Multi-state models: a flexible approach for modelling complex event histories in epidemiology

Ahmadou Alioum

Centre de Recherche INSERM U897 "Epidémiologie et Biostatistique
Institut de Santé Publique, d'Epidémiologie et de Développement (ISPED)
Université Bordeaux Segalen, France

March 7, 2013

Short-Course "Statistique Appliquée pour le Développement en Afrique 2013" (SADA'13), 4-8 mars 2013, Cotonou, Bénin

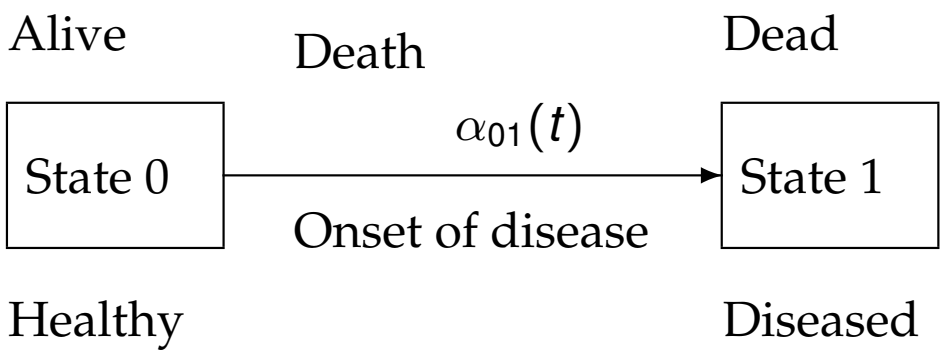
Organization of the course

1. Two-states (survival) models
2. Multistate models - Examples
3. Multistate processes - Markov processes
4. Transition probabilities and transition intensities
5. Patterns of observation
6. Likelihood
7. Likelihood inference
8. Non-parametric, semi-parametric estimation (Package mstate)
9. Problem with interval-censored data (Package SmoothHazard)

Two-state (survival) model

- Survival analysis in epidemiology = observing individuals over time, focusing on the occurence of a event of interest (death, occurrence of a disease or a complication,...) and time of occurrence

- The occurrence of the event of interest is a transition from one state to another



Two-state (survival) model

- Positive random variable T = time from a given origin (in state 0 at time 0) to the occurrence of the event "death"
- If $X = (X_t, t \geq 0)$ is a stochastic process with X_t the state occupied by an individual at time $t \geq 0$, $X_t \in \mathcal{S} = \{0, 1\}$, then T is the smallest time at which the process is not in the initial state 0 anymore

$$T := \inf\{t : X_t \neq 0\}$$

- Hazard rate function $\rightarrow$ transition intensity

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \alpha_{01}(t)$$

Two-state (survival) model

- Cumulative hazard function $\rightarrow$ cumulative transition intensity

$$\Lambda(t) = \int_0^t \lambda(u)du = A_{01}(0,t)$$

- Survival function or distribution function $\rightarrow$ transition probabilities

$$S(t) = P(T > t) = \exp[-\Lambda(t)] = P(X_t = 0|X_0 = 0) = P_{00}(0,t)$$

$$F(t) = P(T \leq t) = 1 - S(t) = P(X_t = 1|X_0 = 0) = P_{01}(0,t)$$

- Transition probability matrix

$$P(0,t) = \begin{pmatrix} P_{00}(0,t) & P_{01}(0,t) \\ 0 & 1 \end{pmatrix}$$

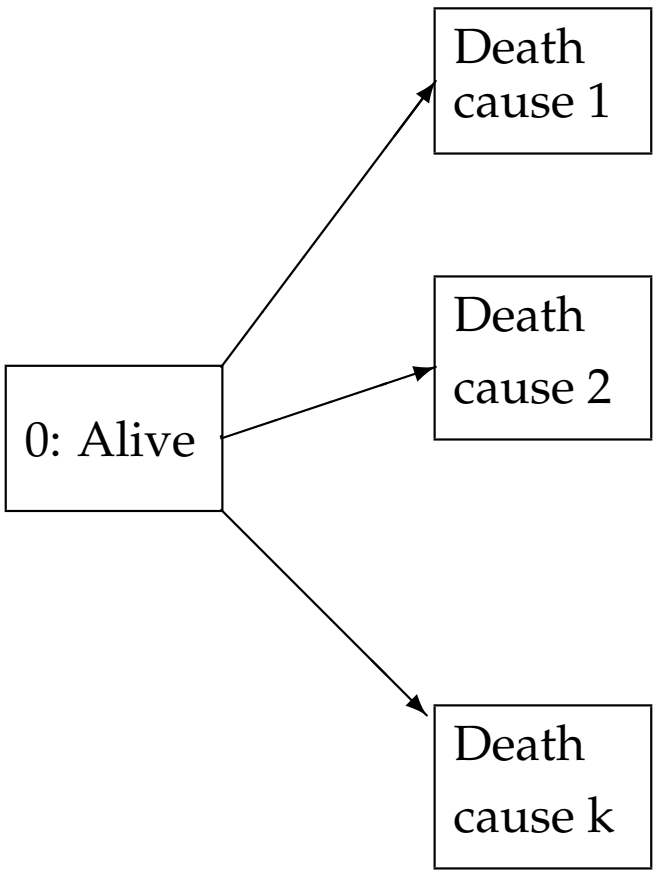
- Inference for hazard rate and survival functions taking into account incomplete observation of event histories (right-censoring, left-truncation, ...)

Multi-states model : examples

- Generalization of survival models (two-state models) : more than one transition between more than two states
- Multi-state model (MSM) = model of continuous-time stochastic process allowing individuals to move among a finite number of states
- Extensive litterature on MSMs
  - Books : ABGK (1993), Hougaard (2000), Aalen, Borgan and Gjessing (2008), Beyersmann, Schumacher and Allignol (2012)
  - Reviews : Commenges (1999), Hougaard (1999), Andersen and Keiding (2002)
  - Special Issues of : Statistical Methods in Medical Research on MSMs (2002) ; Journal of Statistical Software (2011)

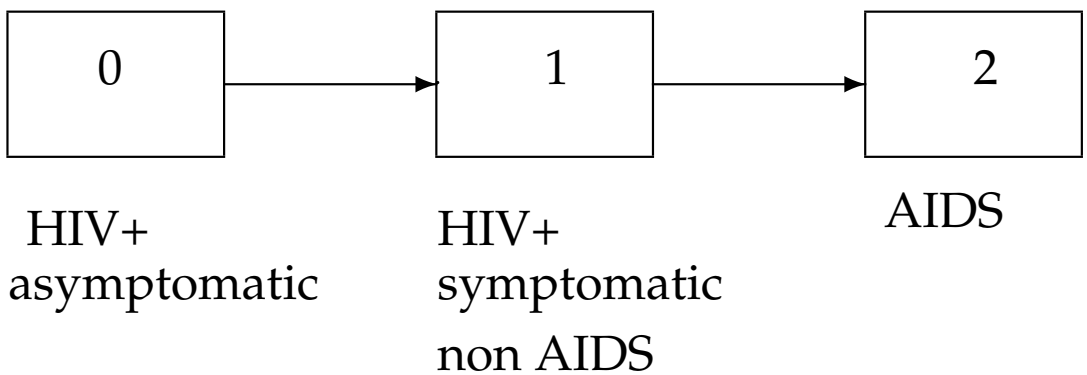
Multi-states model : examples

- Competing risks model : one transient state (0: alive) and several absorbing states representing different causes of "deaths"



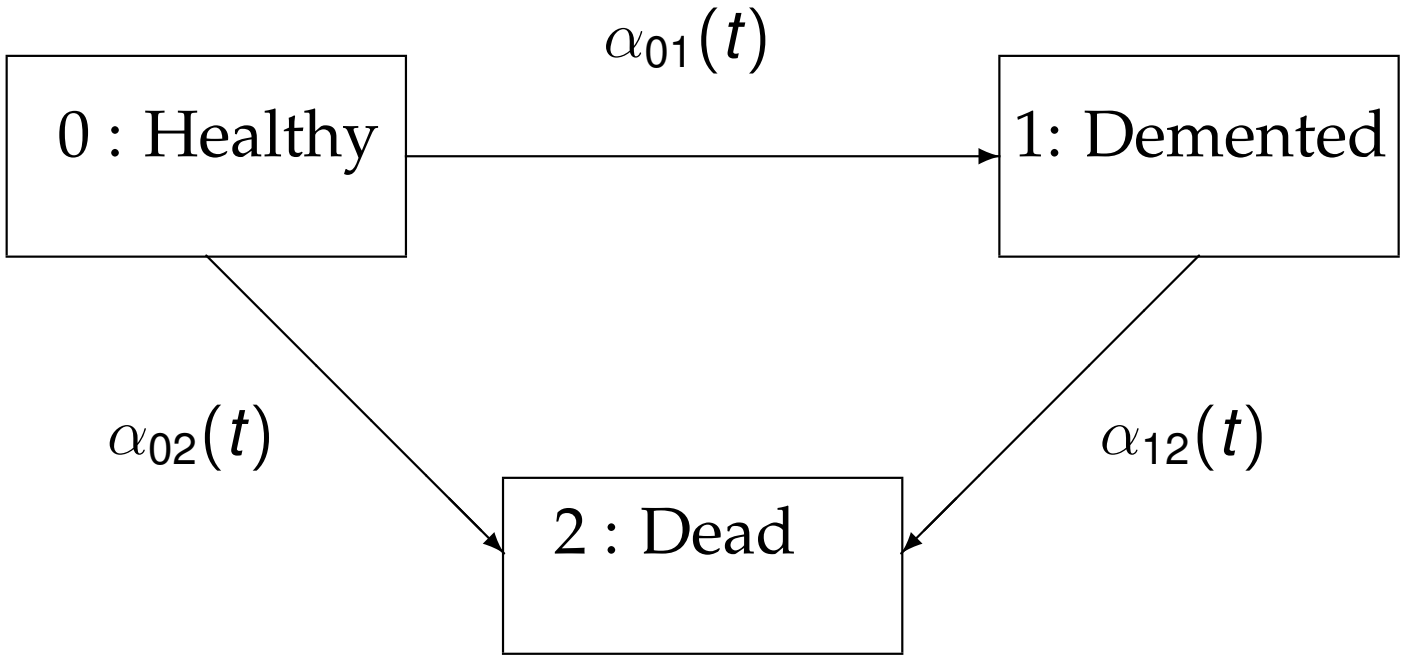
Multi-states model : examples

- Splitting the initial state into two or more states gives a progressive multi-state model



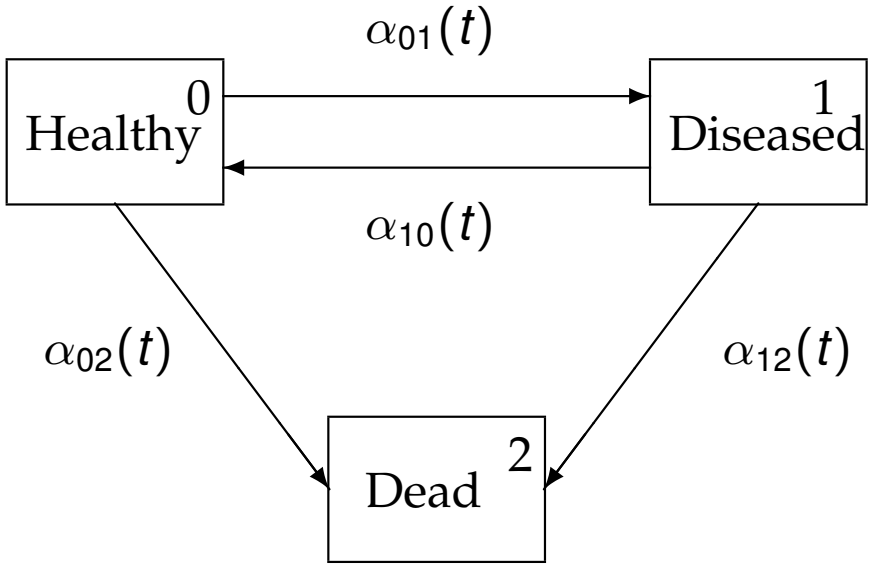
Multi-states model : examples

- Illness-death model : very useful in epidemiology to study both incidence of a disease and the mortality rate
  - $\alpha_{01}(t)$: incidence of dementia
  - $\alpha_{02}(t)$: mortality rate for healthy
  - $\alpha_{12}(t)$: mortality rate for demented

