PERGAMON Mathematical and Computer Modelling 38 (2003) 251–267

MATHEMATICAL AND COMPUTER MODELLING

www.elsevier.com/locate/mcm

The Seroreversion and the Survival Related to HIV Infection Among Children: Statistical Modeling Applied to Retrospective Data Collection

HYUN MO YANG*
IMECC—Departamento de Matemática Aplicada
Universidade Estadual de Campinas
Caixa Postal 6065; CEP: 13081-970; Campinas, S.P., Brazil
hyunyang@ime.unicamp.br

M. DELLA NEGRA, YU CHING LIAN AND W. QUEIROZ
Instituto de Infectologia Emílio Ribas
Av. Dr. Arnaldo, 165; CEP: 01246-900; São Paulo, S.P., Brazil

L. K. HOTTA
IMECC—Departamento de Estatística
Universidade Estadual de Campinas
Caixa Postal 6065; CEP: 13081-970; Campinas, S.P., Brazil

(Received May 2000; revised and accepted April 2003)

Abstract—We are interested in describing the biological phenomenon of the time of seroreversion or death due to AIDS. By the seroreversion, we mean the cleansing of maternally derived antibodies against HIV by seropositive children without HIV infection, and the survival is related to how long HIV infected children survive. In order to perform *a posteriori* descriptions of retrospective data related to babies from HIV infected mothers, we develop a simple statistical methodology. The resulting statistical model is a function of two parameters named half-age and phenomenon-gauge. The model's parameters are fitted to the collection of seropositive children's data by the maximum likelihood estimation method. We compare this parametric estimation with nonparametric Kaplan-Meier estimation with respect to the half-age.

Keywords—HIV, AIDS, Seroreversion, Survival, Stochastic process, Estimation, Phenomenon-gauge, Half-age.

1. INTRODUCTION

The infection of children with HIV by the vertical transmission route has been reported since 1982 [1]. Since then the epidemiological evidence pointed to prenatal acquisition of HIV as the principal route of transmission [2]. Other routes may occur as perinatal transmission *in utero* via transplacental infection of the fetus, during labour and delivery by contact with infected blood and secretion, and postnatally via breast milk [3–5].

*Author to whom all correspondence should be addressed.
Financial support from CNPq Grant No. 300627/93-9(RE).

0895-7177/03/$ - see front matter Typeset by AMS-TEX
doi: 10.1016/S0895-7177(03)00218-8

Most authors believe that the majority of the children are infected during labour, due to the contact with maternal infected blood and secretions. In a multicentric survey with twins born to HIV infected mothers, it was observed that the first twin showed a higher rate of HIV infection than the second one, probably due to longer contact with maternal secretions [6]. The HIV infected (pregnant) women were initially those belonging to high HIV at-risk group, or those in contact with recipients of blood or blood products, or those whose sexual partners belonged to a high HIV at-risk group. However, nowadays, many other groups of women are at-risk to HIV infection, such as housewives, for instance. The most important risk factor for HIV infection among female population in Brazil is through sexual relationships [7].

Since the HIV transmission from an infected mother to her baby occurs most probably late in pregnancy or during labour and delivery, some drugs can be used to decrease this transmission. Hence, antiretroviral prophylaxis of the neonate during the time of HIV exposure at birth can be done by zidovudine or nevirapine, because techniques as the CD4 cell counts, HIV RNA PCR assays and HIV culture permit the confirmation of a vertical transmission of HIV [8]. However, when an anti-HIV antibodies screening of the babies born from HIV infected mothers is performed, not all the seropositive children are effectively HIV infected because some of them may harbour only maternally derived antibodies against HIV. Moreover, if we are restricted only to the anti-HIV antibody test by EIA and western blot in children, the following-up of these seropositive newborns permits us to decide about the status of HIV infection. The reason is that this follow-up of newborns provides us with two mutually exclusive reachable situations by a seropositive newborn: the cleansing of maternally derived antibodies against HIV (*serorevertion* phenomenon, hereafter) and the vertically transmitted HIV infection (*survival* phenomenon, hereafter). The HIV infected newborn generally develops into a pediatric AIDS case, which shortens the life span due to death caused by an infection or other opportunistic diseases.

