AJAS / SPAS

African Journal of Applied Statistics
Vol. 7 (1), 2020, pages 915 - 932.
DOI: http://dx.doi.org/10.16929/ajas/2020.915.247

ISSN 2316-0861

Using Markov assumption with covariates to assess the Plasmodium falciparum malaria serological markers evolution

Niass Oumy [1,2,*] **, Diongue Abdou Kâ** [1]**, Philippe Saint-Pierre** [3] **and Touré Aissatou** [2]

[1] LERSTAD : Laboratoire d'Etudes et de Recherches en Statistiques et Développement, UFR Sciences Appliquées et Technologie, UGB, BP 234, Saint-Louis, (SENEGAL)
[2] Unité d'Immunologie, Institut Pasteur de Dakar, Dakar (SENEGAL)
[3] Institut Mathématique de Toulouse (FRANCE)

Received on December 05, 2019; Accepted on N0vember 2, 2020

Abstract. In this study, we develop three Markov models which are continuous time-homogeneous Model, time piecewise constant intensities Markov model and semi-Markov model with Weibull distribution as the waiting time distribution to evaluate malaria serology evolution. We consider two-state model describing antibody reactivity defined by immunologists. We discuss in detail the application of these models to identify relationships between malaria control program and serological measurements of malaria transmission.

Key words: longitudinal serological malaria data; continuous time Markov processes; Plasmodium falciparum; piecewise constant intensities; maximum likelihood estimation; semi-Markov process
AMS 2010 Mathematics Subject Classification Objects: 62P10; 60Jxx;

*Corresponding author: Oumy Niass (oumyniass@gmail.com)
Diongue Abdou Kâ : abdou.diongue@ugb.edu.sn
Philippe Saint-Pierre: philippe.saint-pierre@math.univ-toulouse.fr
Aissatou Touré: atoure@pasteur.sn, atoure85@yahoo.fr

Résumé (Abstract in French) Dans cette étude, nous développons trois modèles markoviens qui sont des modèles continus homogènes dans le temps, des modèles de Markov à intensité constante par morceaux de temps et des modèles semi-Markov avec distribution de Weibull comme distribution du temps d'attente pour évaluer l'évolution sérologique du paludisme. Nous considérons un modèle à deux états décrivant les réactivités des anticorps définies par les immunologistes. Nous discutons en détail de l'application de ces modèles pour identifier les relations entre le programme de lutte contre le paludisme et les mesures sérologiques de la transmission du paludisme.
The authors.

Oumy Niass, Ph.D., is lecturer at Université Virtuelle du Sénégal, Dakar, SEMEGAL.

Abdou Kâ Diongue, Ph.D., is full professor at Department de Mathematiques Appliquées, UFR Sciences Appliquées et Technologie (SENEGAL).

Philippe Saint-Pierre, Ph.D., is associated professor at: Institut Mathématique de Toulouse, (FRANCE)

Aissatou Touré, Ph.D., is a Doctor of Immunology, head of the Immunology Laboratory at: Institut Pasteur de Dakar, Dakar (SENEGAL).

1. Introduction

In epidemiology, cohort studies are often seen. In these studies, some subjects are irregularly monitored and the information collected is in the form of states of health. Particularly in sero-immunology where reactivity of antibodies against antigen appears in state form for some discrete time. It is therefore important to assess the passage of various immunological markers for the different immune reactivity state for a better understanding of the body's protective mechanisms. The multi-states models provide a complete representation of the evolution of biological phenomena over time (Longini *et al.* (1989); Alioum *et al.* (1998)). Reasons for constructing multi-state models are to provide a comprehensive view of the disease process and to allow estimation of proportions of individuals who will be in the various states at some time in the future.
In many biological application Markov assumption was applied successfully to extend classical survival model to multi-states analysis and taking into account that exact times between diseases states are generally not observed for many patients. Indeed Markov process has been used for evaluation of diseases progression in chronic healthy. For instance, to model different states of cancer (Kay (1986); Hsieh *et al.* (2002)), states of HIV infection (Gentleman *et al.* (1994); Longini *et al.* (1989); Alioum *et al.* (1998); Jackson *et al.* (2003)) or diabetes (Marshall and Jones (1995)). Markov models have also been applied to asthma controls (Saint-Pierre *et al.* (2003); Combescure *et al.* (2003)). In malaria context, Markov assumption was

developed to estimate the incidence and recovery rates of *Plasmodium falciparum* parasitaemia (Bekessy *et al.* (1976)). Recently, studies have shown that predicted intensities rate of *Plasmodium falciparum* serology in area with low transmission endemicity may be used to assess malaria transmission intensities (Drakeley *et al.* (2005); Corran *et al.* (2007); Stewart *et al.* (2009); Supargiyono *et al.* (2009)).

However, in most application, time-homogeneous assumption implying that transition intensities are time-independent are usually used to model evolution in chronic disease. These assumptions are so strong in many clinical problems. Since transition intensities seem unlikely to be constant over long periods for most disease processes and since Markov assumptions may be violated, it is important to use another model and to have assessment tools. Time-piecewise constant intensities Markov and semi-Markov models can be used to extend homogeneous Markov models with continuous time and discrete finite states. Time-piecewise constant intensities which preserve constantness intensities hypothesis had successful application in healthy environment (Alioum and Commenges (2001); Saint-Pierre *et al.* (2003)). Several authors have discussed its proprieties in analysis of health data (Alioum and Commenges (2001); Lindsay and Ryan (1993)).

In this study, our goal is to apply homogeneous and time-period homogeneous Markov process to study evolution of antibodies responses against crude extract schizonts 07/O3 of *Plasmodium falciparum* and deal with how malaria serological markers can be assessed the transmission level. The consideration of time-piecewise constant intensities is allowing to take into account not only the differentiate immunogeneicity of subject according to the age-dependence (time) of malaria antibody responses (Diop *et al.* (2014)).

Regarding to the literature on serological evolution for assessing malaria transmission, some previous studies have focused on immunological malaria markers evolution using a Markov background (Drakeley *et al.* (2005); Corran *et al.* (2007); Stewart *et al.* (2009); Supargiyono *et al.* (2009); Rosas-Aguirre *et al.* (2013)). All these studies used a continuous-time Markov model without covariates to model serological malaria evolution. In this study, we use both continuous-time Markov model without and with covariates. In this paper, we proceed as follows. In section 2 we describe theories of time-homogeneous Markov and time-piecewise constant intensity Markov process without covariates and its extension to models with covariates. The statistical methods such as the likelihood estimation, model assessment and goodness of fit are also given. In Section 3, we propose an illustration using *P.falciparum* malaria serological data and in the last section (section 4), results are discussed.