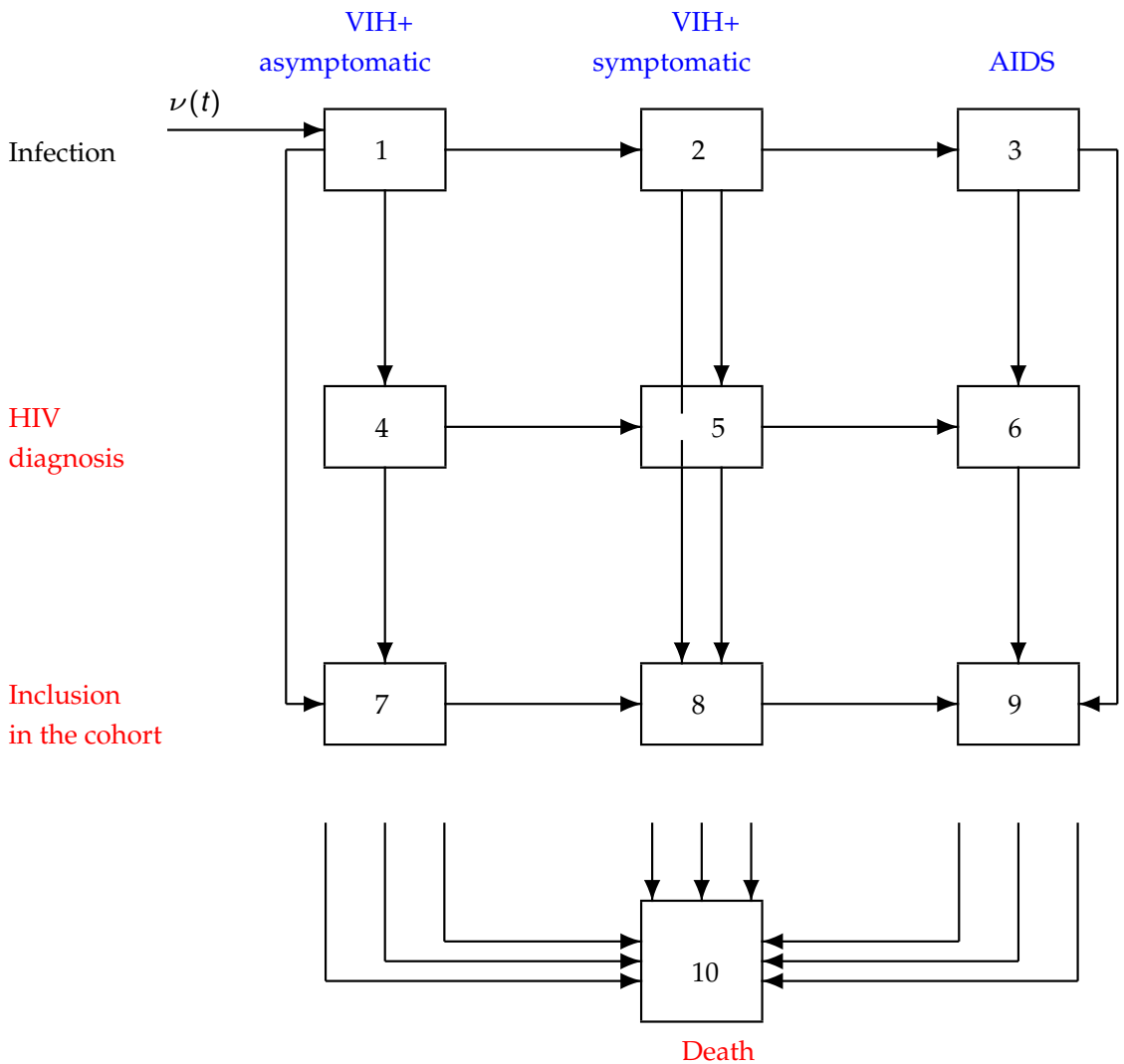


Multi-states model : examples

- Bi-directional illness-death model : possibility of recovery from diseased state



Multi-states model : progression of HIV infection, HIV diagnosis, inclusion in a cohort and death



Stochastic processes and filtration

- Continuous-time stochastic process $X = (X_t, t \geq 0)$
- We should think of a filtration as a flow of information. The σ-algebra $\mathcal{F}_t$ contains the events that can happen "up to time t". The filtration defined by

$$\mathcal{F}_t^X = \sigma(X_s : s \leq t)$$

  is called the filtration generated by X or the natural filtration of X (keep tracks of the "history" of the process)
- A multi-state process is a process which can take a finite number of states; that is, for any t, the variable X_t has values in $\mathcal{S} = \{0, 1, \ldots, K\}$.
- The law of a multi-state process is defined by its finite dimensional distribution:

$$P(X_{t_1} = j_1, X_{t_2} = j_2, \ldots, X_{t_n} = j_n), n \in \mathbb{N}$$

Specification of multi-state process

A multi-state model is fully characterized through

- Transition probabilities between state h and state j :

$$P_{hj}(s,t;\mathcal{F}_{s-}) = P(X_t = j | X_s = h, \mathcal{F}_{s-}), \text{ for } s < t$$

- Transition intensities

$$\alpha_{hj}(t;\mathcal{F}_{t-}) = \lim_{\Delta t \to 0} \frac{P_{hj}(t, t+\Delta t;\mathcal{F}_{t-})}{\Delta t}$$

- Cumulative (integrated) transition intensities

$$A_{hj}(t;\mathcal{F}_{t-}) = \int_0^t \alpha_{hj}(u;\mathcal{F}_{u-})du$$

Markov processes

- Markov assumption : the future and the past of the process are independent given its present state

  $P_{hj}(s,t;\mathcal{F}_{s-}) = P_{hj}(s,t) = P(X_t = j | X_s = h)\ s < t,\ h,j \in \mathcal{S},\ h \neq j$

  and $P_{hh}(s,t) = 1 - \sum_{j \neq h} P_{hj}(s,t)$

  $P(s,t) = \{P_{hj}(s,t)\}$ matrix of transition probabilities

- Transition intensities are then defined as:
  $\alpha_{hj}(t) = \lim_{\Delta t \to 0} P_{hj}(t, t+\Delta t)/\Delta t,\ h \neq j$, and
  $\alpha_{hh}(t) = -\sum_{j \neq h} \alpha_{hj}(t)$; $\sum_{j \neq h} \alpha_{hj}(t)$ is the hazard function associated with the distribution of the sojourn time h.

  $a(t) = \{\alpha_{hj}(t)\}$ matrix of transition intensities

- $A(t) = \{A_{hj}(t)\}$ matrix of cumulative transition intensities, where $A_{hj}(t) = \int_0^t \alpha_{hj}(u)du$

Relationship between transition probabilities and intensities

- Chapman-Kolmogorov equation:
  $P(s,t) = P(s,u)P(u,t),\ s < u < t$
- $P(s,t)$ is the unique solution of the Kolmogorov forward differential equation: $\frac{\partial}{\partial t}P(s,t) = P(s,t)a(t); P(s,s) = I$
- $P(s,t)$ can be recovered from the transition intensities through product integration

$$P(s,t) = \prod_{u\in(s,t]}(I + dA(u))$$

where $\prod$ is a product integral

Relationship between transition probabilities and intensities

- For simple models, such as the illness-death model, the transition probabilities has explicit expression

$$P_{00}(s,t) = e^{-[A_{01}(s,t)+A_{02}(s,t)]}$$
$$P_{11}(s,t) = e^{-A_{12}(s,t)}$$
$$P_{01}(s,t) = \int_s^t P_{00}(s,u)\alpha_{01}(u)P_{11}(u,t)du$$

- Particular case : homogeneous Markov chain
  $a(t)$ is constant; that is $\alpha_{hj}(t) = \alpha_{hj}$ does not depend on time t, then

$$P(s,t) = P(0,t-s) = P(t-s)$$

  The solution of Kolmogorov equation is : $P(t) = e^{ta}$
  If the transition intensities matrix a can be diagonalized, then $P(t) = V\ \mathrm{diag}(e^{\rho_0 t}, ..., e^{\rho_K t})\ V^{-1}$
  Explicit formulas for the illness-death model

Aalen-Johansen estimator

$N_{hj}(t)$: number of observed direct transitions from state h to state j up to time t; $Y_h(t)$: number of individuals under observation in state h just before time t

$dA(t)$ matrix of elements $d(A_{hj}(t))_{h,j} = (\alpha_{hj}(t))_{h,j}dt$

Nelson-Aalen estimator of the cumulative transition intensities:

$$d\hat{A}_{hj}(t) = \frac{dN_{hj}(t)}{Y_h(t)}$$

$$\hat{A}_{hj}(t) = \sum_{u \leq t} d\hat{A}_{hj}(t), \ h \neq j \text{ and } \hat{A}_{hh}(t) = -\sum_{h \neq j} \hat{A}_{hj}(t)$$

Aalen-Johansen estimator of the transition probabilities : $\hat{A}(t)$ is a matrix of step-functions with finite number of jumps in $(s, t]$

$$\hat{P}(s, t) = \prod_{s < u \leq t} (I + d\hat{A}(u))$$

Special case : Illness-death model

- Aalen-Johansen estimators of $P_{00}(s,t)$ and $P_{11}(s,t)$ are equal to the corresponding Kaplan-Meier estimators

  $\hat{P}_{00}(s,t) = \prod_{s<u\leq t}(1 - d\hat{A}_{01}(u) - d\hat{A}_{02}(u))$

  $\hat{P}_{11}(s,t) = \prod_{s<u\leq t}(1 - d\hat{A}_{12}(u))$

- $\hat{P}_{01}(s,t) = \sum_{s<u\leq t} \hat{P}_{00}(s,u-)d\hat{A}_{01}(u)\hat{P}_{11}(u+,t)$

- Then since $\hat{P}_{02}(s,t) = 1 - \hat{P}_{00}(s,t) - \hat{P}_{01}(s,t)$ and $\hat{P}_{12}(s,t) = 1 - \hat{P}_{11}(s,t)$

  $\hat{P}_{02}(s,t) = \sum_{s<u\leq t} \hat{P}_{00}(s,u-)d\hat{A}_{01}(u)\hat{P}_{11}(u+,t)$
  $+ \sum_{s<u\leq t} \hat{P}_{00}(s,u-)d\hat{A}_{02}(u)$

  $\hat{P}_{12}(s,t) = \sum_{s<u\leq t} \hat{P}_{11}(s,u-)d\hat{A}_{12}(u)$

Semi-Markov process

- The most used semi-Markov process can be specified by:

$$P_{hj}(s,t,T_h) = P(X_t = j | X_s = h, T_h), s < t$$

$$\alpha_{hj}(t,T_h) = \lim_{\Delta t \to 0} P_{hj}(t, t+\Delta t, T_h)/\Delta t$$

  These quantities are random and we do not have the Kolmogorov equations.
  In general we may still be able to derive transition probabilities from transition intensities but the theory is more complex.