The serorevertion phenomenon is characterized by the continuous decaying of maternally derived antibodies against HIV, since the newborn effectively uninfected by HIV does not produce antibodies. This decaying can be approximated by an exponential function [9], that is, the decreasing rate in the amount of anti-HIV antibodies is proportional to the actual amount of antibodies. In turn, the death of vertically infected HIV children can be understood as a result of the cumulative destruction of the helper/inducer subpopulations of CD4 T-lymphocytes by HIV. In the beginning of the infection, the immune system is activated and controls the disease, but after a variable period of time, the immune system is weakened and initiates the period when the rate of depletion of T cells by HIV is higher than their production, leading to the deficiency in the immune response. This immune-deficiency is the gate by which other infections can be initiated and rapidly spread out, leading sometimes to death. This diminishing of T cells population can roughly be described by an exponential function, although with lower depletion rate than the antibodies degrading rate. Observe that in both phenomena, there is continuous depletion of antibodies and CD4 cells with time, that is, we have occurrences flowing in only one-direction, contrary to many physical phenomena where the flow of the variation can be reversed, for instance, in physics, a particle can have its energy either increased or decreased.

Our goal is the development of a methodology to treat the previously discussed phenomena concerning HIV infection among children, which is a unidirectional flow completely dependent on the observations along the time. This methodology is applied to records obtained from follow-ups of seropositive children from the Emílio Ribas Infectology Institute, located in São Paulo State, Brazil. These data (a retrospective collection) consisted of seropositive newborns followed-up from December 1985 until September 1996, which reported only the age at which they seroreverted or died (further details about the data are presented in Section 3). Although the factors that determine the serorevertion and survival, which are, respectively, the cleansing of maternally derived anti-HIV antibodies and cumulative destruction of CD4 T-lymphocytes by HIV, can now be measured [8], the data were not available in the study.

To deal with this kind of retrospective data collection, we must develop an appropriate device. The level of anti-HIV antibodies concentration (in the seroreversion) and the remaining amount of CD4 T-lymphocytes cells (in the survival), which can be related to appropriate biological conditions, were not available in the data collection. Nevertheless, the seroreversion and the survival can be completely assessed by following up, and for this reason, we will introduce a latent variable to assess the unavailable biological conditions to help us to construct an age-specific methodology to treat the data. So, conditionally to the latent variable, the biological conditions are irrelevant. Another feature of these data collections is related to the great variability among the followed-up children with respect to the occurrence of the seroreversion and the death due to HIV/AIDS.

Taking into account the above considerations, in this paper, we derive a probabilistic survival distribution to deal with the seroreversion and the survival phenomena. Regarding the phenomenon of the cleansing of maternally derived antibodies against HIV by uninfected children, we deal with the change of status of seropositive children (initial state) to seroreverted children (final state). On the other hand, to treat the phenomenon of vertically transmitted HIV infection, we deal with the change of status of HIV infected children (initial state) to children dying due to HIV/AIDS (final state). To analyse the above two transitions from the initial to final states, we are assuming that all the followed-up children have been breast-feeding equally and are in the same environment and under the same care (hospitals or care centers) conditions.

The statistical model is based only on two parameters of easy interpretation and is used to describe both seroreversion and survival phenomena. The first parameter is the *half-age*, i.e., the age at which the probability of occupying the initial state (or final state, which are the seroreversion or the death of HIV/AIDS children) is one half, and the other is the *phenomenon-gauge*, i.e., a parameter to discriminate the seroreversion and survival phenomena, which occur in distinct regions of ages largely separated. However, inside a specified phenomenon, the latter parameter also measures the heterogeneous children responses facing the same external stimuli applied on them. These two parameters may depend on many factors [10–12]. Among them, we state the risk category of their mothers, for instance, if she belongs to an injecting drug user (IDU) group or if her sexual partner is HIV infected or both; the individual response to cleanse the maternally derived antibodies or to the development to AIDS pediatric; the treatment in course and the BCG vaccination. In this paper, we estimate both parameters by the maximum likelihood method.