2. Methods

In this section, we present some properties of continuous-time Markov process that we are needed to define our interest models. For the theoretical development of Markov processes, we refer the reader to Cox and Miller (1977).

2.1. Time-homogeneous Markov models

Consider a model consisting of k states belonging in the state space $E = \{1, 2, ..., k\}$, with a subject being unequivocally in some state at time t. Let $X(t)$ denote the state occupied at time t by a given subject. Assume that subjects independently move among the k states according to a continuous-time Markov model. The transition between states are given by the transition probability defined by :

$$p_{ij}(s,t) = P[X(t) = j | X(s) = i], \quad \text{for } 0 \leq s \leq t \quad \text{and} \quad i, j \in E \tag{1}$$

A Markov process can be completely specified by the transition intensities:

$$\lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{p_{ij}(t, t+\Delta t)}{\Delta t}, \quad j \neq i \tag{2}$$

$$\lambda_{ii}(t) = -\sum_{i \neq j} \lambda_{ij}(t), \quad i = 1, 2, \cdots, k \tag{3}$$

Let $P(s,t)$ the $k \times k$ transition probability matrix with entries $p_{ij}(s,t)$. This process is telling time-homogeneous Makov if these transition intensities $\lambda_{ij}(t)$ are constant ($\lambda_{ij}(t) = \lambda_{ij}$, $\forall t$). In this case transition probabilities are stationary and can be calculated as

$$P(s, s+t) = P(0,t) = P(t) = \exp(\Lambda t) \tag{4}$$

where Λ is the transition intensities matrix with entries λ_{ij} for $(i \neq j)$ and $\lambda_{ii} = -\sum_{i \neq j} \lambda_{ij}$. Note that $\lambda_{ij} \geq 0$ and $\sum_{j \neq \infty} \lambda_{ij} = 0$. Cox and Miller (1977) suggested a simple procedure for computing the transition probability matrix $P(t) = \exp(\Lambda t)$ in terms of eigenvalues and eigenvectors of the transition intensities matrix Λ. Clearly, $P(t)$ is calculated as

$$P(t) = A \; diag(e^{d_1 t}, e^{d_2 t}, \cdots, e^{d_k t}) \; A^{-1} \tag{5}$$

where $d_1, d_2, \cdots, d_k$ are the eigenvalues of Λ and A is the matrix whose jth column is the eigenvector associated with d_j $(j = 1, 2, \cdots, k)$. Note also that the transition probabilities satisfy the Chapman-Kolmogorov equation defined by

$$P_{ij}(s,t) = \sum_{l \in E} P_{il}(s,u) P_{lj}(u,t), \quad s \leq u \leq t. \tag{6}$$

In many healthy studies, for each subject some covariates are measured and it is interesting to assess the relationship between these covariates and the transition intensities λ_{ij} in the Markov model. The model can be extended in a straightforward way to allow for regression modelling of Λ. If we assume that the proportional intensities regression models hold, then the transition intensities can be expressed as

$$\lambda_{ij}(Z) = \lambda_{ij0} \exp(\beta_{ij}^{'} Z), \quad j \neq i \tag{7}$$

where Z is a p-dimensional vector of covariates, β_{ij} a vector of p regression coefficients relating the influence of Z to the instantaneous rate of transition from state i to state j and λ_{ij0} represents the baseline intensity relating to the transition from state i to state j. These regression coefficients can be interpreted similarly to those in the proportional hazards regression model (Cox (1972)). Note that the covariates can be time-dependent. In this context it is interesting to assume that between the two consecutive times, if the time-dependent covariate remains constant then the transition intensities are defined as

$$\lambda_{ij}(t|Z(t)) = \lambda_{ij0} \exp\{\beta_{ij}^{'} Z(t)\}, \quad j \neq i \tag{8}$$

2.2. Time-piecewise constant intensities Markov

A time-homogeneous Markov model with time-dependent covariates allows one to deal with non-homogeneous Markov model and particularly with a piecewise constant intensities model. Let us divide the time axis into intervals $[\tau_{l-1}; \tau_l)$, where $l = 1, 2, \cdots, r+1$ and $\tau_{r+1} = +\infty$ and assume constant intensity for each type of transition in each interval.

Consider a vector $z^*(t) = (z_0^*(t), z_1^*(t), \cdots, z_r^*(t))$ of artificial time-dependent covariates defined by

$$\begin{array}{rcl} \hat{z_0}(t) & = & 0 \qquad\qquad\qquad\quad \forall\, t \\ z_l^*(t) & = & \begin{cases} 0 & \text{if} \quad \tau_{l-1} \leq t < \tau_l \quad \text{for} \quad l = 1, 2, \cdots, r \\ 1 & \text{if} \quad t \geq \tau_l \end{cases} \end{array} \tag{9}$$

and the model with transition intensities

$$\lambda_{ij}(t|z^*(t)) = \lambda_{ij0} \exp\{(\beta_{ij}^*)^{'} z^*(t)\}, \quad i \neq j. \tag{10}$$

In this model, the intensities vary with time t as step-functions defined on the prespecified intervals $[\tau_{l-1}, \tau_l)$, $l = 1, 2, \cdots, r+1$; time is measured from the beginning of the process under study. The parameters of this model are the baseline intensities q_{ij0}, which represent the transition intensities in the interval $[\tau_0, \tau_1)$ and the vector of regression coefficients β_{ij}^* associated with the artificial time-dependent covariates. Remark that this model leads to a non-homogeneous Markov model in which the transition intensities are step-functions of time. Clearly, we have:

$$\lambda_{ij}(t|z^*(t)) = \begin{cases} \lambda_{ij_0} & \tau_0 \leq t < \tau_1, \\ \lambda_{ij_1} = \lambda_{ij_0} \exp\{\beta_{ij,1}^*\} & \tau_1 \leq t < \tau_2, \\ \quad \vdots & \\ \lambda_{ij_r} = \lambda_{ij_0} \exp\{\sum_{l=1}^{r} \beta_{ij,l}^*\} & t > \tau_r. \end{cases} \tag{11}$$

This model is a generalization of the homogeneous-Markov model in that when $r = 0$ we obtain an homogeneous model.

2.3. Semi-Markov model with Weibull distribution

Basically, the homogeneous Markov model assumes that evolution of the process doesn't dependent on the spent time in the state (no memory). The semi-Markov model consider the waiting time distribution in the modelling and this distribution can be estimated explicitly. Thus it can be extended the time homogeneous model. Indeed in the particular situation where the waiting time found to be an exponential distribution, the semi-Markov model is equivalent to the homogeneous model.