- Clock reset models : the relevant time is the time spent in the current state (the clock is "reset" to 0 each time a patient enters a new state)

Semi-Markov model : an example

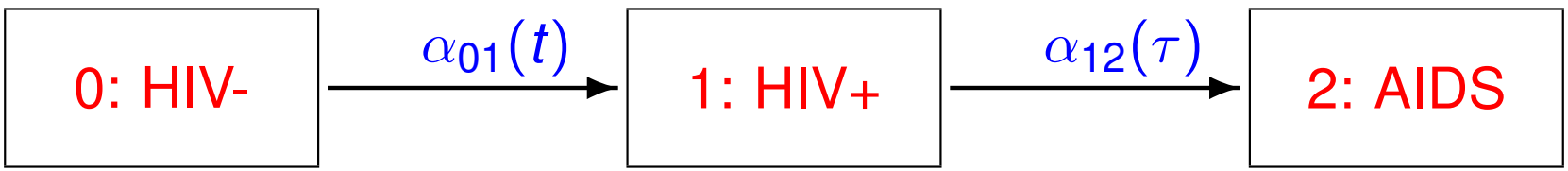


Figure: Progressive three-state Model: if $\tau = t$ this is a Markov model; if τ is the time since the $0 \to 1$ transition this is a semi-Markov model.

Multistate models for a population

- States and time
  Associate to each subject i of a population a multistate process $X^i = (X_t^i)_{t \geq 0}$ with law specified by $a^i(.)$.
  States of the X^is and time t have same meaning for all the subjects.
  However, several choices of t are possible: t may be the calendar time or time since a subject-specific event; if the event is birth, t is age.
- Heterogeneity of intensities
  The population may be homogeneous or heterogeneous.
  Heterogenity may be modelled as depending of observed covariates $Z_i(t)$: $\alpha_{hj}^i(t) = \phi_{hj}(\alpha_{hj0}(t), Z_i(t))$
  - Proportional intensity assumption:
    $\alpha_{hj}(t, Z_i(t)) = \alpha_{hj0}(t) e^{\beta_{hj} Z_i(t)}$
  - Additive intensity $\alpha_{hj}(t, Z_i(t)) = \alpha_{hj0}(t) + \beta_{hj}(t) Z_i(t)$

Patterns of observation

- Selection of the sample

  For estimation purpose we have to draw a sample of size n from a population.

  It often happens that subjects can enter the sample only if they are in a given state.

  For instance in an illness-death model the study sample may be a random sample of healthy subject. The condition is:

  $$X^i_{t_{i0}} = 0$$

  This is a kind of left-truncation which must be taken into account for inference.

Patterns of observation

- Left- and right-censoring
  Subject i may be first observed at time L_i until time C_i.

  We observe:

  $$(X_t^i, L_i \leq t \leq C_i)$$

  If $L_i = 0$ this the extension of right-censoring to multistate processes.

Patterns of observation

- Discrete-time observations

  We may consider that the state of the process i is observed at only a finite number of times $V_0^i, V_1^i, \ldots, V_m^i$.

  This typically happens in cohort studies where fixed visit times have been planned. In such cases the exact times of transitions are not known; it is only known that they occurred during a particular interval; these observations are said to be interval-censored.

  It is also possible that the state of the process is not exactly observed but it is known that it belongs to a subset of $\{0, 1, \ldots, K\}$

Patterns of observation

- Mixed continuous and discrete-time observations

  The most common pattern of observation is in fact a mixing of discrete and continuous time observations.

  This is because most multistate models include states which represent clinical status and one state which represents death: most often clinical status is observed at discrete times (visits) while the (nearly) exact time of death can be retrieved.

  It may also happen that some events other than death are observed exactly.

Likelihood

- Using the likelihood approach for inference taking into account selection and censoring.
- The likelihood can be written easily in terms of the $P_{hj}(.,.)$ and the $\alpha_{hj}(.)$ under the assumption of
  - ignorability
  - conditional on the state at the first observation time V_0
  - the processes X^is are independent
  - the processes are Markov

Likelihood

- Likelihood: continuous-time observations

  The process X is continuously observed from V_0 to C. We observe that transitions have occured at (exactly) times $T_1, T_2, \ldots, T_m$. With the convention $T_0 = V_0$, the likelihood (conditional on $X(V_0)$) is:

$$\mathcal{L} = [\prod_{r=0}^{m-1} P_{X(T_r),X(T_r)}(T_r, T_{r+1}-)\alpha_{X(T_r),X(T_{r+1})}(T_{r+1})]P_{X(T_m),X(T_m)}(T_m, C).$$

- Likelihood: discrete-time observations Observations of X

  are taken at $V_0, V_1, \ldots, V_m$. We write the likelihood as if the V_j were fixed (ignorability):

$$\mathcal{L} = \prod_{r=0}^{m-1} P_{X(V_r),X(V_{r+1})}(V_r, V_{r+1}).$$

Likelihood

- Likelihood: Mixed discrete-continuous time observations

  For simplicity we give the likelihood when the process is observed at discrete times but time of transition towards one absorbing state, representing generally death, is exactly observed or right-censored. Denote by K this absorbing state.
  Observations of X are taken at $V_0, V_1, \ldots, V_m$ and the vital status is observed until C ($C \geq V_m$). Let us call $\tilde{T}$ the follow-up time that is $\tilde{T} = \min(T, C)$, where T is the time of death; we observe $\tilde{T}$ and $\delta = I\{T \leq C\}$.

  The likelihood is:

$$\mathcal{L} = \prod_{r=0}^{m-1} P_{X(V_r),X(V_{r+1})}(V_r, V_{r+1}) \sum_{j \neq K} P_{X(V_m),j}(V_m, \tilde{T}-)\alpha_{j,K}(T)^{\delta}.$$

Example: likelihood in the illness-death model

- If the subject starts in state “health”, has never been observed in the “illness” state and was last seen at visit L (at time V_L) the likelihood is:

$$\mathcal{L} = P_{00}(V_0, V_L)[P_{00}(V_L, \tilde{T})\alpha_{02}(\tilde{T})^{\delta} + P_{01}(V_L, \tilde{T})\alpha_{12}(\tilde{T})^{\delta}]$$

- If the subject has been observed in the illness state for the first time at V_J then the likelihood is:

$$\mathcal{L} = P_{00}(V_0, V_{J-1})P_{01}(V_{J-1}, V_J)P_{11}(V_J, \tilde{T})\alpha_{12}(\tilde{T})^{\delta}.$$

Likelihood inference

- Once we have a likelihood $\mathcal{L}(a(.), \mathcal{X})$, several approaches for getting estimators can be used
  - Non-parametric: no restriction on $a(.)$
  - Parametric: $a(.) \in \mathcal{M}$
  - Smooth non-parametric by smoothing a non-parametric estimate
  - Smooth non-parametric by penalized likelihood
  - Bayesian approach
- The non-parametric methods may be combined with a parametric form for effects of covariates leading to semi-parametric models.

$$\alpha_{hj}(t, Z) = \alpha_{hj,0}(t) \exp(\beta_{hj} Z)$$

$$\hat{P}(s, t; Z) = \prod_{s<u\leq t} (I + d\hat{A}(u; Z))$$

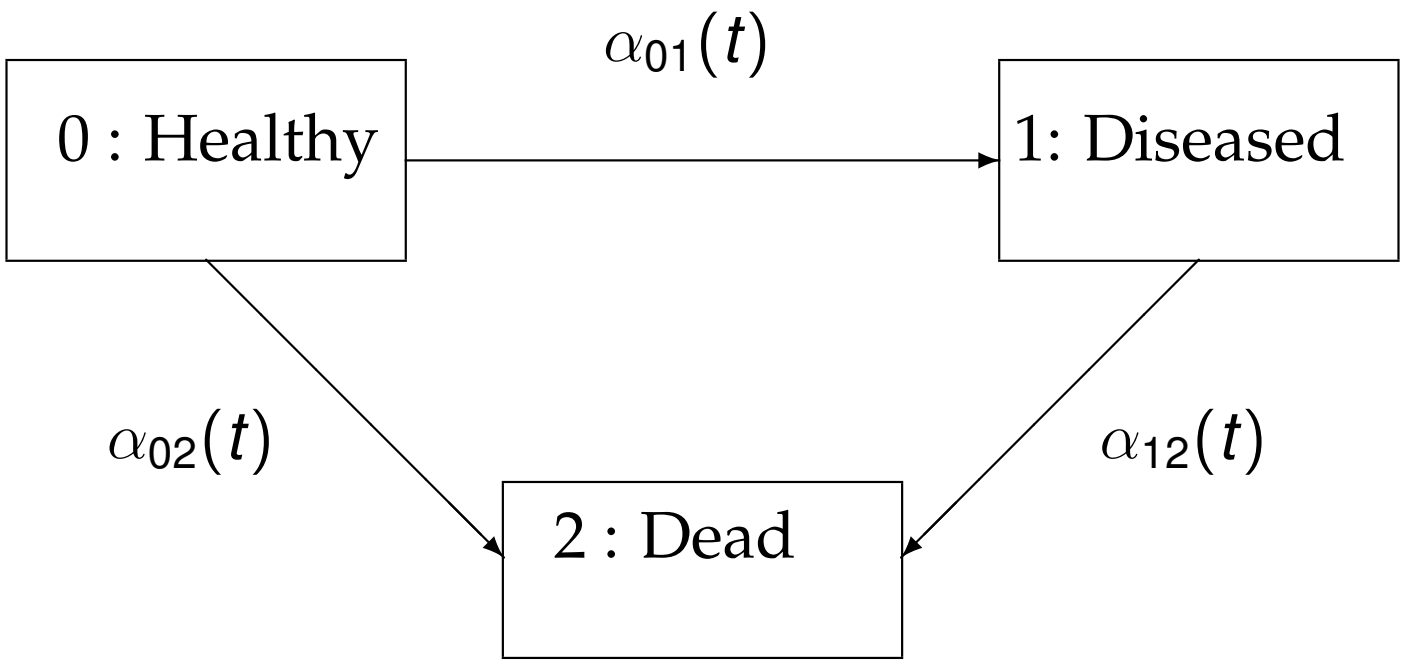
Case with continuous time observations

- If transition times are exactly observed or right-censored, the problem can be split in as many survival problems as there are transitions.
  - For instance if subject *i* enters state h at T_h^i and makes a transition from state *h* to state *j* at T_j^i his likelihood contribution to estimation of $\alpha_{hj}(.)$ is

