Far East J. Theo. Stat. **19**(2) (2006), 185-201
(A special volume on Biostatistics and related areas)
This paper is available online at http://www.pphmj.com

A SEMI-MARKOV FRAILTY MODEL FOR MULTISTATE AND CLUSTERED SURVIVAL DATA

YOHANN FOUCHER, PHILIPPE SAINT-PIERRE,
JEAN-PIERRE DAURES and JEAN-FRANCOIS DURAND*

Institut Universitaire de Recherche Clinique
Laboratoire Epidemiologie et Biostatistique
UFR Medecine Site Nord, 640
av du Doyen Gaston Giraud
34295 Montpellier Cedex 05, France
e-mail: yohann.foucher@iurc.montp.inserm.fr

*Departement des Maladies Infectieuses
CHU de Nice
Route Saint Antoine de Ginestiere
06202 Nice, France

Abstract

In the longitudinal analysis of chronic diseases, investigators are often interested in studying multiple transitions between various states of disease progression. Semi-Markov models may be used for modeling this type of process. We propose a novel semi-Markov model with random effects. This allows for the dependence of transition times within subgroups. Another originality consists in the introduction of a generalized Weibull distribution for the hazard functions, offering a more global parametric method than those frequently used. Laplace transform is used to define a marginal likelihood function in order to estimate the regression parameters. A real dataset for HIV is analyzed to illustrate the methodology. Correlation subgroups are defined

---

2000 Mathematics Subject Classification: 60K15, 62P10.

Keywords and phrases: semi-Markov process, gamma frailty, generalized Weibull, Laplace transform, HIV.

Received May 2, 2006

according to the subjects, because a same transition can be observed several times for a given subject. Based on our model, we estimate individual effects and test the independence of observed transition times. The results demonstrate that all the observations are independent whoever the subject and that traditional semi-Markov models may thus be used in our application.

1. Introduction

Multilevel survival data are becoming increasingly common in epidemiological and longitudinal studies. Indeed, the progression of a disease cannot be summarized by one terminal event. For example, in HIV (Human Immunodeficiency Virus) follow-up, a patient may evolve through various states of gravity. This type of method has recently been applied with success by Alioum et al. [3] or Jackson et al. [10]. There are numerous other clinical examples such as cancerology or asthma by Combescure et al. [6] or Saint-Pierre et al. [14].

Homogeneous Markov models are often used in this context, but the hazard rates of transition are assumed to be constant over the time spent in the state (exponential distribution). The evolution of the process is thus independent from the waiting times. Semi-Markov models make the dependence possible, by modeling waiting time distributions. Perez-Ocon and Ruiz-Castro [12] thus proposed a semi-Markov model based on Weibull distribution, constituting the theoretical background of the extension presented in this paper.

Another modeling issue concerns the repeated occurrence of the same type of event during patient follow-up. In practice, the assumption made in modeling a time-dependent process, is that the observed transition times are independent, given the observed covariates. Escolano et al. [7], Yau and Huzurbazar [15] or Foucher et al. [8] all made this assumption. It may be the case that the transition times of individuals in some subgroups of the population are associated since members of these groups share a common unobserved trait. For example, there may be an association in times to events such as disease or death between members of the same family. In this paper, we study the association of repeated events for a particular individual. Models with random effects provide an interesting method for dealing with such a correlated structure. In survival analysis, those random effects, which act multiplicatively on the baseline hazard functions, are called *frailties*. Several authors propose various frailty models, according to the distribution of frailties and estimation procedure. Klein [11] proposed an EM algorithm in order to estimate a Cox model with random effects. Aalen and Husebye [2] used the Laplace transform to define a marginal likelihood function of a Weibull model with frailty. This theory was recently extended to multivariate survival data by Hougaard [9].

In this paper, we propose a novel semi-Markov frailty model for multi-state and clustered longitudinal analysis. The development of such a type of model was motivated by the study of the evolution of patients HIV-infected, using a cohort followed up at Nice University Hospital (France). The hazard functions with the associated covariates are specific to each transition. The shape of the hazard function generalizes the Weibull distribution, by allowing a U or inverse U shape. Frailties are assumed to be gamma distributed. The same transitions are allowed to be independent given the individual frailty. All these choices of distribution are argued in this paper. Parameters and their variances are estimated by maximizing the likelihood. The following developments are based on the Laplace transform, as defined by Aalen and Hougaard. To the best of our knowledge, such a model has never been published. Only Ripatti et al. [13] defined a three-state frailty model although they used piecewise constant hazard functions without covariates.

This paper is organized as follows: Section 2 introduces the semi-Markov frailty model and the estimation procedure. In Section 3, these methods are applied to a cohort of patients HIV-infected. Section 4 contains the discussion.