This paper is divided as follows. Section 2 contains the development of a statistical model in terms of the parameters half-age and phenomenon-gauge. Section 3 compares the nonparametric Kaplan-Meier estimator, which is in Section 3.1, with the proposed parametric model, which is in Section 3.2. Finally, the results are commented on and discussed in Section 4.

2. THE STATISTICAL APPROACH

In this section, we deal with the time duration between the birth and the time of seroreversion or death, stochastically.

Let us assume that the transition time is related to a latent variable, say W, which is based on the biological conditions of a child and is not necessarily observable. We have two distinct biological conditions, according to the phenomenon in the study. With respect to the survival of the child infected with HIV, the main biological condition is related to concentration of CD4 cells in the blood stream. In this case, high concentration of CD4 cells indicates a better health condition, and its lowering results in AIDS disease and death. On the other hand, the seroreversion deals with children with high levels of maternally derived anti-HIV antibodies that diminish progressively until their cleansing, while their health condition is unchanged because they are uninfected children. Note that both phenomenon consider the transition from high level to lowering of CD4 cells (anti-HIV antibodies). Therefore, the latent variable W relates inversely with

the biological conditions, that is, lower values of W indicate high probability of maintaining the initial state. Additionally, let us suppose that the distribution of this latent variable at the transition is supposed to be an exponential function of the hazard rate λ [13]. The interpretation of the parameter λ follows below.

Let us assume that conditionally to the biological conditions ω, the distribution of the age A is proportional to the distribution of $\ln(X+1)$, where X has an exponential distribution with (constant) hazard rate ω. The rationale behind it is that the expectation of the transition age, conditionally to the biological conditions, decreases proportionally to the logarithm of the biological conditions ω, that is,

$$D(A \mid W = \omega) = D\left[\xi^{-1} \ln(X+1)\right]. \tag{1}$$

$D(A \mid W = \omega)$ is the distribution of age (the time duration between the birth and the transition) conditionally to the biological conditions and ξ is a positive auxiliary parameter relating age and biological conditions. The exponential relation between the elapsed time since the birth and the biological conditions (the decrease in the maternally derived antibodies anti-HIV and the reduction in the number of T-lymphocyte cells) were taken according to the empirical observations. On the other hand, the data presented a large range of variation and the log-transformation provided a better fitting [14].

Conditionally to w, the probability function of the age A at the transition is given by

$$f_{(A|W)}(a \mid \omega) = \omega \xi e^{\xi a} e^{-\omega\left(e^{\xi a}-1\right)}, \qquad a > 0 \text{ and } \omega > 0. \tag{2}$$

Thus, the probability that a child keeps the initial state, conditionally to $W = \omega$, is given by

$$S_{(A|W)}(a \mid \omega) = e^{-\omega\left(e^{\xi a}-1\right)}, \qquad a > 0 \text{ and } \omega > 0, \tag{3}$$

with the marginal survival function given by

$$S_A(a) = \frac{1}{1 + \lambda^{-1}\left(e^{\xi a} - 1\right)}, \tag{4}$$

where weight is taken as the biological conditions [15].

Now we can interpret the parameter λ. Let us define the half-age τ as the age at which the probability of keeping the initial state is half. Then, we must have at this half-age $S_A(\tau) = 1/2$, which is satisfied if we have $\lambda = e^{\xi\tau} - 1$. Substituting this in expression (4), we obtain

$$S_A(a) = \frac{1}{1 + \left(e^{\xi a} - 1\right)/\left(e^{\xi\tau} - 1\right)}. \tag{5}$$

Therefore, the hazard rate λ, which is related to the biological conditions, can be set in terms of the half-age τ by

$$\tau = \frac{\ln(\lambda + 1)}{\xi}, \tag{6}$$

according to relation (1).