$$e^{-\int_{T_h^i}^{T_j^i} \alpha_{hj}(u)du} \alpha_{hj}(T_j^i)$$

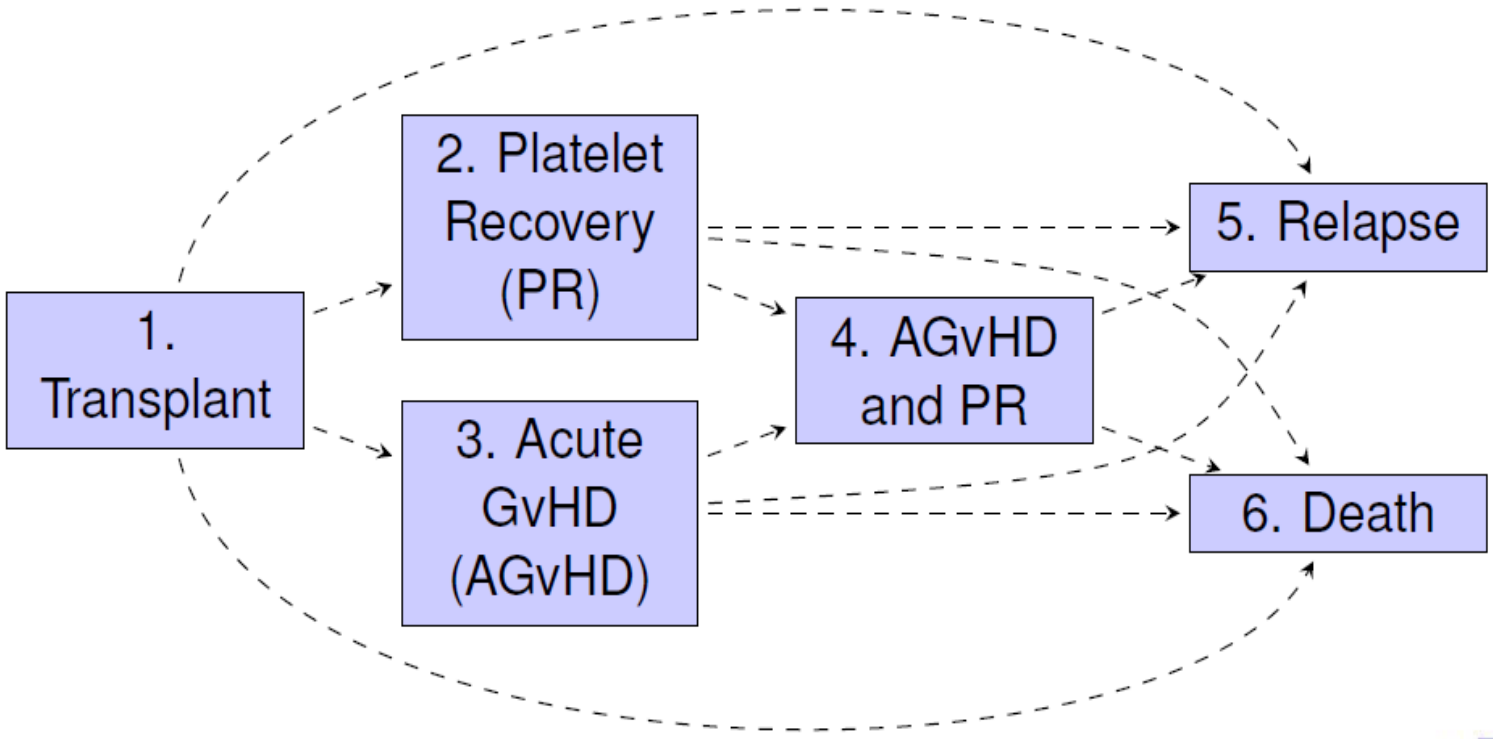
    - If a subject makes a transition to another state it is considered as censored.
- Estimation of $\alpha_{hj}(.)$ can be done by survival methods allowing left-truncation (delayed entry) and right-censoring.
- Estimation may be parametric or semi-parametric (Cox model).

Illness-death model



An example : Transplant treatment of patients with blood cancer

From de Wreede, Fiocco and Putter (Journal of Statistics Software, 2011, vol 38, Issue 7)



An example : Transplant treatment of patients with blood cancer

Pronostic factors for all patients

Pronostic factor (name in data)	Categories	*n*(%)
Donor recipient (match)	no gender mismatch	1734 (76)
	gender mismatch	545 (24)
Prophylaxis (proph)	no	1730 (76)
	yes	549 (24)
Year of transplant (year)	1985-1989	634 (28)
	1990-1994	896 (39)
	1995-1998	749 (33)
Age at transplant (years) (agec1)	≤ 20	551 (24)
	20-40	1213 (53)
	> 40	515 (23)

Package *mstate*

- Estimation and prediction (non-parametric and semi-parametric) : cumulative transition intensities, transition probabilities, prediction
- Need to make a dataset with a line for each transition (data preparation)

Journal of Statistical Software

January 2011, Volume 38, Issue 7. *http://www.jstatsoft.org/*

mstate: An R Package for the Analysis of Competing Risks and Multi-State Models

Liesbeth C. de Wreede
Leiden University
Medical Center

Marta Fiocco
Leiden University
Medical Center

Hein Putter
Leiden University
Medical Center

Package *mstate* : semi-parametric models

- Incorporate covariate information in Cox PH models
  - Separate Cox models for each of the transitions

$$\alpha_{hj}(t; Z) = \alpha_{hj,0}(t) \exp(\beta_{hj} Z)$$

  - Build appropriate subsets of data and estimate covariate vector and baseline intensity for each transition (standard software)
- Transition-specific covariates
  - Model all transitions in a single Cox regression model, using stratified regression and transition-specific covariates

$$\alpha_{hj}(t; Z_{hj}) = \alpha_{hj,0}(t) \exp(\beta Z_{hj})$$

  - Obtain different covariate effect estimates accross transition (transition-specific covariates)
  - Obtain a single effect estimate assumed to be common for all transitions (stratified Cox regression using the global covariate)

Package *mstate*

```
> library(mstate)
> data("ebmt4")
> ebmt <- ebmt4
> head(ebmt4)
  id  rec rec.s   ae ae.s recae recae.s  rel rel.s  srv srv.s      year agecl
1  1   22     1  995    0   995       0  995     0  995     0 1995-1998 20-40
2  2   29     1   12    1    29       1  422     1  579     1 1995-1998 20-40
3  3 1264     0   27    1  1264       0 1264     0 1264     0 1995-1998 20-40
4  4   50     1   42    1    50       1   84     1  117     1 1995-1998 20-40
5  5   22     1 1133    0  1133       0  114     1 1133     0 1995-1998   >40
6  6   33     1   27    1    33       1 1427     0 1427     0 1995-1998 20-40
  proph              match
1    no no gender mismatch
2    no no gender mismatch
3    no no gender mismatch
4    no    gender mismatch
5    no    gender mismatch
6    no no gender mismatch
```

```
> tmat <- transMat(x = list(c(2,3,5,6), c(4,5,6), c(4,5,6), c(5,6), c(), c()),
names=c("Tx", "Rec", "AE", "Rec+AE", "Rel", "Death"))
> tmat
        to
from     Tx Rec AE Rec+AE Rel Death
  Tx     NA   1  2     NA   3     4
  Rec    NA  NA NA      5   6     7
  AE     NA  NA NA      8   9    10
  Rec+AE NA  NA NA     NA  11    12
  Rel    NA  NA NA     NA  NA    NA
  Death  NA  NA NA     NA  NA    NA
```

Package *mstate*

```
> msebmt <- msprep(data=ebmt, trans=tmat, time=c(NA, "rec", "ae", "recae", "rel",
"srv"), status=c(NA, "rec.s", "ae.s", "recae.s",  "rel.s", "srv.s"),
keep=c("match", "proph", "year", "agecl"))
> msebmt[msebmt$id == 1, c(1:8, 10:12)]

An object of class 'msdata'

Data:
  id from to trans Tstart Tstop time status proph      year agecl
1  1    1  2     1      0    22   22      1    no 1995-1998 20-40
2  1    1  3     2      0    22   22      0    no 1995-1998 20-40
3  1    1  5     3      0    22   22      0    no 1995-1998 20-40
4  1    1  6     4      0    22   22      0    no 1995-1998 20-40
5  1    2  4     5     22   995  973      0    no 1995-1998 20-40
6  1    2  5     6     22   995  973      0    no 1995-1998 20-40
7  1    2  6     7     22   995  973      0    no 1995-1998 20-40
```

```
> events(msebmt)
$Frequencies
        to
from       Tx  Rec   AE Rec+AE  Rel Death no event total entering
  Tx        0  785  907      0   95   160       332          2279
  Rec       0    0    0    227  112    39       407           785
  AE        0    0    0    433   56   197       221           907
  Rec+AE    0    0    0      0  107   137       416           660
  Rel       0    0    0      0    0     0         0             0
  Death     0    0    0      0    0     0         0             0
```

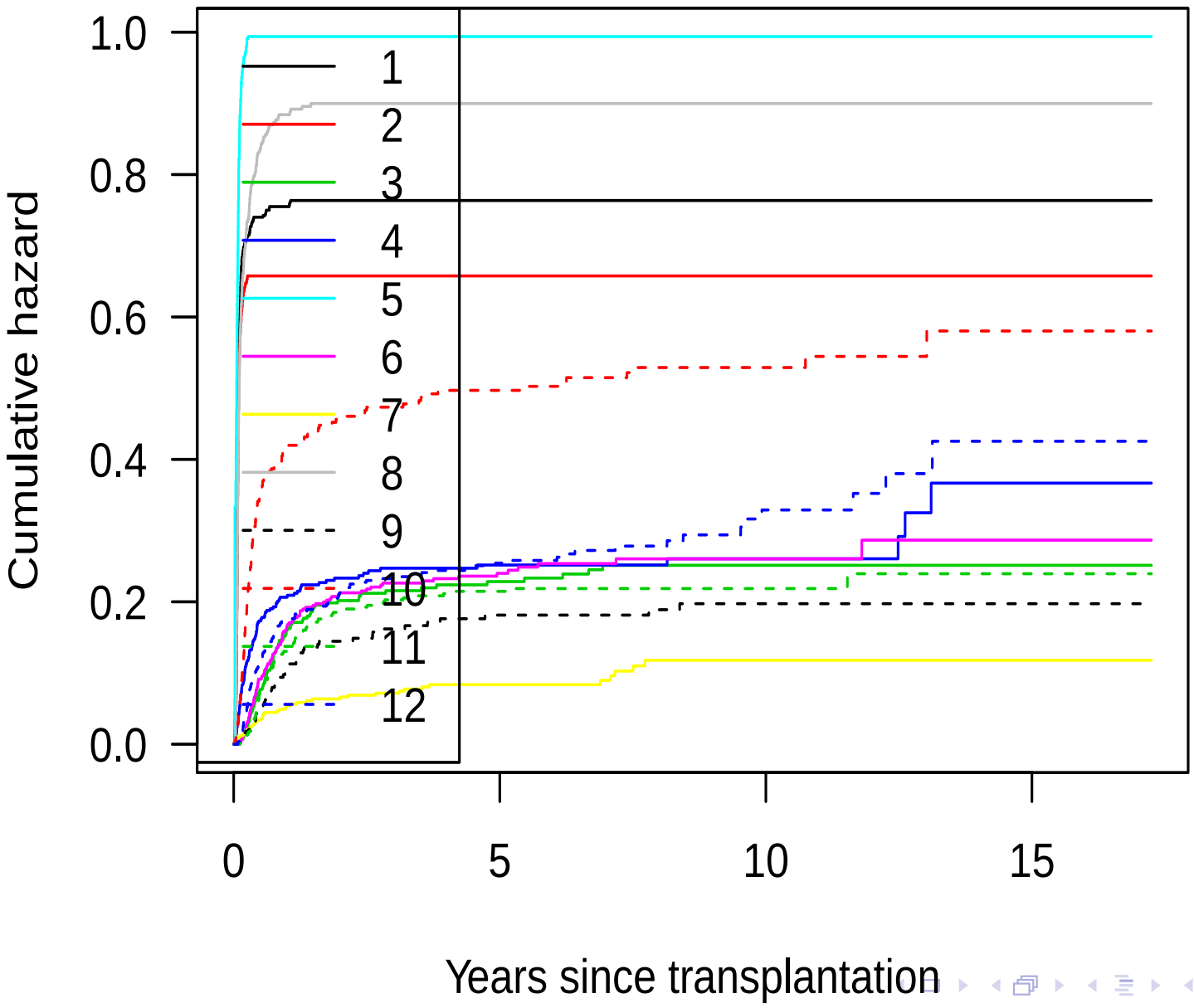
Package *mstate*

```
> covs <- c("match", "proph", "year", "agecl")
> msebmt <- expand.covs(msebmt, covs, longnames=FALSE)
> msebmt[msebmt$id == 1, -c(9, 10, 12:48, 61:84)]
An object of class 'msdata'

Data:
  id from to trans Tstart Tstop time status      year year2.1 year2.2 year2.3
1  1    1  2     1      0    22   22      1 1995-1998       1       0       0
2  1    1  3     2      0    22   22      0 1995-1998       0       1       0
3  1    1  5     3      0    22   22      0 1995-1998       0       0       1
4  1    1  6     4      0    22   22      0 1995-1998       0       0       0
5  1    2  4     5     22   995  973      0 1995-1998       0       0       0
6  1    2  5     6     22   995  973      0 1995-1998       0       0       0
7  1    2  6     7     22   995  973      0 1995-1998       0       0       0
  year2.4 year2.5 year2.6 year2.7 year2.8 year2.9 year2.10 year2.11 year2.12
1       0       0       0       0       0       0        0        0        0
2       0       0       0       0       0       0        0        0        0
3       0       0       0       0       0       0        0        0        0
4       1       0       0       0       0       0        0        0        0
5       0       1       0       0       0       0        0        0        0
6       0       0       1       0       0       0        0        0        0
7       0       0       0       1       0       0        0        0        0
```