2. Model Formulation

2.1. Modeling semi-Markov process without frailty

Let $E = \{1, 2, ..., r\}$ be a finite state space. Consider the random processes $(T, X) = \{(T_n, X_n) : n \geq 0\}$, in which $0 = T_0 < T_1 < \cdots < T_n$

are the consecutive times of entrance into the states $X_0, X_1, \ldots, X_n \in E$, with $X_{p+1} \neq X_p$, $\forall p \geq 0$ and X_p not persistent. n represents the number of jumps. The sequences $X = \{X_n, n \geq 0\}$ form an embedded homogeneous Markov chain. The probabilities of jumping from i to j, associated with this chain, can be written as $P_{ij} = P(X_{n+1} = j \mid X_n = i)$. If state i is not persistent, then $P_{ij} \geq 0$ for $i \neq j$ and $P_{ij} = 0$ for $i = j$. Otherwise, if state i is persistent, then $P_{ij} = 0$ for $i \neq j$ and $P_{ij} = 1$ for $i = j$. In the following developments, we will suppose that state i is transient. As we can see, the Markov chain does not deal with the duration of states. The waiting times are explicitly defined. These processes (T, X) are called *semi-Markovian*, if the distribution of waiting times $(T_{n+1} - T_n)$ satisfies $P(T_{n+1} - T_n \leq x, X_{n+1} = j \mid X_0, T_0, X_1, \ldots, X_n, T_n)$ $= P(T_{n+1} - T_n \leq x, X_{n+1} = j \mid X_n)$. The density probability function of the waiting time in state i before passing into state j, is given as:

$$f_{ij}(x) = \lim_{h \to 0^+} \frac{P(x < T_{n+1} - T_n < x + h \mid X_{n+1} = j, X_n = i)}{h}. \tag{1}$$

The density distribution and corresponding value of parameters can vary between each type of transition. As usual in survival analysis, we deduce from $f_{ij}(x)$ the corresponding distribution function, survival function and hazard function, that is to say $F_{ij}(x)$, $S_{ij}(x)$ and $\lambda_{ij}(x)$, respectively:

$$\begin{aligned} \lambda_{ij}(x) &= \lim_{h \to 0^+} \frac{P(x < T_{n+1} - T_n < x + h \mid T_{n+1} - T_n \geq x, X_{n+1} = j, X_n = i)}{h} \\ &= \frac{f_{ij}(x)}{S_{ij}(x)}. \end{aligned} \tag{2}$$

The marginal density probability function is deduced from (2)

$$f_{i.}(x) = \sum_{j \neq i} P_{ij} f_{ij}(x). \tag{3}$$

By definition, the hazard function of the semi-Markovian process corresponds to the probability of jumping towards state j, given that the process occupies state i for a duration x:

$$\alpha_{ij}(x) = \lim_{h \to 0} \frac{P[x \le T_{n+1} - T_n < x + h,\ X_{n+1} = j \mid T_{n+1} - T_n \ge x,\ X_n = i]}{h}$$

$$= \frac{P_{ij} f_{ij}(x)}{S_{i.}(x)} \quad \text{with } i \ne j,\ i,\ j \in E \ \text{ and } \ \alpha_{ii}(x) = -\sum_{j \ne i} \alpha_{ij}(x). \tag{4}$$

In order to take covariates in the model into account, we use the assumption of risk proportionality. The additional hypothesis is that covariates act on the waiting time distributions. Suppose n_{ij} transitions $i \to j$ for a patient, $\forall i \ne j \in E$. These transitions occur at the following waiting times $x_{ij1}, x_{ij2}, \ldots, x_{ijn_{ij}}$. At these times, covariates are respectively $z_{ij1}, z_{ij2}, \ldots, z_{ijn_{ij}}$, in which $z_{ijp} = (z_{ijp}^{1}, z_{ijp}^{2}, \ldots, z_{ijp}^{k_{ij}})$ is the vector of k_{ij} covariates specific to the pth transition $i \to j$. The *hazard function* is thus defined by $\lambda_{ij}(x_{ijp}, z_{ijp}) = \lambda_{0,ij}(x_{ijp}) \exp(\beta_{ij}^{T} z_{ijp})$ to obtain a strictly positive function, in which $\beta_{ij} = (\beta_{ij}^{1}, \beta_{ij}^{2}, \ldots, \beta_{ij}^{k_{ij}})$ is the vector of k_{ij} regression parameters associated with z_{ijp}, and $\lambda_{0,ij}(\ )$ is the baseline hazard function.