Let $E = \{1, 2, \cdots, k\}$ an finite state space and $(T, X) = \{(T_m, X_m), m \geq 0\}$ a random process in which $0 = T_0 < T_1 < \cdots < T_m$ (m correspond to the number of transition) is the consecutive entry time to the state $X_0, X_1, \cdots, X_m$, such as $X_{p+1} \neq X_p, \forall p \geq 0$ and X_p is not persistent. The process $(T, X) = \{(T_m, X_m), k \geq 0\}$ is called Semi-Markov if the sequence $X = \{X_m : m \geq\}$ is an homogeneous Markov chain which transition probability from state i to state j is defined by:

$$P_{ij} = P[X_{m+1} = j | X_m = i], \tag{12}$$

and the distribution of the waiting time $D_m = (T_{m+1} - T_m)$ satisfies the following condition:

$$P[D_m \leq d, X_{m+1} = j | X_0, T_0, X_1, T_1, \cdots, X_m, T_m] = P[D_m \leq d, X_{m+1} = j | X_m] \tag{13}$$

Note that the chain $X = x_m, m \geq 0$ deals only on the sequence of states and not on the chronological time. In addition the distribution of the waiting time depends only on the contiguous states, knowing the sequence of states X.
The density of the waiting time in the state i before jumping to the state j is given by:

$$f_{ij}(d; \theta_{ij}) = \lim_{h \to 0^+} \frac{P[d \leq D_m \leq d + h | X_{m+1} = j, X_m = i]}{h} \tag{14}$$

where θ_{ij} is the parameter vector of the waiting time density f_{ij}. Note that the distribution and the parameter vector can vary as function of transitions. In the survival analysis field, we often deduce from f_{ij} the corresponding distribution function, the survival and the hazard functions, respectively denoted by F_{ij}, S_{ij} and α_{ij}:

$$\alpha_{ij}(d) = \lim_{h \to 0^+} \frac{P[d < D_m \leq d + h | D_m \geq d, X_{m+1} = j, X_m = i]}{h} \tag{15}$$

We deduce the marginal density and the marginal survival functions from equations 12 and 15 and by using the Bayesian theorem:

$$f_{i.}(d) = \sum_{j \neq i} P_{ij} f_{ij}(d), \quad \text{because} \quad P_{ii} = 0 \tag{16}$$

$$S_{i.}(d) = \sum_{j \neq i} S_{ij}(d) P_{ij} \tag{17}$$

The semi-Markovian hazard function which corresponding to the probability of passing towards state j knowing that the process stays in state i during a duration d, is defined as:

$$\lambda_{ij}(d) = \lim_{h\to 0^+} \frac{P[d < D_m \leq d+h, X_{m+1} = j | D_m \geq d, X_m = i]}{h}$$

$$= \frac{P_{ij} f_{ij}(d)}{S_{i.}(d)} \quad \text{with} \begin{cases} i \neq j \\ S_{i.}(d) > 0 \\ \lambda_{ii}(d) = -\sum_{i\neq j} \lambda_{ij}(d) \end{cases} \tag{18}$$

We consider the Weibull distribution ($W(\sigma_{ij}, \nu_{ij})$) like distribution of the waiting time on our modelling. The hazard function of the waiting time is given by :

$$\alpha_{ij}(d) = \nu_{ij} \left(\frac{1}{\sigma_{ij}}\right)^{\nu_{ij}} d^{\nu_{ij}-1}, \ \forall d > 0, \ \sigma_{ij} > 0, \ \nu_{ij} > 0 \tag{19}$$

If $\nu_{ij} = 1$, we found the exponential distribution which is the distribution of the waiting time on the time-homogeneous model (without memory). Covariates are incorporated in the modelling by assuming the semi-proportional risk which id the ratio between hazard waiting functions and it's expression is defined as:

$$\alpha_{ij}(d, Z_{ij}) = \alpha_{ij,0}(d) \exp\left(\beta_{ij}^{'} Z_{ij}\right), \ \forall d > 0, \ \sigma_{ij} > 0, \ \nu_{ij} > 0 \tag{20}$$

Where $\beta_{ij} = (\beta_{ij,1}, \cdots, \beta_{ij,n_{ij}})$ is the vector of the n_{ij} regression parameters associated with Z_{ij}.

2.4. Parameter estimation Method

2.4.1. Parameter estimation of time-homogeneous Markov model

Suppose a sample constituted by n subjects, denoted by $h (h = 1, 2, \cdots, n)$.The h-th subject moves n_h times. The observed data for subject h consist of $t_0^h < t_1^h < \cdots < t_{n_h}^h$ which are the successive follow-up times, the different states occupied $x^h = (x_0^h, x_1^h, x_2^h, \cdots, x_{n_h}^h)'$ and the values of the covariates vectors $(z_0^h, z_2^h, \cdots, z_{n_h}^h)$, where $x_j^h = X(T_j^h)$ and $z_j^h = Z(T_j^h)$, for $j = 1, 2, \cdots, n_h$. It is supposed that z_{j-1}^h remains constant between the two consecutive times t_{j-1}^h and t_j^h . For fixed covariates, $z_j^h = z_0^h$ for all j. The full likelihood function is the product of all individual contributions that are the result of the product of the contribution from each observed transition. Conditionally on the fact that $X(t_0^h) = x_0^h$, the likelihood can therefore be written as

$$L = \prod_{h=1}^{n} \left[\prod_{j=1}^{n_h} P_{x_{j-1}^h, x_j^h}(t_j^h - t_{j-1}^h | z_{j-1}^h) \right]. \tag{21}$$

Parameters are the baseline rate $\lambda_{ij,0}$ and regression coefficients β_{ij}. Maximum likelihood estimates for these parameters can be obtained by maximizing the likelihood function given by equation (21).

Quasi-Newton algorithm can be used to find maximum likelihood estimates using only an analytical expression for the likelihood function and using finite differences to obtain numerical approximations of the derivatives (Kalbleisch and Lawless (1985)). Kalbleisch and Lawless (1985) proposed a complete discussion of the scoring procedure estimation methods. It is important to note that in a model with a single absorbing state, the likelihood function is slightly different. This case is developed in detail by Kay (1986).