Package *mstate*

```
> msebmt[, c("Tstart", "Tstop", "time")] <- msebmt[, c("Tstart", "Tstop",
"time")]/365.25
> c0 <- coxph(Surv(Tstart, Tstop, status) ~ strata(trans), data=msebmt,
method="breslow")
> msf0 <- msfit(object=c0, vartype="greenwood", trans=tmat)

> head(msf0$Haz)
         time         Haz trans
1 0.002737851 0.000000000     1
2 0.008213552 0.000000000     1
3 0.010951403 0.000000000     1
4 0.013689254 0.000000000     1
5 0.016427105 0.000443066     1
6 0.019164956 0.001333142     1
> tail(msf0$Haz)

         time       Haz trans
6199 12.48460 0.3800455    12
6200 12.61602 0.3800455    12
6201 13.02396 0.3800455    12
6202 13.10609 0.3800455    12
6203 13.12799 0.4255001    12
6204 17.24572 0.4255001    12

> plot (msf0, las=1, lty= rep(1:2, c(8,4)), xlab="Years since transplantation")
```

Package *mstate*



Package *mstate*

```
> pt0 <- probtrans(msf0, predt=0, method="greenwood")
> summary(pt0)
An object of class 'probtrans'

Prediction from state 1 (head and tail):

         time   pstate1      pstate2      pstate3 pstate4 pstate5      pstate6          se1
1 0.000000000 1.0000000 0.0000000000 0.0000000000       0       0 0.0000000000 0.0000000000
2 0.002737851 0.9995610 0.0000000000 0.0004389816       0       0 0.0000000000 0.0004388852
3 0.008213552 0.9978051 0.0000000000 0.0021949078       0       0 0.0000000000 0.0009805148
4 0.010951403 0.9956102 0.0000000000 0.0039508341       0       0 0.0004389816 0.0013851313
5 0.013689254 0.9907814 0.0000000000 0.0079016681       0       0 0.0013169447 0.0020023724
6 0.016427105 0.9863916 0.0004389816 0.0118525022       0       0 0.0013169447 0.0024274584
           se2          se3 se4 se5          se6
1 0.0000000000 0.0000000000   0   0 0.0000000000
2 0.0000000000 0.0004388852   0   0 0.0000000000
3 0.0000000000 0.0009805148   0   0 0.0000000000
4 0.0000000000 0.0013143406   0   0 0.0004388852
5 0.0000000000 0.0018550683   0   0 0.0007598375
6 0.0004388852 0.0022674569   0   0 0.0007598375
...

time   pstate1   pstate2    pstate3   pstate4   pstate5   pstate6         se1
513 12.48460 0.1366292 0.1650151 0.09331023 0.1593697 0.1811612 0.2645145 0.008491577
514 12.61602 0.1320749 0.1650151 0.09331023 0.1593697 0.1811612 0.2690688 0.009350412
515 13.02396 0.1320749 0.1650151 0.08997772 0.1593697 0.1811612 0.2724013 0.009350412
516 13.10609 0.1265718 0.1650151 0.08997772 0.1593697 0.1811612 0.2779044 0.010455556
517 13.12799 0.1265718 0.1650151 0.08997772 0.1521256 0.1811612 0.2851485 0.010455556
518 17.24572 0.1265718 0.1650151 0.08997772 0.1521256 0.1811612 0.2851485 0.010455556
```

Package *mstate*

Tran-sition	Match	Prophylaxis	Year of transplant		Age at transplant	
			1990–1994	1995–1998	20–40	> 40
1	*-0.167* (0.085)	***-0.366*** (0.093)	***0.401*** (0.100)	***0.521*** (0.103)	0.049 (0.089)	0.199 (0.102)
2	-0.111 (0.079)	***-0.278*** (0.083)	0.023 (0.084)	-0.114 (0.091)	0.123 (0.083)	0.067 (0.101)
3	0.196 (0.224)	0.385 (0.227)	0.442 (0.245)	0.221 (0.302)	-0.094 (0.232)	-0.232 (0.322)
4	-0.003 (0.181)	-0.056 (0.179)	-0.359 (0.193)	*-0.476* (0.218)	***0.766*** (0.229)	*0.934* (0.264)
5	0.190 (0.153)	-0.282 (0.196)	-0.095 (0.191)	-0.151 (0.190)	0.292 (0.188)	*0.470* (0.205)
6	*0.426* (0.214)	0.268 (0.221)	-0.210 (0.263)	0.055 (0.259)	-0.255 (0.223)	-0.101 (0.264)
7	0.244 (0.405)	-0.008 (0.378)	*-0.836* (0.398)	*-0.980* (0.442)	0.150 (0.491)	***1.465*** (0.481)
8	0.126 (0.113)	0.125 (0.125)	***0.528*** (0.135)	***0.930*** (0.141)	***-0.393*** (0.116)	*-0.328* (0.142)
9	-0.414 (0.352)	0.159 (0.321)	-0.311 (0.300)	-0.580 (0.433)	0.173 (0.367)	0.423 (0.433)
10	0.008 (0.168)	0.324 (0.166)	*-0.644* (0.173)	-0.213 (0.195)	0.238 (0.205)	*0.495* (0.237)
11	-0.301 (0.248)	0.012 (0.247)	-0.024 (0.253)	-0.390 (0.277)	0.414 (0.250)	0.256 (0.304)
12	***0.572*** (0.179)	-0.118 (0.217)	-0.362 (0.228)	-0.352 (0.238)	***0.760*** (0.272)	***1.337*** (0.287)

The computational problem with interval censored observations

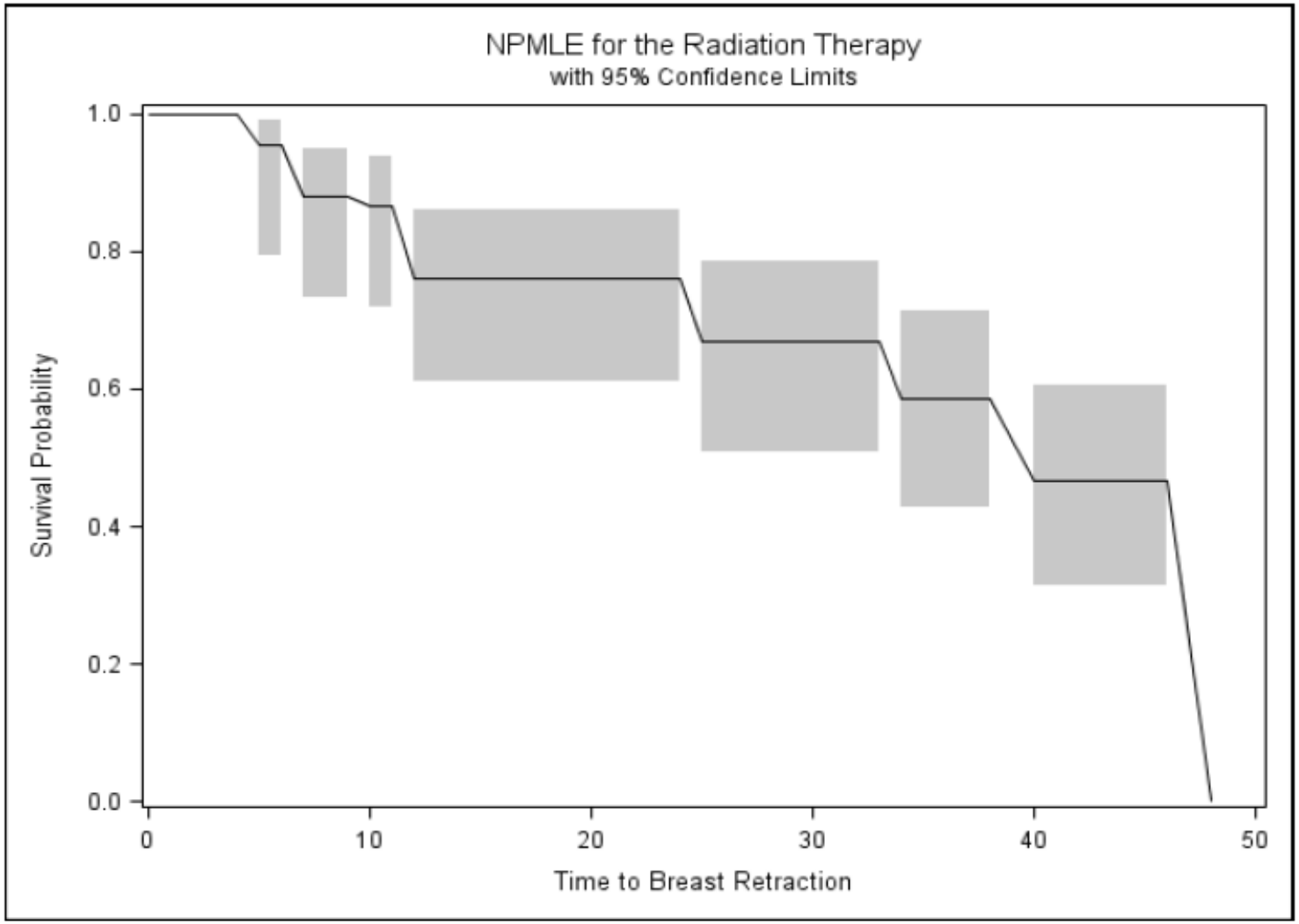
Dealing with interval-censored data in multistate models

Two-state (survival) model with interval-censored data

- Non-parametric estimation of the survival function
- The equivalent of the Kaplan-Meier estimation for interval-censored data was first proposed by Peto (1973), and Turnbull (1976) then improved the numerical algorithm to estimate the survival function (EM algorithm)
- From the data $\{(L_i, R_i], i = 1, 2, ..., n\}$ a set of non-overlapping intervals $\{(q_1, p_1], (q_2, p_2], ..., (q_m, p_m]\}$ is generated over which $S(t)$ is estimated : the NPMLE of $S(t)$ can decrease only on intervals $(q_j, p_j]$.
- PH model $S(t; Z) = S_0(t)^{\exp(\beta Z)}$: estimation (non-parametric) of the baseline survival function $S_0(t)$ and of the regression coefficients (Finkelstein, 1986; Alioum and Commenges, 1996; Goetghebeur and Ryan, 2000).