We shall assume that right censoring can occur only at stopping times. Let y be the censoring time in a state k, and z_y be the covariates at this time y. Defining d as the indicator of censoring ($d = 1$ if a transition is observed, and $d = 0$ otherwise), it is well known (Perez-Ocon and Ruiz-Castro [12]) that the individual contribution to the likelihood is given by:

$$\mathcal{L} = \prod_{ij} \{P_{ij}^{n_{ij}} \lambda_{ij}(x_{ij1}, z_{ij1}) \cdots \lambda_{ij}(x_{ijn_{ij}}, z_{ijn_{ij}}) \exp(-[\Lambda_{ij}(x_{ij1}, z_{ij1})$$

$$+ \cdots + \Lambda_{ij}(x_{ijn_{ij}}, z_{ijn_{ij}})])\} \times \left\{ \sum_{j \ne k} P_{kj} \exp(-\Lambda_{kj}(y, z_y)) \right\}^{1-d} \tag{5}$$

in which $\Lambda_{ij}(x) = \int_0^x \lambda_{ij}(u)du$ is the cumulative hazard function.

2.2. Incorporation of frailties

We define ω_{ij}, as the unobservable random effects on a subject, which are assumed to have independent distributions, specific to the transition $i \to j$, with a mean of 1 (in order to get unique identification of parameters). Given this mixing variable, the transition intensity is therefore $\omega_{ij}\lambda_{ij}(x)$. The frailty term is assumed to describe the dependence of the transition hazard function on an individual. To simplify the notations, we did not take into account the subject index. From (5), the individual and conditional likelihood can be written as:

$$\mathcal{L} = E\Bigg[\prod_{ij}\{P_{ij}^{n_{ij}}\omega_{ij}^{n_{ij}}\lambda_{ij}(x_{ij1}, z_{ij1})\cdots\lambda_{ij}(x_{ijn_{ij}}, z_{ijn_{ij}})\exp(-\omega_{ij}v_{ij})\}$$

$$\times\left\{\sum_{j\neq k}P_{kj}\exp(-\omega_{kj}u_{kj})\right\}^{1-d}\Bigg], \tag{6}$$

where $v_{ij} = \Lambda_{ij}(x_{ij1}, z_{ij1}) + \cdots + \Lambda_{ij}(x_{ijn_{ij}}, z_{ijn_{ij}})$, the total cumulative hazard function and $u_{kj} = \Lambda_{kj}(y, z_y)$. A suitable reformulation can be given by introducing the Laplace transform of ω_{ij}, defined by $L(a) = E[\exp(-a\omega_{ij})]$. Since the rth derivative $L^{(r)}(a)$ equals $(-1)^r E[\omega_{ij}^r \exp(-a\omega_{ij})]$ and the independence of frailties between different transitions, we can rewrite (6). Taking the logarithm, the general formulation of the individual contribution to the likelihood is given by:

$$\ln\mathcal{L} = \sum_{ij}\left\{n_{ij}\ln(P_{ij}) + \sum_{p=1}^{n_{ij}}\ln(\lambda_{ij}(x_{ijp}, z_{ijp})) + \ln((-1)^{n_{ij}}L^{(n_{ij})}(v_{ij}))\right\}$$

$$+ (1-d)\ln\left(\sum_{j\neq k}P_{kj}L(u_{kj})\right). \tag{7}$$

2.3. Model parameterization

The distribution of the mixing variable should be chosen in such a way that it has an explicit Laplace transform. Hougaard [9] describes distributions used in frailty models for multivariate survival data: stable, gamma, power variance function, etc. However, only the gamma distribution offers a simple rth derivative of the transform of Laplace, for r large. For the other distributions, the derivatives use some recursive polynomials, already associated with problems of estimation in a simple multivariate models of survival. Moreover, the gamma frailty distribution has proven to be advantageous in many context based on statistical and practical ground, for example, Aalen [1] or Clayton [5].

For each transition, we assume that the ω_{ij}'s are independent and identically Gamma distributed, with density

$$g(\omega_{ij}) = (\omega_{ij}^{(\delta_{ij}^{-1}-1)} \exp(-\omega_{ij}\delta_{ij}^{-1}))/(\Gamma(\delta_{ij}^{-1})\delta_{ij}^{\delta_{ij}^{-1}}), \quad \forall \delta_{ij} \geq 0. \tag{8}$$

The mean value of ω_{ij} is 1 and the variance is δ_{ij}. The high value of δ_{ij} reflects greater heterogeneity between subjects for the transition $i \to j$. The Laplace transform, $L(a)$ is therefore $(1+\delta_{ij}a)^{-\delta_{ij}^{-1}}$. It may be of interest to test the null hypothesis $\delta_{ij} = 0$, corresponding to the independence between all the observations. It lies on the boundary of the natural parameter space $\delta_{ij} \geq 0$. Aalen and Husebye [2] demonstrated how the family of distributions may be extended to values of δ_{ij} somewhat below zero, making the null hypothesis an interior point. We can thus use the Likelihood Ratio Statistic (LRS) for testing this hypothesis.

