subjects living at home as a function of age: women (—) and men (- - -).

Table 2: Estimates of  $\theta$  (with standard errors) for women and men

	Women		Men	
$\theta_{14}$ (Dm $\rightarrow$ D)	1.13	(0.18)	2.25	(0.17)
$\theta_{24}$ (I $\rightarrow$ D)	1.29	(0.17)	1.79	(0.18)
$\theta_{34}$ (Dm+I $\rightarrow$ D)	1.83	(0.11)	1.57	(0.18)
$\theta_{13}$ (Dm $\rightarrow$ Dm+I)	1.83	(0.15)	2.39	(0.24)
$\theta_{23}$ (I $\rightarrow$ Dm+I)	0.92	(0.19)	-0.84	(1.04)

Figures 4 and 5 display the life expectancy for women and men, respectively for subjects developing dementia and entered in institution. In Figure 4, we plot the life expectancy for women and men demented at home as a function of age. In Figure 5 we plot the life expectancy for women and men in institution and not demented as a function of age. Whatever the age, the life expectancy is higher for women than for men but the difference is decreasing with age. We give some selected values in Table 3. There seems to be a contrast between institutionalized women and men. While women in institution have a higher estimated life expectancy when they are non demented versus when they are demented, the reverse is true for men. Subjects in institution may have some other diseases, but this does not explain the contrast between institutionalized women and men. Note however that the difference for men is small with regard to the confidence interval.

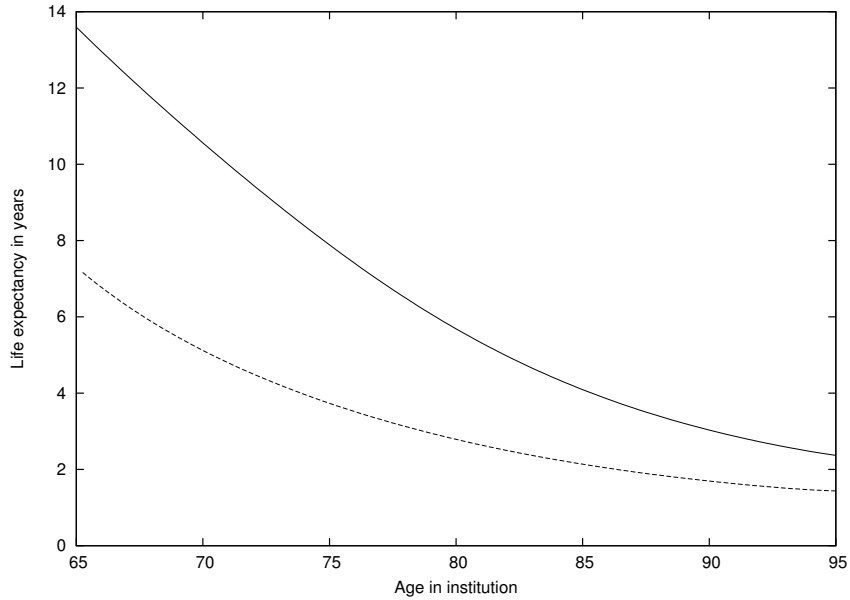


Figure 5: Life expectancy for institutionalized non demented subjects as a function of age: women (—) and men (- -).

Table 3: Estimated life expectancy for subjects, healthy, in institution and/or with dementia, for women and men for different ages with 95% confidence intervals (parametric bootstrap with 2000 replications)

Women				
Age	Healthy	Demented	In institution	Dem+Inst
70	15.92 [15.07;16.20]	10.99 [9.58;12.14]	10.54 [9.07;11.65]	8.19 [7.03;9.38]
80	9.02 [8.70;9.33]	5.91 [5.31;6.50]	5.68 [5.01;6.31]	4.40 [3.83;5.01]
90	5.19 [4.82;5.60]	3.12 [2.79;3.47]	3.02 [2.65;3.40]	2.29 [1.99;2.64]
Men				
Age	Healthy	Demented	In institution	Dem+Inst
70	12.95 [12.43;13.37]	6.94 [5.76;8.09]	4.90 [3.72;6.23]	5.76 [4.39;7.31]
80	7.80 [7.41;8.22]	3.87 [3.30;4.50]	2.66 [2.03;3.42]	3.19 [2.38;4.16]
90	5.33 [4.79;5.94]	2.35 [1.98;2.80]	1.60 [1.22;2.10]	1.95 [1.44;2.64]