NPMLE of the survival curve

Figure 3 NPMLE for the RT Group with Confidence Intervals



Illness-death model with interval-censored data

- Observations pattern : times of transition from 0 to 1 are interval-censored, but times to the absorbing state are assumed to be known exactly or to be right-censored
- If you're just interested in α_{01}, you can use survival methods for interval censored data, but α_{01} will be underestimated (e.g. incidence of dementia ; Joly et al., 2002)
- Problem with survival analysis: using death as right-censoring leads to a biased estimates of age-specific incidence of dementia
- A more adapted model is an illness-death model (Commenges et al., 1998; Commenges et al., 2004)

Incidence of Dementia : survival and illness-death model

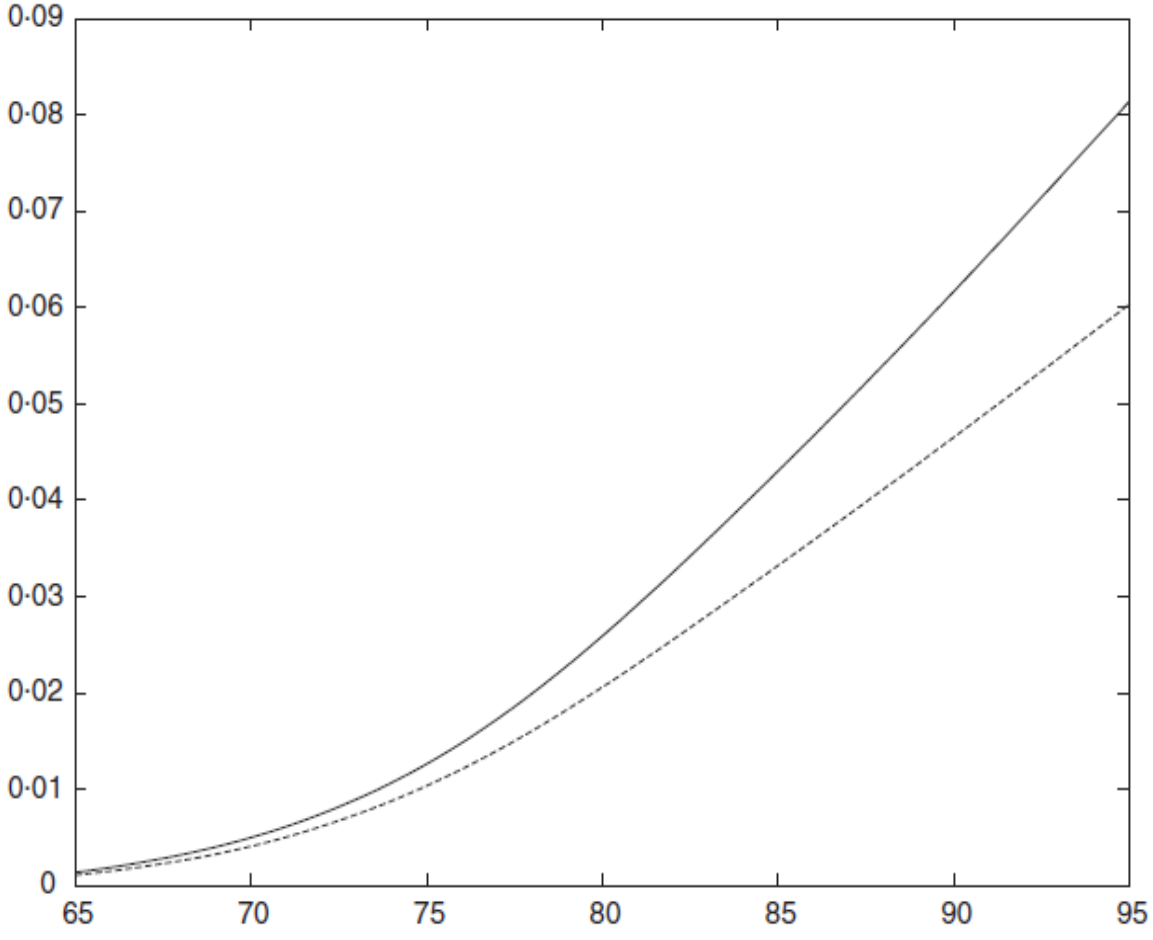


Fig. 5. Estimated age-specific incidence of dementia. The solid curve represents the intensity function α_{01} estimated using the 'illness–death model' and the dotted curve represents the intensity function estimated using a (two-state) survival analysis (Paquid, 1999).

Parametric/Smooth non-parametric

- Parametric approach : easy to implement using likelihood methods
- Smooth non-parametric via penalized likelihood (Joly et al., 2002)

$$\alpha_{hj}(t;Z) = \alpha_{hj,0}(t)\exp(\beta_{hj}Z), hj \in \{01,02,12\}$$

$$pl(\alpha_{01}(.),\alpha_{02}(.),\alpha_{12}(.);\beta_{01},\beta_{02},\beta_{12}) =$$
$$\mathcal{L}(\alpha_{hj}(.);\beta_{hj}) - \kappa_{01}\int_0^{+\infty}[\alpha''_{01,0}(u)]^2du$$
$$-\kappa_{02}\int_0^{+\infty}[\alpha''_{02,0}(u)]^2du - \kappa_{12}\int_0^{+\infty}[\alpha''_{12,0}(u)]^2du$$

Parametric/Smooth non-parametric

- NPMLE $\hat{\alpha}_{01}(.), \hat{\alpha}_{02}(.), \hat{\alpha}_{12}(.)$ are approximated using M-splines

$$\tilde{\alpha}_{hj}(t) = \sum_l \theta_{hj}^l M_l(t)$$

$$\tilde{\Lambda}_{hj}(t) = \sum_l \theta_{hj}^l I_l(t)$$

- $\kappa_{01}, \kappa_{02}, \kappa_{12}$ are estimated simultaneously by maximizing the standard crossvalidation score

SmoothHazard : A new R package to fit illness-death model (Weibull or penalized likelihood) for left-truncated and/or interval-censored data (written by Tourraine, Joly and Gerds)

Package *SmoothHazard*

Package 'SmoothHazard'

February 22, 2013

Title Smooth hazard Weibull regression

Version 1.0.8

Author Celia Touraine, Pierre Joly, Thomas A. Gerds

Description User friendly Weibull regression for interval censored data

---

`illness-death model` *Fit an illness-death model*

Description

Fit an illness-death model using either a semi-parametric approach (penalized likelihood with an approximation of the transition intensity functions by linear combination of M-splines) or a parametric approach (specifying Weibull distributions on the transition intensities). Left-truncated, right-censored, and interval-censored data are allowed. State 0 corresponds to the initial state, state 1 to the transient one, state 2 to the absorbant one. The allowed transitions are: 0 –> 1, 0 –> 2 and 1 –> 2.

Package *SmoothHazard*

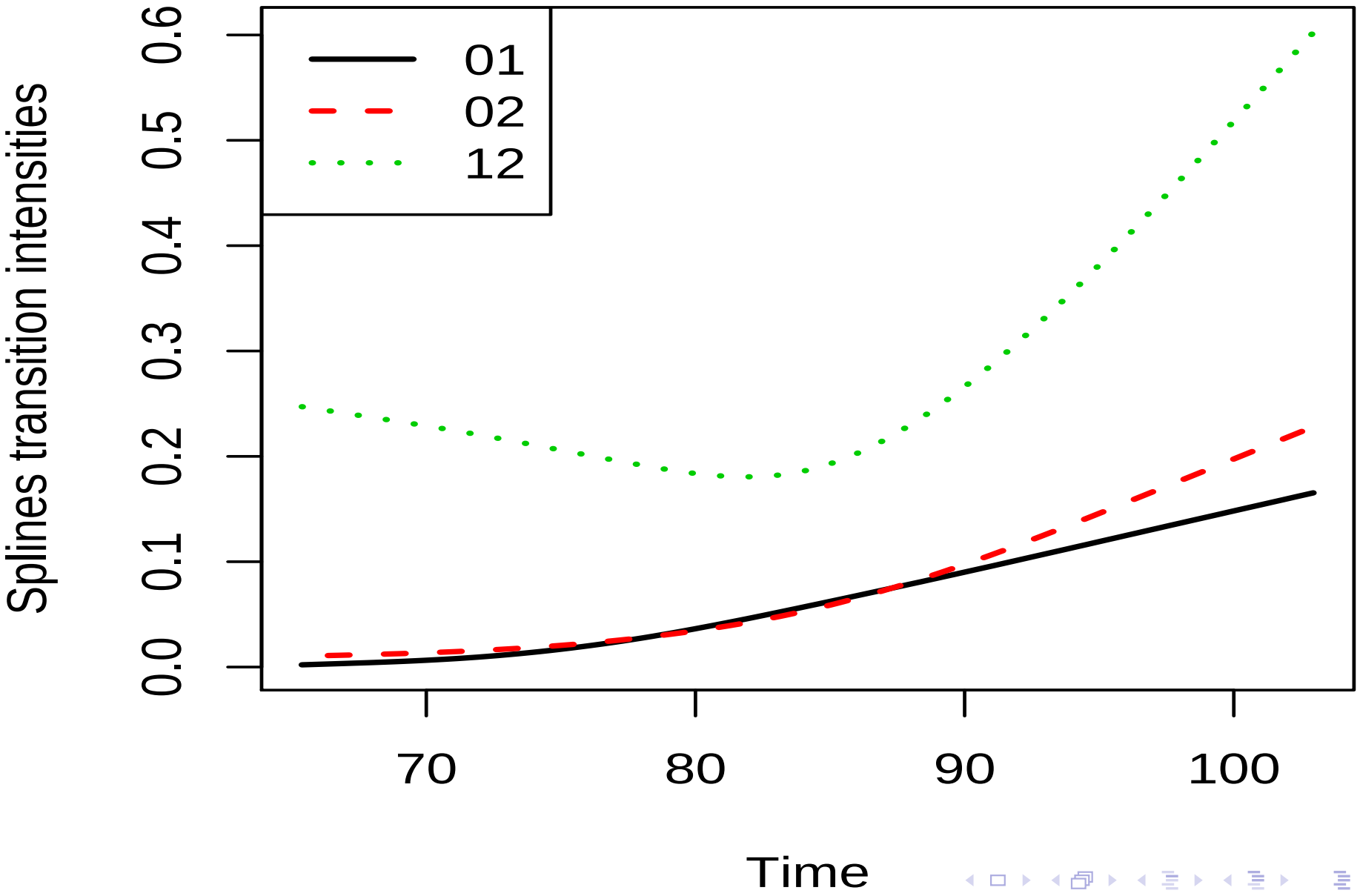
- t0: left truncation time
- t1: for diseased subjects, left endpoint of the censoring interval; for non diseased subjects, right censoring time for 0->1 transition
- t2: for diseased subjects, right endpoint of the censoring interval; for non diseased subjects, right censoring time for the disease event
- t3: for subjects who died, time of death; for alive subjects, censoring time for the death event

```
> library(SmoothHazard)
> data(Paq1000)
> head(Paq1000)
  dementia death      t0       t1       t2       t3 certif gender
1        1     1 72.3333 82.34014 84.73303 87.93155      0      0
2        0     1 77.9167 78.93240 78.93240 79.60048      0      1
3        0     1 79.9167 79.91670 79.91670 80.92423      0      0
4        0     1 74.6667 78.64750 78.64750 82.93501      1      1
5        0     1 76.6667 76.66670 76.66670 79.16636      0      1
6        0     0 66.2500 71.38070 71.38070 84.16975      1      0
```

Package *SmoothHazard*

```
> fit <- idm(formula02=Hist(time,event=death,entry=entry)~certif,
+ formula01=Hist(time=list(L,R),event=dementia)~certif,
+ data=d, hazard="Splines")
> print(fit)
Call:
idm(formula01 = Hist(time = list(L, R), event = dementia) ~ certif,
    formula02 = Hist(time, event = death, entry = entry) ~ certif,
    data = d, hazard = "Splines")

Illness-death model using a penalized likelihood approach with splines
approximation for the intensity functions.

number of subjects:  1000
number of events '0-->1':  186
number of events '0-->2' or '0-->1-->2':  724
number of subjects:  1000
number of covariates:  1 1 1

Smoothing parameters:
      transition01 transition02 transition12
knots        7e+00        7e+00            7
kappa        1e+06        5e+05        20000

             coef SE.coef     HR          CI       Wald p.value
certif_01 -0.4797  0.2053 0.6190 [0.41;0.93] 5.46092875 0.01945
certif_02  0.1564  0.1265 1.1693 [0.91;1.50] 1.52834594 0.21636
certif_12 -0.0682  0.2311 0.9341 [0.59;1.47] 0.08707806 0.76793

                         Without cov  With cov
Penalized log likelihood    -3073.08 -3069.819

> plot(fit)
```