Let the hazard function of waiting times follow a generalized Weibull distribution:

$$\lambda_{0,ij}(x) = \frac{1}{\theta_{ij}}\left(1+\left(\frac{x}{\sigma_{ij}}\right)^{\nu_{ij}}\right)^{\frac{1}{\theta_{ij}}-1} \frac{\nu_{ij}}{\sigma_{ij}}\left(\frac{x}{\sigma_{ij}}\right)^{\nu_{ij}-1}, \quad \forall \sigma_{ij} > 0, \nu_{ij} > 0, \theta_{ij} > 0. \tag{9}$$

This class of distribution has very nice properties (Bagdonavicius and Nikulin [4]). In dependence of parameter values, the hazard function can be constant, monotone (increasing or decreasing), and U or inverse U shape. These distributions have the advantage of being nested, the Likelihood Ratio Statistic (LRS) can thus be used to simplify this distribution. For example, if we fix θ_{ij} at 1, then we find the Weibull formulation. Therefore, this method generalizes the semi-Markov model proposed by Perez-Ocon and Ruiz-Castro [12]. For all these reasons, the generalized Weibull distribution appears to be suitable for modeling the baseline hazard functions of waiting times.

From the equation (7) and following the parameterizations implied by (8) and (9), the final individual contribution to the marginal loglikelihood of this semi-Markov frailty model can be calculated. It is derived in the Appendix.

$$\begin{aligned}
\ln \mathcal{L} = \sum_{ij} \Bigg\{ & n_{ij}[\ln(P_{ij}) - \ln(\theta_{ij}) + \ln(\nu_{ij}) - \nu_{ij}\ln(\sigma_{ij})] \\
& + \sum_{p=1}^{n_{ij}} \Bigg(\left(\frac{1}{\theta_{ij}} - 1\right) \times \ln\left(1 + \left(\frac{x_{ijp}}{\sigma_{ij}}\right)^{\nu_{ij}}\right) \\
& + (\nu_{ij} - 1)\ln(x_{ijp}) + \beta_{ij}^T z_{ijp} + \ln(1 + (p-1)\delta_{ij}) \Bigg) \\
& - (1/\delta_{ij} + n_{ij})\ln\left(1 + \delta_{ij} \sum_{p=1}^{n_{ij}} \left(\left(\left(1 + \left(\frac{x_{ijp}}{\sigma_{ij}}\right)^{\nu_{ij}}\right)^{(1/\theta_{ij})} - 1 \right) \exp(\beta_{ij}^T z_{ijp}) \right) \right) \Bigg\} \\
& + (1-d)\ln\left(\sum_{j \neq k} P_{kj} \left(1 + \delta_{kj} \left(\left(1 + \left(\frac{x_y}{\sigma_{kj}} \right)^{\nu_{kj}} \right)^{1/\theta_{kj}} - 1 \right) \exp(\beta_{kj}^T z_y) \right)^{-1/\delta_{kj}} \right).
\end{aligned} \tag{10}$$

3. Analysis of HIV Data

The development of such a type of model was motivated by the study of the evolution of patients infected with HIV, using a cohort followed up at Nice University Hospital (France). Indeed, a patient can repeat several times the same transition. We mean that the times of a same transition for a particular individual are independent random variables with identical distributions.

Clinicians considered two markers for qualifying the gravity of the disease: viral load (VL) and concentration of CD4 lymphocytes (CD4). They defined the multi-state process as characterized in Figure 1. The possible transitions were selected according to their frequencies in the sample. The purpose was to analyze the progression of HIV disease using this four-state semi-Markov model, according to the eight following factors: gender (women = 1; men = 0), age (1 = over 40 years old; 0 = otherwise), hepatitis B coinfection (1 = yes; 0 = no), hepatitis C coinfection (1 = yes; 0 = no) and the means of contamination which could be either heterosexual (1 = yes; 0 = no), homosexual (1 = yes; 0 = no), by drug addiction (1 = yes; 0 = no), or some other accidental way (1 = yes; 0 = no).

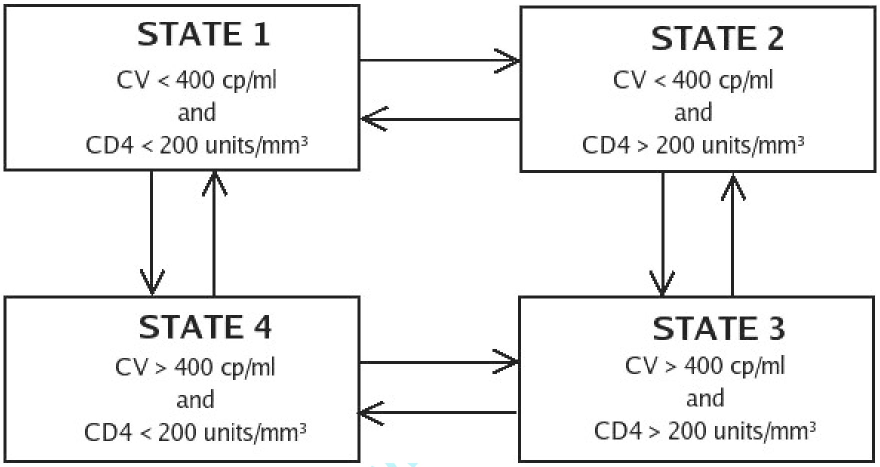

Figure 1. Four-state semi-Markov model.