## 5 Discussion

In this paper we propose an estimation method applied to multi-state model with interval-censoring. Our work was mainly motivated by two major issues: estimation of age-specific incidence of dementia and life expectancy for demented or institutionalized subjects. With interval-censoring and a high risk of death for demented and institutionalized subjects, there is an underestimation of the incidence of dementia using a (two-state) survival analysis approach. It is then important to take interval-censoring into account. The proposed penalized likelihood approach in a multi-state model yields a method for analyzing such data and for providing estimators of the intensity functions, which cannot easily be estimated using conventional non-parametric methods. Once the intensity functions are obtained, age-specific incidence and life expectancy can easily be computed. Moreover, the penalized likelihood approach yields smooth estimates of the transition intensities without making the strong assumption that it belongs to a given parametric model. The proposed approach can be applied to semi-Markov models as well as to non-homogeneous Markov models and can be generalized to regression models to deal with explanatory variables. The approach can easily be extended to models with reversible transitions if there is no interval-censoring in the reversible transitions. If there is interval-censoring, there may be an infinite number of transitions between two observations, which is hard to treat in non homogeneous Markov models. In the same way right-truncation can be accommodated. The program was written in FORTRAN 77 and is available on request.

In France, the higher mortality rate in men compared to women is well known. This work shows that there are also large differences in life expectancy according to sex, even after institutionalization or dementia, although the difference is decreasing with age. These data have public health implications, in the case of demented people and of institutionalized residents.

## References

1. Andersen, P.K. (1988) Multi-state models in survival analysis: a study of nephropathy and mortality in diabetes. *Statistics in Medicine*, 7, 661-670.
2. Commenges, D. et Joly, P. (2004) Multi-state model for dementia, institutionalization and death. *Communications in statistics A*, 33, 1315-1326.
3. Commenges, D., Joly, P., Gégout-Petit, A. and Liqueur, B. (2007) Choice between semi-parametric estimators of Markov and non-Markov multi-state models from generally coarsened observations. *Scandinavian Journal of Statistics*, 34, 33-52.
4. Commenges, D. and Gégout-Petit, A. (2007) Likelihood for generally coarsened observations from multi-state or counting process models. *Scandinavian Journal of Statistics*, 34, 432-450.
5. Efron, B. and Tibshirani, R. J. (1993) An introduction to the Bootstrap. *New York: Chapman & Hall*.
6. Gröger, J., Kay, R. and Schumacher, M. (1991) The validity of inferences based on incomplete observations in disease state models. *Biometrics*, 47, 595-605.
7. 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.
8. 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.
9. Keiding, N. (1991) Age-specific Incidence and Prevalence: a Statistical Perspective. *Journal of the Royal Statistical Society, Series A*, 154, 371-412.
10. Keiding, N., Klein, J.P. and Horowitz, M.M. (2001) Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine*, 20, 1871-1885.
11. Klein, J.P., Keiding, N. and Copelan, E.A. (1993) Plotting summary predictions in multi-state survival models: probabilities of relapse and death in remission for bone marrow transplantation patients. *Statistics in Medicine*, 12, 2315-2332.

12. Launer, L.J., Andersen, K., Dewey, M.E., Letenneur, L., et al. (1999) Rates and risk factors for dementia and Alzheimer's disease: results from EURODEM pooled analyses. EURODEM Incidence Research Group and Work Groups. European Studies of Dementia. *Neurology*, 52(1):78-84.
13. Letenneur, L., Gilleron, V., Commenges, D., Helmer, C., et al. (1999) Are sex and educational level independent predictors of dementia and alzheimer's disease ? Incidence data from the PAQUID project. *Journal of Neurology Neurosurgery and Psychiatry*, 66, 177-183.
14. O'Sullivan, F. (1988) Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and Statistical Computing*, vol. 9, 2, 363-379.
15. Marquardt, D. (1963) An algorithm for least-squares estimation of non-linear parameters. *SIAM Journal of Applied Mathematics*, 11, 431-441.
16. Ramsay, J.O. (1988) Monotone regression splines in action. *Statistical Science*, 3, 425-461.
17. Ramsay, J.O., Silverman, B. W. (1997) Functional Data Analysis. *Springer*.
18. Robine, J.M., Jagger, C., Cambois, E. (2002) European perspectives on healthy aging in women. *J Women Aging*, 14:119-133.