2.4.2. Parameter estimation in time-piecewise constant intensities Markov models

For all $i, j \in E$, let $P_{ij}^{l}(t)$ denote the transition probability associated with all time intervals contained in $[\tau_{l-1}, \tau_l)$, $l = 1, 2, \cdots, r+1$; more precisely, $P_{ij}(s, s+t) = P_{ij}^{l}(t)$ if $\tau_{l-1} \leq s \leq s+t \leq \tau_l$. For any t, let I_t denote the time interval of the form $[\tau_{l-1}, \tau_l)$ which contains t. For notational convenience, we denote by i_1 and i_2 the states occupied by an individual at two consecutive follow-up times t_1 and t_2. Note that the transition intensities are constant in each time interval, but they are different for one interval to another. Then using the Chapman–Kolmogorov equation (6), the contribution of this observation to the modified likelihood can be written as, see Alioum and Commenges (2001) for more details.

$$\begin{aligned} P_{i_1 i_2}[t_1, t_2/z^*(t), t_1 < t < t_2] \;=\; & \textstyle\sum_{k_1 \in E} \sum_{k_2 \in E} \cdots \sum_{k_r \in E} \{ p_{i_1 k_1}^{(I_t)}[\tau_{t_{i_1}} - t_1/z^*(t_1)] \\ & \times\; p_{k_1 k_2}^{(I_{t+1})}[\tau_{I_{t_1+1}} - \tau_{I_{t_1}}/z^*(\tau_{I_{t_1}})] \;\times\; \cdots \\ & \times\; p_{k_v i_2}^{(I_{t_2})}[t_2 - \tau_{I_{t_2-1}}/z^*(\tau_{I_{t_2-1}})] \} \end{aligned} \tag{22}$$

where $v = I_{t_2} - I_{t_1}$. The full likelihood function is obtained as before, and maximum likelihood estimates are computed by maximizing the likelihood function using the scoring procedure Alioum and Commenges (2001).

2.5. Weibull semi-Markovian model

According to the assumptions of the previous section (2.4.1), we have n subjects. The subject h was observed n_h times. He moves $n_h - 1$ time into different states at times, $t_0^{(h)} < t_1^{(h)} < \cdots < t_{n_h-1}^{(h)}$. At these successive follow-up times, he occupies the state $x_0^{(h)}, x_1^{(h)}, \cdots, x_{n_h-1}^{(h)}$, with $x_p^{(h)} \neq x_{p+1}^{(h)},\ p \geq 1$. We are mainly interesting on the last follow-up time $t_{n_h}^{h}$ of the subject h. At this time, the subject h can be censored or move again. The contribution of the h-th subject on the likelihood function is:

$$L_{(h)} = \prod_{j=1}^{n_h} \left\{ P_{x_{j-1}^{(h)}, x_j^{(h)}} f_{x_{j-1}^{(h)}, x_j^{(h)}}(t_j^{(h)} - t_{j-1}^{(h)}, Z_{x_{j-1}^{(h)}, x_j^{(h)}}) \right\}^{\delta_h} \times \left\{ S_{x_{n_h-1}^{(h)}.}(t_{n_h}^{(h)} - t_{n_h-1}^{(h)}) \right\}^{1-\delta_h}$$

The likelihood is the product of all the subject's contributions:

$$L = \prod_{h=1}^{n} L_{(h)} \tag{23}$$

We use the quasi-Newton algorithm to optimize the likelihood function and we adopted the modelling strategy given by Foucher *et al.* (2005) based on three steps. Briefly, we consider three steps of modelling. In the first step called *Stratified modelling* in which one model for each modality of covariates was adjusted. This step allows to evaluate graphically the proportionality assumption of the waiting times hazard functions for each covariate and each transition. Thus we could identify whether a covariate seemed to affect a transition and whether the assumption of risk proportionality was respected by looking for the parallelism between the curves of hazard functions. After this first step, we calculated for each covariate a univariate model. Only covariates which p-value of the LRT was less than 0.05 are considered on the multivariate model. Finally a multivariate model including all covariates selected on the two previous steps was calculated.

2.6. Model assessment and goodness of fit

In many healthy modelling it is important to test some hypotheses about the model. In Markov model for example we can test the hypothesis of the form H_0 : $\lambda_{ij} = 0$ against the alternative hypothesis H_1: $\lambda_{ij} \neq 0$. This approach allows to test the existence of transition from state i to state j. There are several possibilities for constructing a test of H_0. Kay (1986) used an application of Wald's test, which does not require re-computation of the maximum likelihood estimate under H_0. But, when the parameters are near the boundary of the parameter space, the normal approximation is unreliable. Another possibility for constructing a test of H_0 against the general alternative is to use likelihood ratio test Gentleman *et al.* (1994).

It is of great interest to study the relationship between the different covariates and the disease evolution and to demonstrate the influence of covariates on the disease evolution. In this case a Markov model with covariates is used and it is interesting to test hypothesis of the form $\beta_{ij,k} = 0$, which represents the fact that there is no relationship between the transition from state i to state j and the kth covariate value. A likelihood ratio test was used to test if regression coefficients are statistically different to zero. The statistic has an approximately χ^2_1 distribution if H_0 is true. As before, if H_0 is true, it is possible to work with a simpler sub-model where there is no relation between the transition from state i to state j and the kth covariate value ($\beta_{ij,k}$ is taken to be equal to zero).

It is also important to assess the assumption of time homogeneity, when a time-homogeneous-Markov model is using. We used principally three methods to assess this assumption. The first one is to used the piecewise constant intensities model illustrated in Section 2.2. In this case a likelihood ratio test can be used to compare the piecewise constant intensities model with the time-homogeneous model. Under

H_0 (homogeneous model) the test statistic has approximately a χ^2_{k-p} (p is the number of parameters under H_0 and k number of parameters under H_1). The second method is based on a local score test Kalbleisch and Lawless (1989); Bianca and De Stavola (1988). For instance, for any transition $\lambda_{ij}(t)$ with $i \neq j$ we can consider an alternative hypothesis such as $H_1 : \lambda_{ij(t)} = \lambda_{ij} + t\delta$ versus $H_0 : \lambda_{ij(t)} = \lambda_{ij}$. The test statistic is the ratio of the partial derivative of the log-likelihood with respect to δ, evaluated at $(\hat{\lambda_{ij0}}, \hat{\beta_{ij}}, \delta = 0)$, and an estimate of its standard deviation. The advantage of this method is that only the time-homogeneous model has to be fitted. For details in this method, we refer to the paper of Kalbleisch and Lawless (1985). The last method consist to evaluate the significance of parameters of the waiting times distribution. In our case, we test whether $H_0 : \nu_{ij(t)} = 1$ by using the likelihood ratio test (LRT).