We also limited our analysis to observations collected since 1996 and to individuals over 18 years old. The database is thus constituted of 163 HIV-infected patients, followed up in the Hospital of Nice, France (NADIS database). This represents 1478 observations, which are summarized in Table 1. In agreement with the clinicians, the process is assuming sufficiently stable and the visits are rather regular (see Figure 2), in order to consider that most of the transitions are identified, without dealing with interval-censored times. It is also justified by the Table 1. Indeed, the transitions are observed only between consecutive states of gravity.

Table 1. Frequency of the transitions observed

Transition	Number	Percentage	Median[1]	Mean[1]
1 → *censoring*	6	0.41%	0.79	0.78
1 → 2	129	8.73%	0.26	0.38
1 → 4	84	5.68%	0.26	0.38
2 → 1	84	5.68%	0.26	0.51
2 → *censoring*	69	4.67%	0.67	0.96
2 → 3	346	23.41%	0.37	0.53
3 → 2	374	25.30%	0.34	0.51
3 → *censoring*	23	1.56%	0.44	0.61
3 → 4	112	7.58%	0.27	0.48
4 → 1	117	7.92%	0.44	0.65
4 → 3	126	8.52%	0.25	0.42
4 → *censoring*	8	0.54%	0.26	0.33

[1] Waiting times in years

We first select covariates with a univariate strategy ($p \leq 0{,}20$). By defining a factor as a covariate for a specific transition, 10 factors are selected for the multivariate model. Finally, by a descending procedure ($p \leq 0{,}05$), 5 factors seem significant. They are given in Table 2. Hepatitis B seems to be important. It is a significant predictor for the transitions $1 \to 4$, $2 \to 1$ and $2 \to 3$. For example, patients infected by hepatitis B, are 1.93 times likelier to leave State 2, given that State 1 follows. Likewise, patients not infected by homosexual relation, seems to

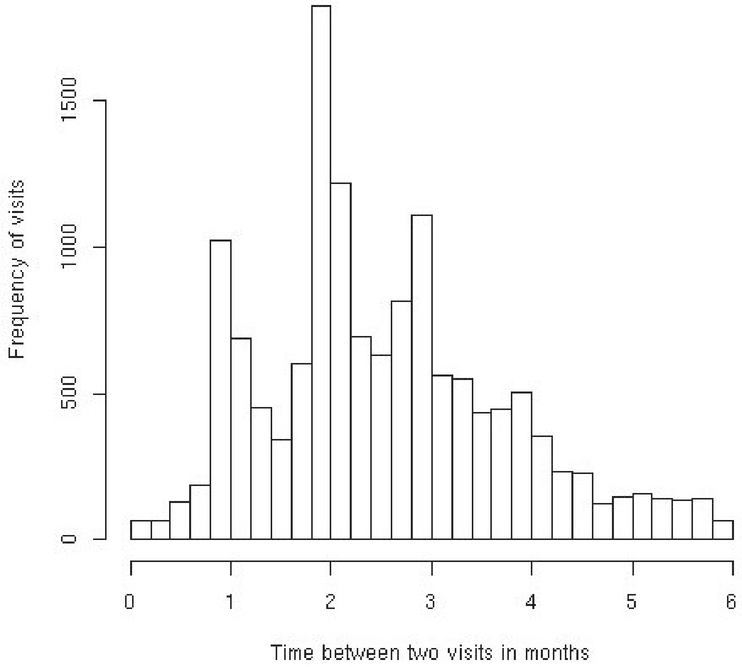

Figure 2. Distribution of time between two consecutive visits.

accelerate the transition 2 → 1. Lastly, hepatitis C seems to be a risk factor of the transition 3 → 2. In order to interpret the effects of factor as relative risk, we are restricted to conditioning on the following state (Equation 2).

Table 2. Regression parameters β_{ij} for the final multivariate model

Covariate	Transition	Estimate	Stand. Deviation	Relative Risk[1]	p-value
Hepatitis B	1 → 4	–0.67	0.30	0.51	0.0243
Hepatitis B	2 → 1	0.66	0.35	1.93	0.0581
Homosexuality	2 → 1	–1.35	0.65	0.26	0.0384
Hepatitis B	2 → 3	–0.52	0.18	0.59	0.0044
Hepatitis C	3 → 2	0.31	0.13	1.36	0.0129

[1] Relative Risk is deduced from risk proportionality assumption: $RR = \exp(\beta)$

Figure 3 presents certain hazard functions of the semi-Markov process (Equation 4). Whereas the ratio between the hazard functions of the waiting times (λ_{ij}) is constant, this diagram shows us that the interpretation of the ratio between the hazard functions of the semi-Markov process (α_{ij}) is not so simple. It depends on waiting times. However, we can identify the effect of the variable without conditioning on the following state. Hepatitis B seems to influence the transition $1 \to 2,$ even if this covariate does not have a direct effect on this last transition.