# Appendix A

## Likelihood

In principle all the indicators and times of observations should be indexed by  $i$  (example  $\mathcal{L}_i, \delta_{Vi}, \dots$ ); we will omit this for the sake of simplicity.

We have three indicators:  $\delta_V = 1$  if the subject is deceased,  $\delta_I = 1$  if the subject is known institutionalised and  $\delta_D = 1$  if the subject is diagnosed demented.  $V_0$  is the age at the entrance in the cohort.  $\tilde{T}_V$  is the age at death or at right-censoring for death,  $T_I$  is the age at the time of institutionalization or at right-censoring for the institutionalization.  $V_L$  is the age at the last visit without dementia and  $V_R$  is the age at the first visit with dementia (for subject diagnosed demented).

### Subject not demented at the time of the last visit $V_L$ ( $\delta_D = 0$ )

- If  $\delta_I = 0$  (not institutionalised  $T_I = \tilde{T}_V$ )

$$\mathcal{L} = p_{00}(V_0, V_L) \left[ \begin{aligned} & p_{00}(V_L, \tilde{T}_V)\alpha_{04}(\tilde{T}_V)^{\delta_V} \\ & + p_{01}(V_L, \tilde{T}_V)\alpha_{14}(\tilde{T}_V)^{\delta_V} \end{aligned} \right]$$

- If  $\delta_I = 0$  (not seen institutionalised but  $T_I < \tilde{T}_V$ )

$$\mathcal{L} = p_{00}(V_0, V_L) \left[ \begin{aligned} & p_{00}(V_L, \tilde{T}_V)\alpha_{04}(\tilde{T}_V)^{\delta_V} \\ & + p_{01}(V_L, \tilde{T}_V)\alpha_{14}(\tilde{T}_V)^{\delta_V} \\ & + p_{01}(V_L, T_I)p_{13}(T_I, \tilde{T}_V)\alpha_{34}(\tilde{T}_V)^{\delta_V} \\ & + p_{00}(V_L, T_I)p_{02}(T_I, \tilde{T}_V)\alpha_{24}(\tilde{T}_V)^{\delta_V} \\ & + p_{00}(V_L, T_I)p_{03}(T_I, \tilde{T}_V)\alpha_{34}(\tilde{T}_V)^{\delta_V} \end{aligned} \right]$$

- If  $\delta_I = 1$  (institutionalised)

- If  $T_I < V_L$  (institutionalised before  $V_L$ )

$$\mathcal{L} = p_{00}(V_0, T_I)\alpha_{02}(T_I)p_{22}(T_I, V_L) \left[ \begin{aligned} & p_{22}(V_L, \tilde{T}_V)\alpha_{24}(\tilde{T}_V)^{\delta_V} \\ & + p_{23}(V_L, \tilde{T}_V)\alpha_{34}(\tilde{T}_V)^{\delta_V} \end{aligned} \right]$$

- If  $T_I > V_L$  (institutionalised after  $V_L$ )

$$\mathcal{L} = p_{00}(V_0, V_L) \left[ \begin{aligned} & p_{00}(V_L, T_I)\alpha_{02}(T_I)p_{22}(T_I, \tilde{T}_V)\alpha_{24}(\tilde{T}_V)^{\delta_V} \\ & + p_{00}(V_L, T_I)\alpha_{02}(T_I)p_{23}(T_I, \tilde{T}_V)\alpha_{34}(\tilde{T}_V)^{\delta_V} \\ & + p_{01}(V_L, T_I)\alpha_{13}(T_I)p_{33}(T_I, \tilde{T}_V)\alpha_{34}(\tilde{T}_V)^{\delta_V} \end{aligned} \right]$$

Subject diagnosed demented at  $V_R$  ( $\delta_D = 1$ )

- If  $\delta_I = 0$  (not institutionalised  $T_I = \tilde{T}_V$ )

$$\mathcal{L} = p_{00}(V_0, V_L)p_{01}(V_L, V_R)p_{11}(V_R, \tilde{T}_V)\alpha_{14}(\tilde{T}_V)^{\delta_V}$$

- If  $\delta_I = 0$  (not known institutionalised but  $T_I < \tilde{T}_V$ )

$$\mathcal{L} = p_{00}(V_0, V_L)p_{01}(V_L, V_R)p_{11}(V_R, T_I) \left[ \begin{array}{l} p_{11}(T_I, \tilde{T}_V)\alpha_{14}(\tilde{T}_V)^{\delta_V} \\ + p_{13}(T_I, \tilde{T}_V)\alpha_{34}(\tilde{T}_V)^{\delta_V} \end{array} \right]$$