3. Application on serological analysis

3.1. Data and Model schema

The population for this study was inhabitants from a rural area in Senegal, Dielmo located in west Dakar. These data are collected as part of the Dielmo project initiated in 1990 by an tripartite agreement between Pasteur Institute of Dakar (IPD), Institut of Research and development (IRD) and the ministry of health of Senegal. Subjects are included after giving their consent of their legal tutor Trape *et al.* (1994). The main objective of this project (Dielmo project) was to study the determinants of protection to malaria infection. At the beginning of the project, Dielmo was a malaria holoendemic locality characterized by a high and perennial parasite transmission where malaria clinical, entomological, parasitological parameters were closely monitored Trape *et al.* (1994); Rogier *et al.* (1999). Different interventions based on the recommendations of Malaria National Program were implemented in this locality. These interventions were, among others, the use of Chloroquine in 2000 which was replaced by the association Amodiaquine plus Sulfadoxine-Pyrimethamine between 2003 and 2006 (November 2003-May 2006), followed by the introduction of artemisinin-based combination therapies (ACT) since 2006 and the deployment of the long-lasting insecticidal impregnated net (LLIIN) since 2008. As a result, the village has experienced a dramatic decrease of the different malaria indicators monitored Diop *et al.* (2014). Between 2000 and 2012, the incidence of clinical malaria dropped from 771 in 2000 to only 17 cases in 2012 Trape *et al.* (2014). Dielmo transmission decreased considerably.

In this study we used serological data from archived blood sample of Dielmo's population collected during the lowest transmission season. We limit the sample to observations collected between 2000 and 2012. Participant's age constitute the chronological time of follow-up. Our sample is therefore constituted of 350 persons, representing a total of 1504 observations (serum). These participant's serum were taking away and tested for presence of antibodies against *P. falciparum* crude extract by clinicians. The level of antibodies reactivity are qualified as term of Optical Density (OD) that are proportional to quantities of antibody against antigens. The state of progression of antibodies response considered for multi-state model are de-

fined based in the ratio of optical density (OD) which is equal to the quotient of the optical density of participant on the average of the control group Diop *et al.* (2014). Subjects are classified as seropositive if their ratio of optical density is greater or equal to 2. Thus clinicians define two reversible states of reactivity. In this context we consider a process characterized by the Figure 1. Only data on participants who have at least two observations during the study period were considered in the analysis. Table 1 described the distribution of transitions. States 2 seems to be the more transitive states, regarding the number of observed transitions.

3.2. Results

In this section, we propose an application of the methodology developed in the previous section and an analysis of malaria serological data. In the first time, models without covariates for all techniques are calculated to evaluate the evolution reactivity of antibodies. Table 2 shows the estimates of the transition intensities and its confidence intervals on the time-homogeneous Markov model. In serological approach, the transition intensity from state 1 to state 2 (λ_{12}) so called seroconversion rate is related to the "force of infection" of malaria, as obtained among the antibodies response from exposed individuals. It might be used to measure the transmission intensity and can be compared to standard transmission measurement such as the entomological inoculation rate (EIR). The transition intensity from state 2 to state 1 (λ_{21}) which is the seroreversion rate gives the persistence of the antibodies response. The log-likelihood of this model is equal to -523.84 corresponding to an AIC of 1051.68.

To check the validity of time-homogeneous hypothesis, time piecewise constant intensities Markov and semi-Markovian models are used. For the time piecewise constant intensities three models were calculated: one with a single artificial time-dependant covariate z_1^* with the cut-time point at ($\tau_1 = 15$), one with two artificial time-dependant covariates z_1^* and z_2^* with two cut-time points at ($\tau_1 = 5$ years and $\tau_2 = 15$ years) and another one with four artificial time-dependant covariates with cut-time point at ($\tau_{i=1,2,3,4} = 5, 15, 25, 40$). Let's noticed that the following time of models is patient's age. In these models, transition intensity is assumed constant over time intervals for each transition. Only the estimates of the model with a single artificial time-dependant covariate were given here. The Table 3 shows the estimates of the baseline intensities, of the regression coefficients, its standard deviations (sd) and maximum of p-value of the Wald test and the likelihood ratio test (LRT) for testing $\beta_{ij}^* = 0$. For each transition, p-value is significant which means thatthe time interval (age groups) has a significant effect on transition intensities. The AIC of Markov piecewise constant intensities model using a single time-dependant covariate is 671.20 corresponding to a log-likelihood equal to -331.60 and it is lower than the AIC of the time-homogeneous model (cf Table4). This result shows that the time piecewise constant intensities model is most parsimonious than the time-homogeneous one (p ¡0.001, LRT). In addition we compare all time piecewise constant intensities models between them by using the AIC criterion. Model with four artificial time dependent covariates was the most

parsimonious (Table 4). Note that the AIC decreases as function of the number cut-time points. Time piecewise constant intensities models results indicate that the time-homogeneous Markov model is so restrictive in these malaria serological data.

Visual assessment of the transition intensities for the semi-Markov model shows their dependence on the waiting times (Figure 2). Transition intensity from state 1 to state 2 increases slightly.It comes out from this results that after entering at state 1, the risk function increases showing an instability of the antibody responses which has just changing on state 2. However, from state 2 to state , the risk decreases exponentially and tafter 5 years duration in the state 2 it becomes equal to zero. Results of the Wald test confirmed the previous ones. All parameters of the waiting times distribution, for each transition, are significantly different to one (Table 5), which means that the transition intensities depend on the waiting times. Thus according to the results of the piecewise constant intensities, the time-homogeneous intensities hypothesis is not verified for our data.

In order to investigate the effect of prognostics factor which can be influenced the different transition of the model, three covariates such as clinical access (healthy=1 no healthy=0), sex and the programs intervention are incorporated into the model. For each covariate, one model was fitted to identify relationship between these factors and any transition by using the Wald test and the LRT. Only sex is not significantly associated with any transition. Each univariate model is compared with the time-homogeneous model without covariate (basic model) using the AIC criterion and the LRT. Both AIC criterion and LRT show that all univariate models are more parsimonious than the basic model. Multivariate time homogeneous, multivariate time piecewise constant intensities models and multivariate semi-Markov models are computed.