From the probability given by equation (5), we can calculate the probability density of the transition at age a, given by

$$f_A(a) = \frac{\xi e^{\xi a}}{\left(e^{\xi\tau} - 1\right)\left(1 + \left(e^{\xi a} - 1\right)/\left(e^{\xi\tau} - 1\right)\right)^2}, \tag{7}$$

which comes from the relation $f_A(a) = -\frac{dS_A(a)}{da}$, and the hazard function, given by

$$h(a) = \frac{\xi e^{\xi a}}{\left(e^{\xi\tau} - 1\right)\left(1 + \left(e^{\xi a} - 1\right)/\left(e^{\xi\tau} - 1\right)\right)}, \tag{8}$$

which comes from the relation $h(a) = f_A(a)/S_A(a)$. Observe that the logistic density can be retrieved if we have $e^{\xi a} \gg 1$ ($a \gg 1$) and $e^{\xi\tau} \gg 1$ ($\tau \gg 1$), where τ and $1/\xi$ are, respectively, parameters of location and scale [16].

We can analyse the limiting distributions of the probability given by equation (5) with respect to ξ to better understand this parameter. In order to do this, we define the age related transition rate ξ as being the inverse of the parameter β (dimension of *time*), that is,

$$\beta = \frac{1}{\xi}. \tag{9}$$

The parameter β can be considered as a measure of the phenomenon-gauge to discriminate between phenomena, and intrinsically inside the specified phenomenon, measures the heterogeneous response of the population.

Let us consider the case $\beta \ll 1$. In this case, equation (5) takes the form

$$S_A(a) = \frac{1}{1 + e^{-(\tau - a)/\beta}}, \tag{10}$$

where we have used $e^{\xi a} - 1 = e^{a/\beta} - 1 \approx e^{a/\beta}$ and $(e^{\tau/\beta} - 1)^{-1} \approx e^{-\tau/\beta}$. The related functions $f_A(a)$ and $h(a)$ can be easily obtained. Note that if we change the variables according to $a \to \varepsilon$, $\tau \to \varepsilon_F$, and $\beta \to kT$, the resulting equation can be compared with the Fermi-Dirac distribution of particles with energy ε, which is given by

$$FD(\varepsilon) = \frac{1}{1 + e^{-(\varepsilon_F - \varepsilon)/kT}},$$

where ε_F is the Fermi energy, k is the constant of Boltzmann, and T is the absolute temperature [17]. Note that when $\beta \to 0$, all the transitions occur at the same age.

The Fermi-Dirac distribution relates the particles that are indistinguishable and obey the principle of exclusion. Also, the newborns we are dealing with are indistinguishable (the initial state is characterized by the seropositive or HIV infected children) and the initial and final states are mutually exclusive. However, the main difference is related to the two-directional energy variation of particles (equilibrium transition) in opposition to the one-directional time flow. Also, note that the phenomenon-gauge parameter was related to describe two phenomena, although the biological conditions are different. However, note that the absolute temperature (kT) in Fermi-Dirac statistics plays a role similar to that we have defined for the phenomenon-gauge (β) if a phenomenon is specified. Therefore, as the temperature disturbs the system of particles to a large spectra of occupied energy states, in some degree the phenomenon-gauge, a parameter dependent on the whole biological constitution of the individual, measures the diversity in the physiological response when the individual is followed up to study or serorevertion or survival.

The other extreme is the case $\beta \to \infty$. This case corresponds to the limiting situation of the survival phenomenon. We have the probability of occupying the initial state given by

$$S_A(a) = \frac{1}{1 + (a/\tau)}, \tag{11}$$

and the related functions $f_A(a)$ and $h(a)$ can be easily obtained. This distribution takes the shape of hyperbolic curve (vertical asymptote at $a = -\tau$).