- If  $\delta_I = 1$  (institutionalised)

- If  $T_I < V_L$

$$\mathcal{L} = p_{00}(V_0, T_I)\alpha_{02}(T_I)p_{22}(T_I, V_L)p_{23}(V_L, V_R)p_{33}(V_R, \tilde{T}_V)\alpha_{34}(\tilde{T}_V)^{\delta_V}$$

- If  $V_L < T_I < V_R$

$$\mathcal{L} = p_{00}(V_0, V_L) \left[ p_{01}(V_L, T_I)\alpha_{13}(T_I)p_{33}(T_I, V_R) \right. \\ \left. + p_{00}(V_L, T_I)\alpha_{02}(T_I)p_{23}(T_I, V_R) \right] p_{33}(V_R, \tilde{T}_V)\alpha_{34}(\tilde{T}_V)^{\delta_V}$$

- If  $T_I > V_R$

$$\mathcal{L} = p_{00}(V_0, V_L)p_{01}(V_L, V_R)p_{11}(V_R, T_I)\alpha_{13}(T_I)p_{33}(T_I, \tilde{T}_V)\alpha_{34}(\tilde{T}_V)^{\delta_V}$$

## Appendix B

Sub-sample of 40 men followed during 13 years: Paquid research programme.

Subject	$V_{0i}$	$\delta_{Di}$	$\delta_{Ii}$	$\delta_{Vi}$	$V_{Li}$	$V_{Ri}$	$T_{Ii}$	$\tilde{T}_{Vi}$
01	85.25	1	0	0	90.61	95.50	98.27	98.27
02	74.17	0	1	0	86.86	86.86	85.76	86.86
03	71.50	0	0	1	74.73	74.73	74.73	78.67
04	66.00	0	0	0	79.07	79.07	79.07	79.07
05	73.75	0	0	1	75.18	75.18	76.10	82.33
06	66.58	0	0	0	79.64	79.64	79.64	79.64
07	70.33	0	0	1	75.56	75.56	77.56	77.56
08	79.67	0	0	1	84.93	84.93	84.93	86.23
09	66.92	0	0	0	77.11	77.11	77.11	79.79
10	78.17	0	0	1	78.17	78.17	82.54	82.54
11	71.92	0	0	1	82.06	82.06	82.06	84.73
12	87.00	0	0	1	87.97	87.97	87.97	90.30
13	66.08	0	0	1	73.92	73.92	73.92	75.73
14	82.00	0	1	1	87.49	87.49	87.40	87.95
15	73.92	0	0	0	86.81	86.81	86.81	86.81
16	87.58	0	0	1	88.58	88.58	88.58	89.94
17	65.58	1	0	1	73.43	75.77	75.77	77.83
18	73.25	0	0	1	76.64	76.64	77.48	77.48
19	83.17	0	0	1	83.17	83.17	90.07	90.07
20	68.67	0	0	1	70.23	70.23	70.23	78.98
21	86.00	0	1	1	89.47	89.47	88.85	93.28
22	65.92	0	0	0	66.96	66.96	68.34	79.04
23	81.67	1	0	1	81.67	82.65	82.65	83.97
24	88.42	0	0	1	93.84	93.84	93.99	93.99
25	66.42	0	0	0	79.55	79.55	79.55	79.55
26	67.17	0	0	0	75.05	75.05	75.05	80.02
27	74.83	0	0	1	74.83	74.83	77.21	88.09
28	65.58	0	0	0	78.46	78.46	78.46	78.46
29	73.92	1	1	1	74.82	76.72	77.92	78.69
30	73.83	0	0	0	86.44	86.44	86.44	86.44
31	76.50	0	0	1	81.69	81.69	81.69	84.38
32	74.92	1	1	1	80.12	83.01	81.66	86.25
33	76.42	0	0	1	77.51	77.51	78.61	79.55
34	73.25	0	0	0	86.11	86.11	86.11	86.12
35	80.75	1	0	1	83.87	85.94	85.94	89.16
36	69.58	0	0	1	70.53	70.53	72.28	79.35
37	79.25	1	0	1	80.22	84.38	86.95	87.97
38	75.83	0	0	1	75.83	75.83	75.83	77.32
39	65.67	0	0	0	78.48	78.48	78.48	78.48
40	77.25	1	1	0	84.93	87.42	87.37	90.02