Package *SmoothHazard*

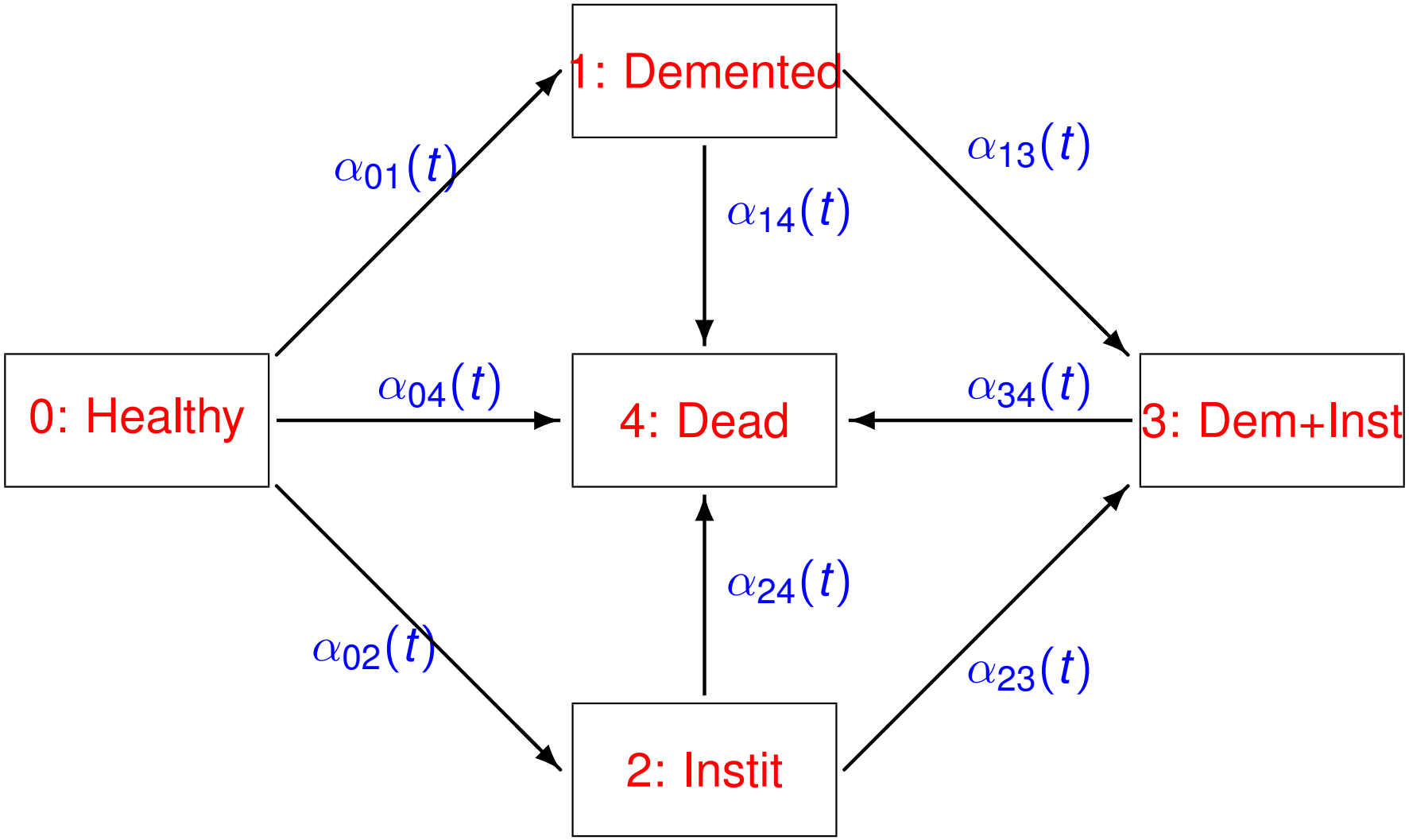


Package *SmoothHazard*

Other features of SmoothHazard for both Weibull model and penalized likelihood

- Plot baseline survival functions
- Predict transition probabilities
- Predict life expectancies (in state 0; for a non diseased subject; for a diseased subject) at time s for a subject with a given vector of covariates
- Fit survival model using either a penalized likelihood approach or a parametric (Weibull) approach from left-truncated and interval-censored data

More complex models : Dementia-Institution-Death



More complex models : Dementia-Institution-Death

- Smooth non-parametric by penalized likelihood
- Need some additional assumptions (Joly et al., 2009)

$$\alpha_{14}(t) = \alpha_{04}(t)e^{\theta_{14}} \qquad \alpha_{23}(t) = \alpha_{01}(t)e^{\theta_{23}}$$
$$\alpha_{24}(t) = \alpha_{04}(t)e^{\theta_{24}} \qquad \alpha_{13}(t) = \alpha_{02}(t)e^{\theta_{13}}$$
$$\alpha_{34}(t) = \alpha_{04}(t)e^{\theta_{34}}$$

$$pl(\alpha_{01}(.), \alpha_{02}(.), \alpha_{04}(.); \theta) =$$
$$\mathcal{L}(\alpha_{01}, \alpha_{02}, \alpha_{04}; \theta) - \kappa_{01} \int_0^{+\infty} [\alpha''_{01}(u)]^2 du$$
$$-\kappa_{02} \int_0^{+\infty} [\alpha''_{02}(u)]^2 du - \kappa_{04} \int_0^{+\infty} [\alpha''_{04}(u)]^2 du$$

Panel data

States $X^i(t)$ of individual $i (i = 1, 2, ..., n)$ is only known at discrete consecutive follow-up times $t = (t_{i,0}, t_{i,1}, ..., t_{i,m_i})$, and the exact time of transitions from one state to another is not observed

- Likelihood conditionally on $X(t_{i,0})$

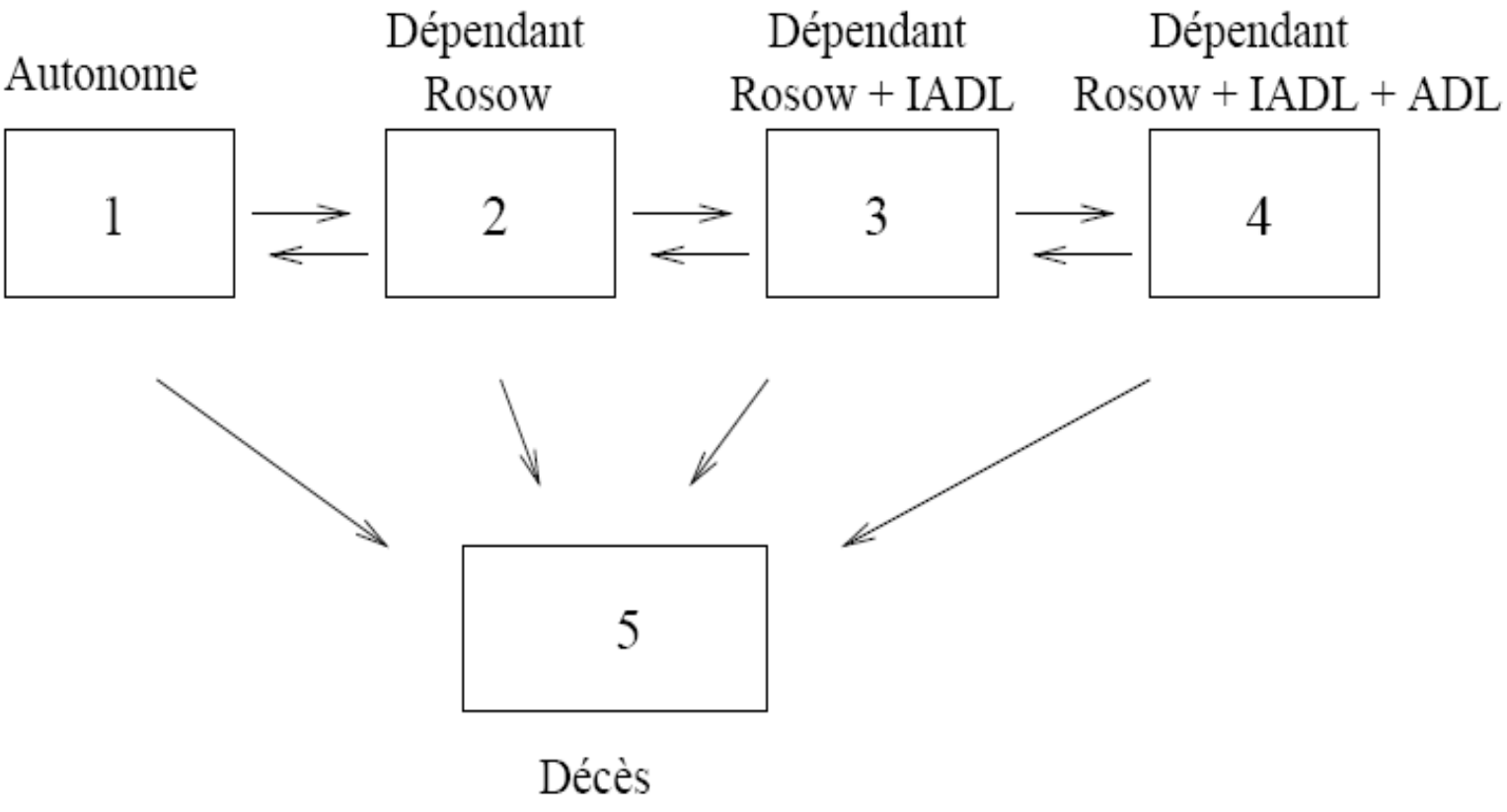
$$\prod_{i=1}^{n} \prod_{j=1}^{m_i} P_{X(t_{i,j-1}),X(t_{i,j})}(t_{i,j-1}, t_{i,j}; Z_i(t_{i,j-1}))$$

If $X(t_{i,m_i}) = K$ (single absorbing state) and t_{i,m_i} is the exact time of transition to K , then the expression of the last term of the product is given by

$$\sum_{l \neq K} P_{X(t_{i,m_i-1}),l}(t_{i,m_i-1}, t_{i,m_i}; Z_i(t_{i,m_i-1}))\alpha_{l,K}(t_{i,m_i})$$

Example of more complex multi-state models

Disablement process in the elderly (Paquid study)



HMM/NHMM with PCI

- NHMM with piecewise constant intensities: partitioning time into 2 or more intervals and assume constant intensity in each interval
  - PIM : $\alpha_{hj}(t; Z(t)) = \alpha_{hj0}(t) \exp(\beta' Z(t))$ where $\alpha_{hj0}(t) = \alpha^{l}_{hj0}$ for $\tau_{l-1} < t \leq \tau_l, (l = 1, 2, ..., r)$
  - The expressions of transition probabilities $P_{hj}(s, t)$ become more complicated (even if we keep the idea of fitting of homogeneous models to a series of r intervals)
  - This method needs prior specification of the cutpoints; the choice of cutpoints should make sure that an appropriate number of observations fall in all intervals

Programs and Packages for fitting HMM/NHMM with PCI

MKVPCI

- Fortran program (Alioum and Commenges, 2001)
- Fitting general k-state Markov models ($k-1$ transient states and the kth state can be optionally chosen as an absorbing state) with PCI and covariates
- The output includes estimates of the baseline transition intensities, regression coefficients, transition intensities in time intervals $(\tau_{l-1}, \tau_l], l = 1, 2, ..., r$, together with their estimated standard errors, and the results of the multivariate hypotheses tests.