In the next section, we apply the probability given by equation (5) to retrospective collection of data.

3. FITTING THE DATA

Preliminarily, we describe the 583 followed-up children from the Emílio Ribas Infectology Institute, from December, 1985 until September, 1996. The collected data consist of retrieving

from the hospital's records the age at which the serorevertion or death occurred. With respect to the drug administration, a fraction of these followed-up children were treated, and the remaining children without treatment are comprised by the earliest cases, when the drug treatment was not available. In this first analysis, we are not taking into account the effects of treatment in the surviving children.

Of the total of 583, we discharged 22 children due to the incomplete information related to the serorevertion or death. A summary of 561 followed-up children is shown in Table 1.

Table 1. This data is from the Emílio Ribas Infectology Institute. We present the number of samples (#) by gender, with their average (mean) and standard deviation (St. Dev.) with respect to ages, expressed in months. The states of children are seropositive (0), HIV/AIDS (1), seronegative (2), and death (3).

	Both			Female			Male		
State	#	Mean	St. Dev.	#	Mean	St. Dev.	#	Mean	St. Dev.
0	27	7.59	5.58	11	7.96	5.99	16	7.34	5.47
1	192	67.48	39.95	85	59.60	31.00	107	73.74	45.00
2	172	12.56	7.44	91	12.81	7.37	81	12.28	7.56
3	170	39.98	39.07	88	38.93	36.00	82	41.10	42.33

In this paper, we are not considering the pregnant women's risk group or the age at which the drug administration was started. These factors may influence mainly on the half-age, and the analysis is left to a further paper. Here we restrict our analysis to the effect of gender. The data are fitted by nonparametrized and parametrized methods.

For the estimation of the parameters in the serorevertion study, we considered 27 seropositive children as censored and 172 seronegative children as transited, in a total of 199 children. The mean and the standard deviation for these data are 11.89 ± 7.41 months. If we consider gender, for 102 females, we found 12.29 ± 7.36 months, and for 97 males, 11.47 ± 7.46 months. The maximum observed age of female children followed up is 37.3 months, while for males, 56.9 months.

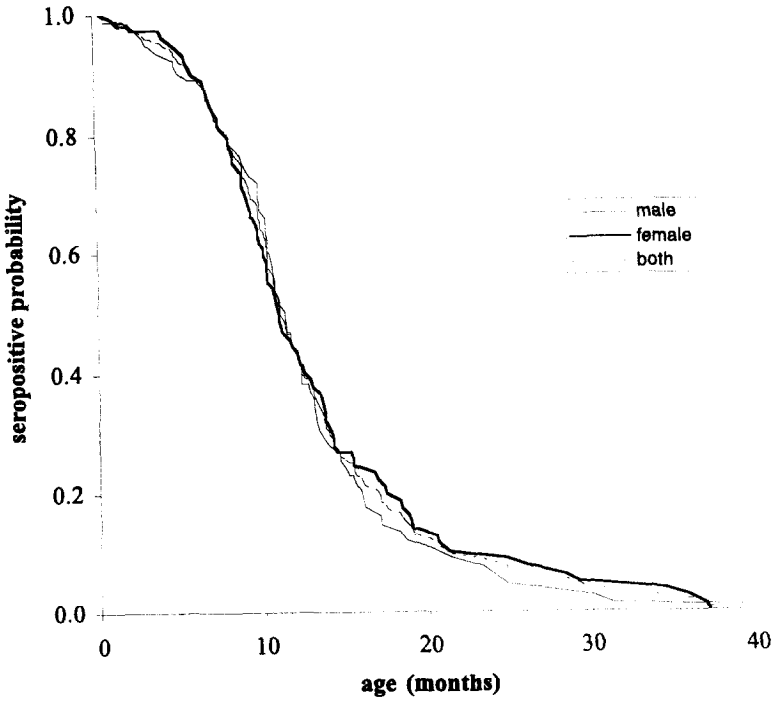
For the estimation of parameters related to vertically HIV infected children (survival study), we considered 27 seropositive and 192 HIV and AIDS symptomatic children as censored (partial sum 219) and 170 HIV/AIDS children who died as transited, in a total of 389 children. The mean and the standard deviation for these data are 51.30 ± 42.10 months. If we consider gender, for 184 females, we found 46.63 ± 35.43 months, and for 205 males, 55.50 ± 46.97 months. The maximum observed age of female children followed up is 170.4 months, while for males, 203.2 months.