The maximum likelihood estimates of the other parameters, with the corresponding standard errors, are given in Table 3. It is interesting to note that the random effect terms, $\delta_{ij},$ are close to zero, confirming the lack of significant variation between individuals. A logarithmic transformation was even necessary to remove the skewness for this parameter. Using the LRS to test the null hypothesis according to which δ_{ij} equal 0, $\forall i \neq j,$ we conclude that the incorporation of frailties is not informative ($\chi^2 = 1.89,$ $ddl = 8$ and $p = 0.9841$).

Table 3. Parameters of waiting time distribution for the final multivariate model

	ν_{ij}		σ_{ij}		θ_{ij}		δ_{ij}	
Transition	Estim	SD	Estim	SD	Estim	SD	Estim	exp(Estim)
$1 \to 2$	0.12	0.02	3.04	0.54	4.93	1.13	–8.18	0.0003
$1 \to 4$	0.13	0.02	3.29	0.76	4.52	1.38	–7.13	0.0008
$2 \to 1$	0.13	0.02	5.08	1.26	10.82	3.75	–1.11	0.3296
$2 \to 3$	0.18	0.03	2.64	0.34	4.24	0.94	–2.63	0.0721
$3 \to 2$	0.15	0.02	2.62	0.34	4.64	0.85	–9.14	0.0001
$3 \to 4$	0.11	0.02	2.30	0.44	4.50	1.17	–7.22	0.0007
$4 \to 1$	0.13	0.02	3.58	1.07	7.44	2.66	–8.29	0.0003
$4 \to 3$	0.11	0.02	2.81	0.59	5.15	1.39	–7.94	0.0004

Finally, still in Table 3, we note that the estimates of θ_{ij} and their standard deviation justify our choice of modeling strategy based on generalized Weibull distribution. For example, a 95% confidence interval for the parameter θ_{12} is [2.72; 7.14], which does not include value 1. This distribution is more informative than the Weibull one. This also justified by the Figure 3, where the hazard functions are not monotone.

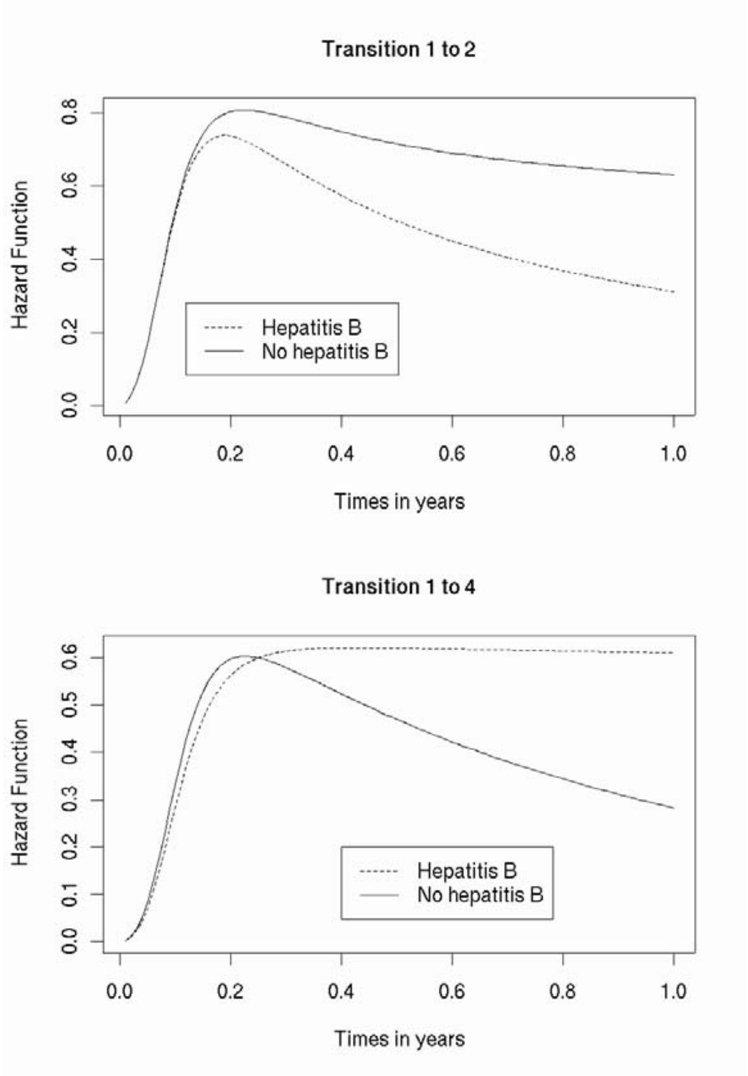

Figure 3. Hazard function of the semi-Markov process, α_{12} and α_{14}, according to Hepatitis B.