Effect of clinical episodes (coefficient $\hat{\beta}_{12}$) is positive for all models, which means that healthy accelerate the passage of participant's serological markers from state 1 to state 2. Let note that, for the homogeneous Markov model, coefficient $\hat{\beta}_{12}$ which was significant on the univariate model becomes no significant on the multivariate model (p = 0.255, Table 6). For the transition $1 \rightarrow 2$, the effect of this covariate ($\hat{\beta}_{21}$ of clinical effect)is negative which means that disease reduce the reversion rate of antibody responses . For the use of mosquito nets (Intervention), results show that passage from state 2 to state 1 is not significantly affected ($p = 0.96$) and coefficient $\hat{\beta}_{21}$ is positive, which suggest that transition intensity from 2 to 1 is significantly accelerated by the use of mosquito nets bites.

For time piecewise constant intensities, all time artificial covariates excepting the fourth one are significantly associated with transition $1 \rightarrow 2$, which means that the time homogeneous is wrong in malaria serological case. For transition $1 \rightarrow 2$, practically same results of time homogeneous model are found for clinical access and intervention covariates. However, coefficient $\hat{\beta}_{21}$, for clinical access which is not significant for time homogeneous model becomes significant for the piecewise constant intensities one, which mean that clinical access is time-dependant

covariate and it may have a time-varying effect. Note that some coefficients which are not significantly different to zero may be constraint to equal to zero in the model, so they will not be estimated.

In the semi-Markov model, results of the stratified strategy are given in Figure 3. We observe gaps between curves for each transition. Graphical comparison shows a substantially parallelism between waiting hazard risk curve for each strate and each transition, which means that the proportional seems to be verify for covariates. For instance, for the transition $1 \rightarrow 2$ with intervention (Figure 3(C)) (respectively with clinical episodes (Figure 3(A))) such as stratification variable, there is not intersection in curves which reflect the proportionality. It is also noted that in the same transition, the shape of the curves of the intensities of the waiting times does not differ from one modality to another for each covariate.
According to the results of the stratified and the univariate strategies, two covariates are selected then six parameters are to be estimated for the multivariate model. In this model we obtained an AIC equals to 2079.87. We observe the positive effect of the healthy which confirm the results of the homogeneous et the piecewise constant intensities models. However for the intervention, the effect is negative for the transition $2 \rightarrow 1$ in contrary of the results of the other models.

For model selection, the LRT and the AIC show that time-piecewise constant intensities is better than time homogeneous. For assessing the goodness of fit of models, we compare observed data to the fitted one for each model. Figure 3 shows curves of transition probabilities from state 1 to state 2 both for time homogeneous and for piecewise constant intensities with four artificial time-dependant covariates. This figure confirms results of the LRT and the AIC. It also shows the importance to incorporate covariates into the model. Indeed covariates modified the evolution of transition probabilities in malaria serological case.

4. Discussion

In this paper we are interested on the study of multi-state Markov models without and with covariates. We summarize methods and estimation techniques for these models. We apply these processes to malaria serological data. These data described antibody activities against crude extract *P. falciparum* shizonts. For assessing the homogeneity assumption, we use the time-piecewise constant intensities with some appropriate time-cut points and semi-Markov approaches. We show that homogeneous Markov model often used is restrictive in malaria sero-epidemiological situation.

The LRT and the Wald test show that time is associated with any transition. This result confirms what is known in malaria context such as the dependence of antibodies response with age (time scale) (Perraut *et al.* (2000, 2003); Diop *et al.* (2014)). In view of these results it is obvious that homogeneous model is so restrictive to model malaria serological markers evolution. However, using

homogeneous model we demonstrate that the EIR combined with intervention programs is correlated to antibodies seroconversion rate (transition intensity from seronegative state to seropositive state). This association is recently observed in many studies in locality with low malaria transmission intensities (Cook *et al.* (2012); Stewart *et al.* (2009); Corran *et al.* (2007)). Results show also that malaria control programs and clinical access are associated with serological markers evolution. Moreover results on univariate models are confirmed when model with two or more covariates is adjusted. The fit of time piecewise constant model allows us to investigate time-homogeneous assumption and to consider eventually the use of no-homogeneous model for malaria sero-epidemiological markers evolution. In addition, the use of time piecewise constant model give interesting results. It allows us to confirm heterogeneity in the malaria transmission in the study site (Trape *et al.* (2014)). This work has many suggestion. Firstly, Markov models can be applied to assess malaria serological markers evolution as the previous study (**?**). Secondly, incorporation of covariates in the model is important to adjust calculation and give a better precision of transition probabilities in different states taking into account participant's individuality.

We meet some limitations on the use of Markov models in continuous-time. In the first hand, the assessment of the Markov assumption is not always easy. In the second hand, time-homogeneous Markov model is so restrictive and for the use of time piecewise constant model, it is not clear how to choose the number of time-cut points. It could be interesting to know how to deal with the number of time periods. Indeed, we encounter convergence problems in the estimation computation when the number of time-cut points is greater than four. Thirdly, in the use of time piecewise constant intensities, it could be interesting to calculate a model with interaction such as in time-homogeneous model. But estimation computation is so difficult. Indeed number of parameters increase proportionality with the number of covariates. For instance, in malaria serological case, there are sixteen parameters for the model with three covariates and eighteen parameters for model with interaction (four covariates). Some authors were proposed to use simplify versions of the model to overcome this difficulty (Marshall and Jones (1995); Cook *et al.* (2002)).

In view of results of piecewise constant intensities, it could be suitable to fit a non-homogeneous Markov model for malaria serological markers evolution. It is also interesting to consider other Markov model to take into account time spent in the state. Finally, it could be important to extend Markov models with time-independent covariate effects to time-dependant effects in most disease history studies specially in malaria serological markers evolution.

Acknowledgements

This research was supported by grant from EDCTP,Pasteur Institute of Dakar and grant from SCAC (Service de coopération et d'Action Culturelle) of French embassy. The authors thank all the staff of Immunology unit (Pasteur Institute of Dakar),

of LSTA (Laboratoire de Statistique Théorique et Appliqué,Paris and of LERSTAD (Laboratoire d'Étude et de Recherche en Statistique et Développement, Saint-Louis, Senegal and the CEA-MITIC. Authors wish also to acknowledge Dielmo populations for their collaborations.

References

Longini I. M. J., Clark W. S., Byers R. H., Ward J. W., Darrow W. W., Lemp G. F. and Hethcote H. W.(1989) Statistical analysis of the stages of hiv infection using a markov model. *Statistics in Medecine*, 8(7):831-843.

Alioum A, Leroy V., Commenges D., Dabis F. and Salamon R. (1998). Effect of gender, age, transmission category, and antiretroviral therapy on the progression of human immunodeficiency virus infection using multistate markov models. *groupe d'épidemiologie clinique du sida en aquitaine. Epidemiology*, 9:605-612.