The reason for including the 27 seropositive children, whose destiny is not yet defined, in both studies are as follows. The first is our total ignorance about their situation, since we are taking into account only the level of anti-HIV antibodies, and the other is the relatively small number of children in this class, which is around 10% of the sample size. Therefore, instead of developing a complicate probability distribution to state these children are in one of the two studies, we simply added these children in both studies. On the other hand, with the inclusion of the censored data, the comparison between parametric and nonparametric models becomes more appealing.

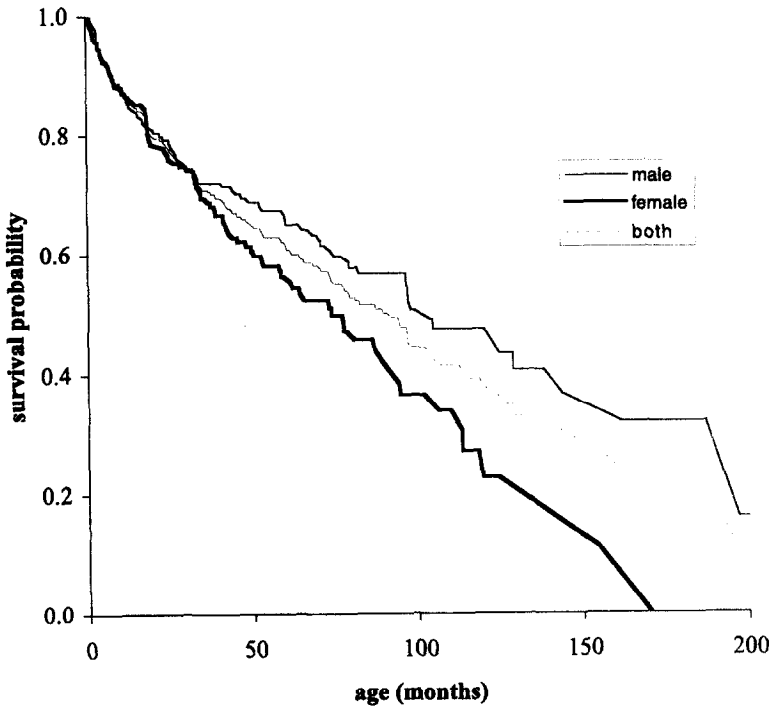
In the next sections, we deal with the nonparametric and parametric estimation of the model's parameters. In this paper, we are not considering the estimation of the vertical transmission rate [18] nor the reduction in the risk of vertical transmission due to drug treatment, but simply following up seropositive children to assess the half-age τ and the phenomenon-gauge β.

3.1. Nonparametric Kaplan-Meier Estimation

The data are fitted by the nonparametrized Kaplan-Meier estimation method [19]. In Figures 1a and 1b, we show the Kaplan-Meier estimation.



(a)



(b)

Figure 1. The nonparametrized Kaplan-Meier estimation of the serorevertion and survival phenomena: the probability function of keeping the seropositive state (a) and the alive state (b).

Observe that the serorevertion probability follows a reversed sigmoidal shape, while the survival probability can be described by either decaying exponential or reciprocal function.

From Figure 1, we can determine numerically the half-age τ provided by the Kaplan-Meier estimation method. The result is shown in Table 2.