4. Discussion

In this paper, we introduced a general semi-Markov model with random effects and generalized Weibull distribution of waiting times. Random effects were assumed to be Gamma distributed. Estimations were obtained by maximizing the marginal likelihood, which was made possible by using Laplace transform. The advantage was to base all our modeling strategy on the LRS and on the traditional maximum likelihood estimation theory.

As illustrated by the example on HIV, the model provides a useful method for analyzing clustered multi-state processes. As we supposed, a subject may be assumed to be a cluster, because a transition $i \to j$ can occur for any given subject. Unobserved covariates could then constitute variability between subjects. Even if the frailties do not significantly increase the likelihood in our application, that could be the case for other analysis. Moreover, our strategy has at least the advantage of testing the independence of the transition times. This assumption is too strong to be formulated a priori.

The general formulation of the model and the corresponding likelihood (10) may be used for modeling different structures of cluster. For example, to take into account center, geographical or family effects in a multi-state process, random effects should constitute an adequate modeling solution.

The model is also original concerning the choices of parameterization. Factor effect is specific to each transition and baseline hazard functions are chosen to be generalized Weibull distributed. The application shows the parsimony obtained for the analysis of the dynamics of the HIV.

All these results are in agreement with the medical literature, except for the effects of hepatitis. It would be relevant to propose a distinct model for CV and CD4, in order to have a more precise analysis of this covariate.

However, the model has a few limits. First, the probability that an interval is selected is proportional to the length of this interval (length-biased sampling). As it is discussed in Section 3, a patient HIV-infected is rather stable and carries out frequent biological measurements. For the majority of the visits, the results of biological measurements show that the patient remains in the same state. Therefore, this problem should be limited on our application. Secondly, model checking techniques need to be developed to investigate the fit of the model. This is a very important issue for these types of highly parametric models.

Appendix: Likelihood of the Semi-Markov Model

The rth derivated from the Laplace transform for Gamma distribution is

$$L^{(r)}(a) = -\delta_{ij}^{r-1}(1+\delta_{ij}a)^{(-1/\delta_{ij}-r)}\prod_{p=1}^{r-1}(-1/\delta_{ij}-p).$$

The equation (7) can therefore be written as

$$\begin{aligned}
\ln \mathcal{L} = {} & \sum_{ij}\Bigg\{n_{ij}\ln(P_{ij}) + \sum_{p=1}^{n_{ij}}\ln(\lambda_{ij}(x_{ijp}, z_{ijp})) \\
& \quad + \ln\Bigg((-\delta_{ij})^{n_{ij}-1}(1+\delta_{ij}v_{ij})^{(-1/\delta_{ij}-n_{ij})}\prod_{p=1}^{n_{ij}-1}(-1/\delta_{ij}-p)\Bigg)\Bigg\} \\
& + (1-d)\ln\Bigg(\sum_{j\neq k}P_{kj}(1+\delta_{kj}u_{kj})^{-1/\delta_{kj}}\Bigg) \\
= {} & \sum_{ij}\Bigg\{n_{ij}\ln(P_{ij}) + \sum_{p=1}^{n_{ij}}\ln(\lambda_{ij}(x_{ijp}, z_{ijp})) + (-1/\delta_{ij}-n_{ij})\ln(1+\delta_{ij}v_{ij}) \\
& \quad + \sum_{p=1}^{n_{ij}-1}(\ln(-\delta_{ij}) + \ln(-1/\delta_{ij}-p))\Bigg\} \\
& + (1-d)\ln\Bigg(\sum_{j\neq k}P_{kj}(1+\delta_{kj}u_{kj})^{-1/\delta_{kj}}\Bigg).
\end{aligned}$$

However, we have

$$\sum_{p=1}^{n_{ij}-1}(\ln(-\delta_{ij})+\ln(-1/\delta_{ij}-p))=\sum_{p=1}^{n_{ij}-1}(\ln(1+\delta_{ij}p))$$

$$=\sum_{p=1}^{n_{ij}}(\ln(1+(p-1)\delta_{ij})).$$

So we obtain

$$\ln\mathcal{L}=\sum_{ij}\left\{n_{ij}\ln(P_{ij})+\sum_{p=1}^{n_{ij}}(\ln(\lambda_{ij}(x_{ijp},\,z_{ijp}))+\ln(1+(p-1)\delta_{ij}))\right.$$

$$\left.-\,(1/\delta_{ij}+n_{ij})\ln(1+\delta_{ij}v_{ij})\right\}+(1-d)\ln\left(\sum_{j\neq k}P_{kj}(1+\delta_{kj}u_{kj})^{-1/\delta_{kj}}\right).$$