Example 2: disablement process in the elderly

File of input parameter for MKVPCI

disab.dat
15439
5
10
0 1 0 0 1
1 0 1 0 1
0 1 0 1 1
0 0 1 0 1
0 0 0 0 0
1
80.0
1 2 3 4 5 6 7 8 9 10
1
2
gender 0 1 2 3 4 5 6 7 8 9 10
educ 0 1 2 3 4 5 6 7 8 9 10
1
1
0

Example 2: disablement process in the elderly

Multi-state Markov model with piecewise constant intensities
Program developed by Ahmadou ALIOUM
ISPED - University Victor Segalen Bordeaux 2

Markov model with piecewise constant intensities : two periods

Data file : disab.dat
Number of records : 15439
Number of subjects : 3461
Number of states : 5
Number of transitions : 10
Number of artificial covariates : 1
Cut points : 80.000000
Number of observed covariates : 0
Number of parameters : 20
Exact transition times to the absorbing state : 1

Baseline transition intensities

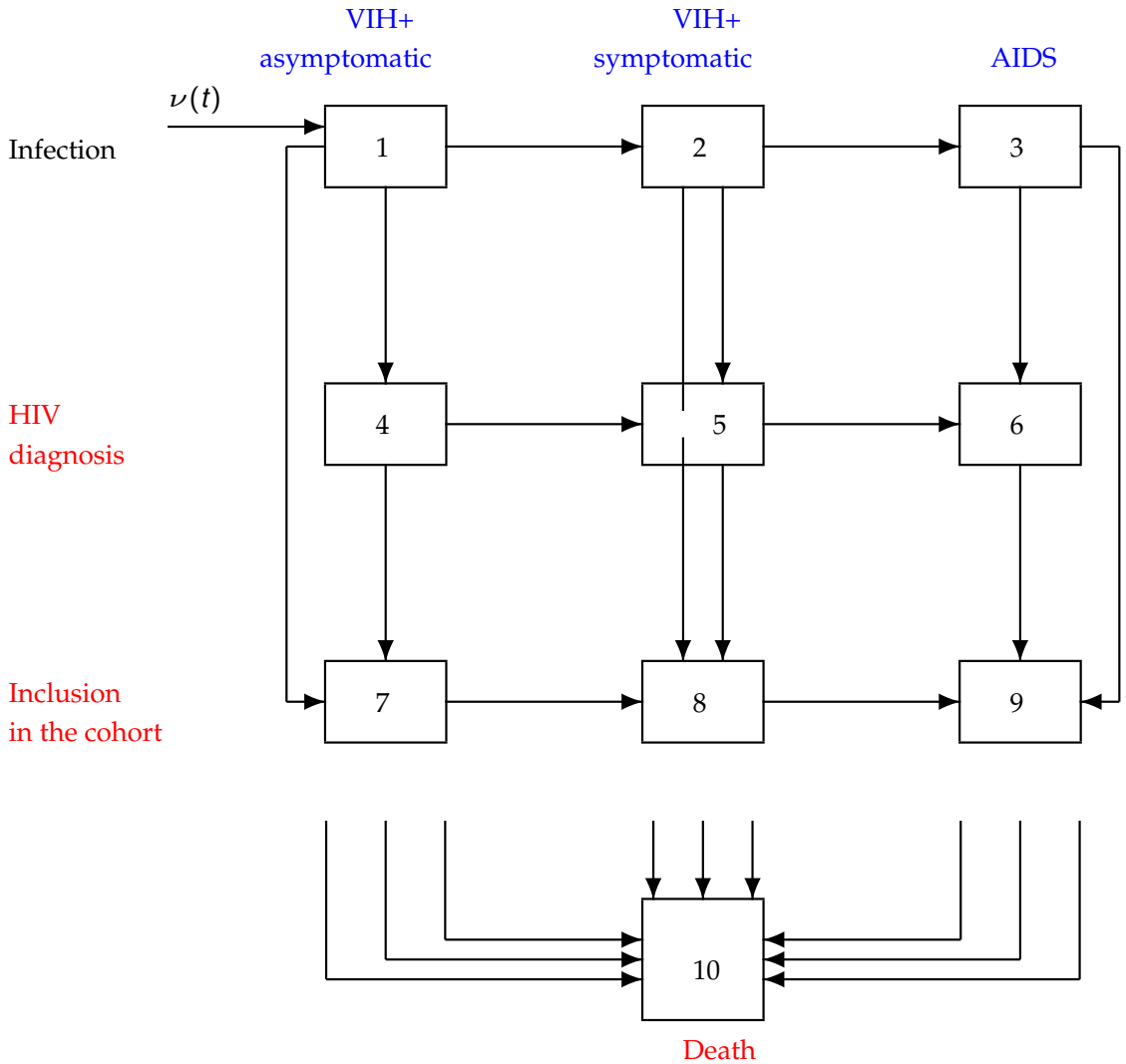
Time intervals

	Tau(0) - Tau(1)		Tau(1) - Tau(2)	
TRANSITION	ESTIMATES	STD ERROR	ESTIMATES	STD ERROR
1 –> 2	.28432	.01180	.57785	.04278
1 –> 5	.01099	.00274	.02167	.01165
2 –> 1	.17916	.00870	.09423	.01089
2 –> 3	.13516	.00661	.29438	.01282
2 –> 5	.02707	.00339	.04016	.00665
3 –> 2	.17757	.01238	.05033	.00475
3 –> 4	.07652	.00834	.14481	.00827
3 –> 5	.05382	.00775	.08822	.00711
4 –> 3	.10510	.02144	.04460	.00794
4 –> 5	.21518	.02759	.35568	.01823

R package *msm*

- Package *msm* developped by Christopher H. Jackson (Journal of Statistical Software, 2011, vol. 38, issue 8)
- Markov model with any number of states and any pattern of transitions to panel data : HMM, NHMM with PCI
- Proportional intensities models : piecewise constant time-dependant covariates
- Exact death times, censored states (alive but in an unknown disease state at the end of the study)
- Hidden Markov models in which the states of the Markov chain are not observed : observed data X_{ij} are governed by some probability distribution conditionally on the unobserved state S_{ij}
- Tools for model assessment

Estimation of HIV incidence from a prevalent cohort



Estimation of HIV incidence from a prevalent cohort

- Infection occurs according to a non-homogeneous Poisson process with intensity $\nu(t)$
- Non-homogeneous Markov process with piecewise constant intensities
- Data from a prevalent cohort study : a total of n subjects were enrolled in the cohort by time T^*
  - s_i time of first HIV positive test
  - $X_i(t_{i0}), X_i(t_{i1}), \ldots, X_i(t_{im_i})$
- Sampling criterion : a subject i is enrolled only if $t_{i0} \leq T^*$
- The number n is the realization of a random variable N, and the observation of $N = n$ brings information on $\nu(t)$

Estimation of HIV incidence from a prevalent cohort

- Conditionally on the occurrence of a certain number of infections in the period $[0, T^*]$, the times of infection a iid with density function
  $\nu^*(t) = \nu(t) / \int_0^{T^*} \nu(u)du$
- Conditional on $N = n$, the likelihood of the trajectories of observed subjects is
  $L_1 = \prod_{i=1}^{n} \mathcal{L}_i / P_I$
  where $P_I = \int_0^{T^*} \nu^*(u) P^*(u) du$,
  and $\mathcal{L}_i = P(s_i, t_{i0}, x_{i0}) \prod_{j=1}^{m_i} P_{X_i(t_{i,j-1}), X_i(t_{i,j})}(t_{i,j-1}, t_{i,j})$

Estimation of HIV incidence from a prevalent cohort

- N follows a Poisson distribution with parameter $A = \int_0^{T^*} \nu(u) P^*(u) du$: $L_2 = P(N = n)$
- Full likelihood : $L = L_1 \times L_2$
- Smooth estimate of HIV is obtained by maximazing the penalized log-likelihood : $pl(\nu, \alpha) = \log L_1 + \log L_2 - \lambda \int \nu''(u)^2 du$

Estimation of HIV incidence from a prevalent cohort

Data from a hospital-based surveillance system of HIV infection implemented since 1987 in Aquitaine region, Southwestern France : homosexuals/bisexuals ($\lambda = 100$; GECSA, 1985-2000)

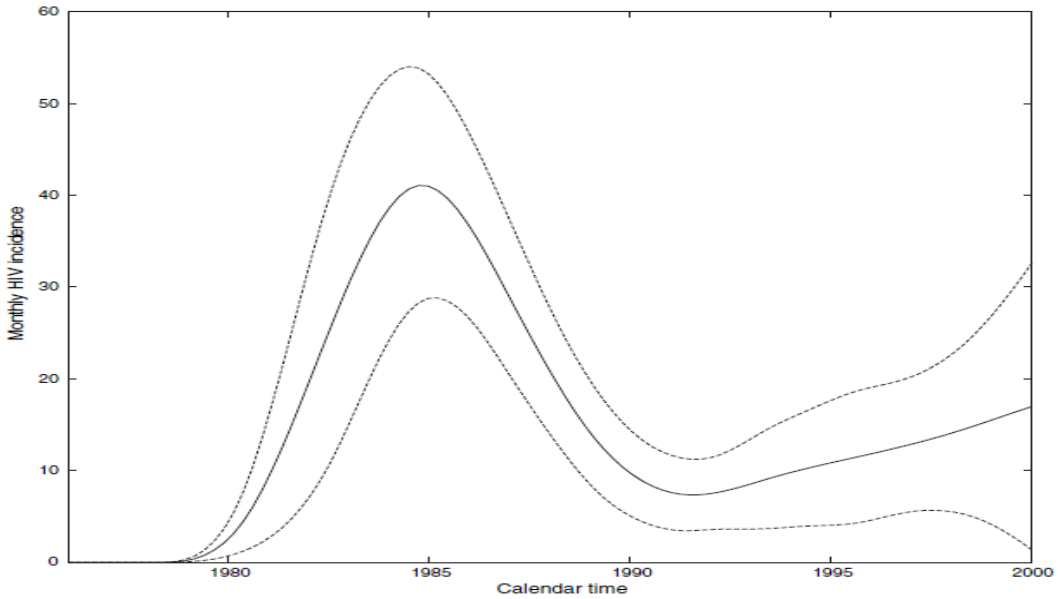


Fig. 3. Estimated monthly incidence of HIV among homosexual and bisexual patients in Aquitaine, south-western France, with 95% pointwise Bayesian confidence limits ($\lambda = 100$), 1985–2000

Concluding remarks

- Methods for inference in MSM from selected sample or/and interval-censored data are useful in epidemiology
- Some software are now available for fitting simple MSM models (mstate, msm, smoothhazard, among others)
- Some challenges for modelling in the MSM framework
  - more complex processes
  - unobserved heterogeneity, clustering
  - dependant LTF or missing data
  - model checking and validation

Some selected references

PK Andersen, O Borgan, RD Gill and N Keiding (1993), *Statistical Models Based on Counting Processes*, Springer, NY

P Hougaard (2000), *Analysis of Multivariate Survival Data*, Springer, NY

Aalen, Borgan and Gjessing (2008), *Event History Analysis*, Springer, NY

Beyersmann, Schumacher and Allignol (2012), *Competing Risks and Multistate Models with R*, Springer, NY

A Alioum et al. (2005), *Applied Statistics* 54, 739-752

PK Andersen and Keiding (2002), *Statist. Methods Med. Res.* 11, 91-115

D Commenges (2002), *Statist. Methods Med. Res.* 11, 167-182

D Commenges et al. (2004), *Statist. Med.* 23, 199-210

P Joly et al. (2002), *Biostatistics* 3, 433-443

de Wreede, Fiocco and Putter (2011), *Journal of Statistical Software* Volume 38, Issue 7,