Kay R. (1986) A markov model for analysing cancer markers and disease states in survival studies. *Biometrics*, 42(4):457-481.

Hsieh H. J., Chen T. H. H. and Chang S. H. (2002) Assessing chronic disease progression using non-homogeneous exponential regression markov models : an illustration using a selective breast cancer screening in taiwan. *Statistics in Medecine*, 21:3369-3382.

Gentleman RC., Lawless JF., Lindsey JC. and Yan P.(1994) Multi-state markov models for analysing incomplete disease history data with illustrations for hiv disease. *Statistics in Medecine*, 13(8):805–821.

Jackson C., Sharples L., Thompson S., Duffy S., and Couto. (2003) Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society Series*, C52:193.

Marshall G. and Jones R. H. (1995) Multi-state models and diabetic retinopathy. *Statistics in Medecine*, 14(18):1975-1983.

Saint-Pierre P,Combescure C, Daurès JP and Godard P. (2003) The analysis of asthma control under markov assumption with use of covariates. *Statistics in Medecine*, 22:3755–3770.

Combescure C., Chanez P., Saint-Pierre P., Daures JP., Proudhon H., and Godard P. (2003) Assessement of variations in control of asthma over time. *European Respiratory Journal*, 22:298-304.

Bekessy A., Molineaux L. and Storey J. (1976) Estimation of incidence and recovery rates of plasmodium falciparum parasitaemia from longitudinal data. *Bull. World Health Organ*, 54:685–693.

Drakeley CJ., Corran PH., Coleman PG., Tongren JE., McDonald SL. et al.(2005) Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proc Natl Acad Sci U S A*, 102(14):5108-5113.

Corran P, Coleman P, Riley E and Drakeley C. (2007) Serology: a robust indicator of malaria transmission intensity? *Trends Parasitol*, 23:575-582.

Stewart L, Gosling R,Griffin J, Gesase S, Campo J, Hashim R, Masika P, Mosha J, Bousema T, Shekalaghe S, Cook J, Corran P, Ghani A, Riley EM and Drakeley, C.(2009) Rapid assessment of malaria transmission using age-specific sero-conversion rates. *PLoS One*, 4:e6083.

Supargiyono S, Bretscher MT, Wijayanti MA, Sutanto I, Nugraheni D, Rozqie R, Kosasih AA, Sulistyawati S, Hawley WA, Lobo NF and Cook J. (2013) Seasonal changes in the antibody responses against plasmodium falciparum merozoite surface antigens in areas of differing malaria endemicity in indonesia. *Malaria Journal*, 12:444.

Alioum A. and Commenges D. (2001) Mkvpci : a computer program for markov models with piecewise constant intensities and covariates. *Comput Methods Programs Biomed*, 64(2):109-119.

Lindsay JC and Ryan LM. (1993) A three-state multiplicative model for rodent tumorigenicity experiments. *Applied Statistics*, 42:283-300.

Diop F., Richard V. et al. (2014) Dramatic declines in seropositivity as determined with crude extracts of plasmodium falciparum schizonts between 2000 and 2010 in dielmo and ndiop, senegal. *Malaria Journal*, 13:83.

Rosas-Aguirre A, Llanos-Cuentas A, Speybroeck N, Cook J, Contreras-Mancilla J, Soto V, Gamboa D, Pozo E, Ponce OJ, Pereira MO, Soares IS, Theisen M, D'Alessandro U and Erhart A. Assessing malaria transmission in a low endemicity area of north-western peru. *Malaria Journal*, 12:339.

Cox DR. and Miller HD. (1977) *The theory of stochastic processes*. CRC Press.

Cox DR. (1972) Regression models and life tables (with discussion). *J. R. Stat. Soc. B*, 34:187-220.

Kalbleisch JD and Lawless, J. (1985) The analysis of panel data under a markov assumption. *journal of American Statistical association*, 392(90):863-871.

Foucher Y., Mathieu E., Saint-Pierre P. and Durand J. (2005) A semi-markov model based on generalized weibull distribution with an illustration for hiv disease. *Biometrical Journal*, 47:825-833.

Kalbfleisch JD and Lawless JF. (1989). Some statistical methods for panel life history data. *Proceedings of the Statistics Canada Symposium on the Analysis of Data in Time, Ottawa, Statistics Canada*, page 85-192.

Bianca L and De Stavola (1988). Testing departures from time homogeneity in multistate markov processes. *Applied Statistics*, 37:242–250.

Trape JF, Rogier C, Konate L, Diagne N, Bouganali H, Canque B, Legros F, Badji A, Ndiaye G, Ndiaye P, et al. The dielmo project: a longitudinal study of natural malaria infection and the mechanisms of protective immunity in a community living in a holoendemic area of senegal. *Am J Trop Med Hyg*, 51(2):123-37.

Rogier C, Ly AB, Tall A, Cisse B and Trape JF. (1999). Plasmodium falciparum clinical malaria in dielmo, a holoendemic area in senegal: no influence of acquired immunity on initial symptomatology and severity of malaria attacks. *Am J Trop Med Hyg*, 60(3):410-20.

Trape JF, Tall A, Sokhna C, Ly AB, Diagne N, et al. (2014) The rise and fall of malaria in a west african rural community, dielmo, senegal, from 1990 to 2012: a 22 year longitudinal study. *Lancet Infect Dis*, 14(6):476-88.

Perraut R, et al. (2000) Seasonal fluctuation of antibody levels to plasmodium falciparum parasitized red blod cell-associated antigen in two senegalese villages with different transmission condition. *Am T Trop Med Hyg*, 62:746-51.

Perraut et al. (2003) Distinct surrogate markers for protection against plasmodium falciparum infection and clinical malaria identified in a senegalese community

after radical drug cure. *J Infect Dis*, 188:1940-1950.

Cook J,Speybroeck N, Sochanta T, Somony H, Sokny M, Claes F, Lemmens K, Theisen M, Soares I.S, Umberto D'Alessandro, Marc Coosemans and Annette Erhart. Sero-epidemiological evaluation of changes in plasmodium falciparum and plasmodium vivax transmission patterns over the rainy season in cambodia. *Malaria Journal*, 1:11:86.

Cook RJ, Lawless JF and Yi GY. A generalized mover-stayer model for panel data. *Biostatistics*, 3:407-420.