In order to conclude the parameterization, the hazard function of waiting times is explicitly introduced. This formulation is given by (9). By integrating this function, the cumulative hazard, $\Lambda_{0,ij}(x)$, is equal to $\Lambda_{0,ij}(x)=\left(1+\left(\frac{x}{\sigma_{ij}}\right)^{\nu_{ij}}\right)^{1/\theta_{ij}}-1.$ Finally, these developments lead to the following individual contribution to the loglikelihood.

$$\ln\mathcal{L}=\sum_{ij}\left\{n_{ij}[\ln(P_{ij})-\ln(\theta_{ij})+\ln(\nu_{ij})-\nu_{ij}\ln(\sigma_{ij})]\right.$$

$$+\sum_{p=1}^{n_{ij}}\left(\left(\frac{1}{\theta_{ij}}-1\right)\times\ln\left(1+\left(\frac{x_{ijp}}{\sigma_{ij}}\right)^{\nu_{ij}}\right)\right.$$

$$\left.+\,(\nu_{ij}-1)\ln(x_{ijp})+\beta_{ij}^{T}z_{ijp}+\ln(1+(p-1)\delta_{ij})\right)$$

$$\left.-\,(1/\delta_{ij}+n_{ij})\ln\left(1+\delta_{ij}\sum_{p=1}^{n_{ij}}\left(\left(1+\left(\frac{x_{ijp}}{\sigma_{ij}}\right)^{\nu_{ij}}\right)^{(1/\theta_{ij})}-1\right)\exp(\beta_{ij}^{T}z_{ijp})\right)\right\}$$

$$+(1-d)d\ln\left(\sum_{j\neq k}P_{kj}\left(1+\delta_{kj}\left(\left(1+\left(\frac{x_{y}}{\sigma_{kj}}\right)^{\nu_{kj}}\right)^{1/\theta_{kj}}-1\right)\exp(\beta_{kj}^{T}z_{y})\right)^{-1/\delta_{kj}}\right).$$

References

[1] O. Aalen, Effects of frailty in survival analysis, Statistical Models for Medical Research 3 (1994), 227-243.

[2] O. Aalen and E. Husebye, Statistical analysis of repeated events forming renewal processes, Statistics in Medicine 10 (1991), 1227-1240.

[3] A. Alioum, V. Leroy, D. Commenges, F. Dabis and R. Salamon, Effect of gender, age, transmission category, and antiretroviral therapy on the progression of human immunodeficiency virus infection using multistate Markov models, Groupe d'epidemiologie clinique du SIDA en Aquitaine, Epidemiology 9 (1998), 605-612.

[4] V. Bagdonavicius and M. Nikulin, Accelerated Life Models, Chapman & Hall/CRC, London, 2002.

[5] D. Clayton, Some approaches to the analysis of recurrent event data, Statistical Models for Medical Research 3 (1994), 244-262.

[6] C. Combescure, P. Chanez, P. Saint-Pierre, J. Daures, H. Proudhon and P. Godard, Assessment of variations in control of asthma over time, European Respiratory J. 22 (2003), 298-304.

[7] S. Escolano, J. Golmard, A. Korinek and A. Mallet, A multi-state model for evolution of intensive care unit patients: prediction of nosocomial infections and deaths, Statistics in Medicine 19 (2000), 3465-3482.

[8] Y. Foucher, E. Mathieu, P. Saint-Pierre, J. Durand and J. Daures, A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease, Biom. J. 47 (2005), 825-833.

[9] P. Hougaard, Analysis of Multivariate Survival Data, Springer, New York, 2000.

[10] C. Jackson, L. Sharples, S. Thompson, S. Duffy and E. Couto, Multistate Markov models for disease progression with classification error, J. Roy. Statist. Soc. Ser. D 52 (2003), 193-209.

[11] J. Klein, Semiparametric estimation of random effects using the Cox model based on the EM algorithm, Biometrics 48 (1992), 795-806.

[12] R. Perez-Ocon and J. Ruiz-Castro, Semi-Markov Models and Applications, Kluwer Academic Publishers, Dordrecht, 1999.

[13] S. Ripatti, M. Gatz, N. Pedersen and J. Palmgren, Three-state frailty model for age at onset of Dementia and death in Swedish twins, Genetic Epidemiology 24 (2003), 139-149.

[14] P. Saint-Pierre, C. Combescure, J. Daures and P. Godard, The analysis of asthma control under a Markov assumption with use of covariates, Statistics in Medicine 22 (2003), 3755-3770.

[15] C. Yau and A. Huzurbazar, Analysis of censored and incomplete survival data using flowgraph models, Statistics in Medicine 21 (2002), 3727-3743.

■