5. Figures and tables

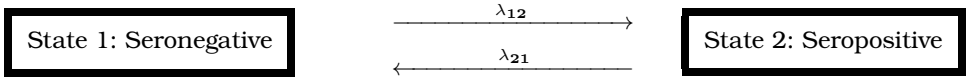

Fig. 1. Two-state Markov model for *P. falciparum* malaria serological markers

Transition	Effectives	Percentage
$1 \rightarrow 1$	139	12.15
$1 \rightarrow 2$	343	29.98
$2 \rightarrow 1$	100	8.74
$2 \rightarrow 2$	562	49.13

Table 1. Distribution of observed transitions

Table 2. Estimates of baseline transition intensities with confidence intervals (CI) for homogeneous Markov model without covariates

Baseline transition intensities	Estimate (CI)
$\lambda_{12}(1 \rightarrow 2)$	$0.284(0.23, 0.35)$
$\lambda_{21}(2 \rightarrow 1)$	$0.066(0.05, 0.08)$

Table 3. Estimates of baseline transition intensities with confidence intervals (CI) and regression coefficients (effect of the time-artificial covariate) with standard deviation and maximum p-value of Wald test and LRT for the piecewise constant intensities Markov model

Baseline transition intensities	Estimate (CI)	Regression coefficients	Estimate (sd)	p-value
λ_{120}	$0.22(0.17 - 0.28)$	β^*_{12}	0.452 (0.164)	0.037
λ_{210}	$0.094(0.08 - 0.11)$	β^*_{21}	-0.789 (0.168)	0.054

	Without covariates		
Model	AIC	-2 log LR	p-value
basic model	1051.671	1	
	671.185	384.487	< 0.001
	651.410	398.261	< 0.001
	660.279	407.392	< 0.001

Table 4. Criterion AIC and log-likelihood for model selection

	parameters of the Weibull Distribution		
Transition	$\sigma_{ij}(sd)$	$\nu_{ij}(sd)$	p-value
$1->2$	26.32 (24.17)	1.42(0.06)	< 0.001
$2->1$	0.013 (0.001)	0.13 (0.01)	< 0.001

Table 5. Estimation of parameters of the distribution of the waiting times on the Weibull Semi-Markovian model without Covariates.

Table 6. Regression coefficients (Coefficients) estimates with standard errors and p-values for homogeneous Markov model with covariate and for time piecewise constant intensities with four artificial time-dependant covariates (z_1^* ,z_2^*,z_3^*, z_4^*) with cut-time point at ($\tau_{i=1,2,3,4} = 5, 15, 25, 40$years) and with covariates.

		Models				
		Time-homogeneous		Time piecewise constant intensities		Semi-Mark
Covariates	Transition	β_{ij} (sd)	p-value	β_{ij} (sd)	p-value	β_{ij} (sd)
Clinical Access	$1->2$	0.437 (0.14)	0.255	0.638 (0.015)	0.004	0.751(0.53)
	$2->1$	-0.001 (0.352)	< 0.0001	$-0.457(0.16)$	< 0.001	-1.622 (0.37)
Intervention control	$1->2$	0.042 (0.049)	0.851	0.009 (0.12)	0.02	0.238(0.11)
	$2->1$	0.091 (0.306)	0.767	$-0.011(0.25)$	0.56	-4.662 (0.24)

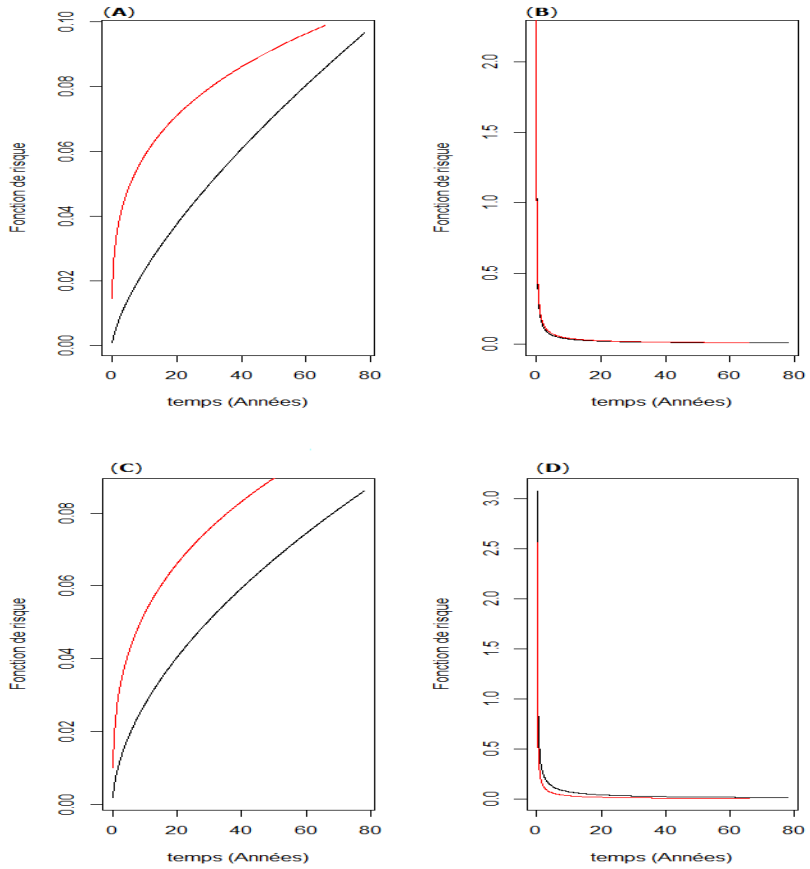

Fig. 2. Hazard function of the waiting time distribution

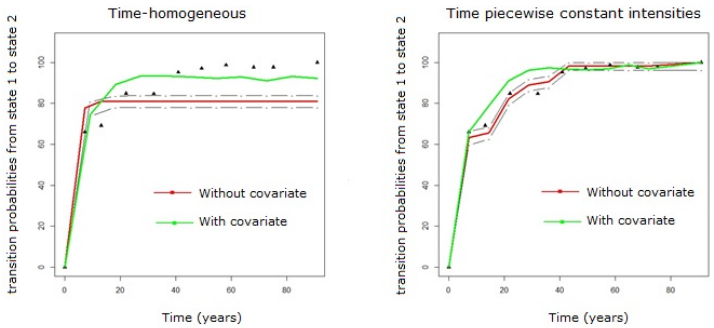

Fig. 3. transition probabilities from state 1 to state 2 curves. black triangles represent the observed data, green curves the predicted transition probabilities for multivariate models, unbroken lines represent maximum likelihood transition probabilities from state 1 to state 2 curves and broken lines 95% confidence intervals