Table 2. The Kaplan-Meier estimates of the half-age (τ, in months). The estimation is based on the data presented in Table 1, for the serorevertion and the survival.

Gender	Serorevertion	Survival
	τ	τ
Both	11.10	91.23
Female	11.03	73.81
Male	11.44	101.31

In relation to the serorevertion, the probability of maintaining the initial state attains one half around 11 months for both genders and, thereafter, the probability decreases quickly. The age at which the probability of surviving attains one half differs significantly with the gender. We obtained the half-age 73.8 and 101.3 months for, respectively, females and males. We observe that all children are seroreverting at almost the same ages (11.03 for females and 11.44 for males), but with respect to those surviving, the female children die earlier than males.

We can conclude from these findings that there are indications that gender is not important with respect to the seroreverting phenomenon, but plays an important role in the survival phenomenon.

3.2. Parametric Estimation

The parameters of the probability, given by equation (5), and the probability density of transition, given by equation (7), can be estimated by the maximum likelihood method. By applying the invariance principle, we estimate both the half-age τ and the phenomenon-gauge β.

We maximize the log-likelihood l, given by

$$l = \sum_t \ln\left[f_A\left(a_t\right)\right] + \sum_c \ln\left[S_A\left(a_t\right)\right], \tag{12}$$

where $S_A(a)$ and $f_A(a)$ are, respectively, given by equations (5) and (7) and the subscripts t and c stand, respectively, for children who transited from initial state to final state and who are censored.

Table 3 presents the estimated model's parameters considering the raw data.

Table 3. The fitted half-age (τ in months and variance σ_τ^2), ξ (in months^{-1} and σ_ξ^2) and the covariance $\eta_{\tau\xi}^2$, with corresponding minus the logarithm of the likelihood ($-l$). df stands for the degree of freedom. The estimation is based on the data presented in Table 1, for the serorevertion and the survival.

Parameters	Serorevertion			Survival		
	Both	Female	Male	Both	Female	Male
τ	12.01	12.07	11.87	81.16	67.60	98.03
σ_τ^2	0.32	0.70	0.59	52.88	51.97	215.7
ξ	0.244	0.232	0.258	0.0045	0.011	0.0008
σ_ξ^2	0.0002	0.0005	0.0004	0.000007	0.00002	0.00001
$\eta_{\tau\xi}^2$	0.003	0.007	0.004	-0.007	-0.01	-0.02
-l	572.8	304.6	268.0	978.9	491.0	484.6
df	196	99	94	386	181	202

First, we observe a much higher value for $\xi(= 1/\beta)$ in the serorevertion phenomenon in comparison with the survival phenomenon. The interpretation of this parameter as a phenomenon-gauge is very appropriate because this parameter can be scaled according to the phenomenon in study. Second, the early serorevertion of males in relation to females shown by parametrized estimation can be better understood by looking to the estimated phenomenon-gauge. Observe that the male children ($\beta = 3.88$) are more homogeneous than females ($\beta = 4.31$). This parameter justifies the

reversion in the seroreversion half-age when the parametrized estimation is compared with the nonparametrized one. But, for survival, we observe a great heterogeneity among males. Observe that the estimated values of β for the seroreversion and survival phenomena are situated at a very distant interval of age.

Note that, to deal with the raw data, we must calculate the values for the exponential function for large ages. For this reason, we transform the age by taking the logarithm (Appendix), according to $\ln(a+1)$. Table 4 shows the estimates related to the log-transformed data

Table 4. The likelihood estimation using the log-transformed data, $\ln(a+1)$. The legends are equals to those described in Table 3.

Parameters	Seroreversion			Survival		
	Both	Female	Male	Both	Female	Male
τ'	2.52	2.53	2.52	4.44	4.24	4.65
$\sigma^2_{\tau'}$	0.0013	0.0025	0.0029	0.01	0.02	0.03
ξ	3.64	3.61	3.68	1.03	1.15	0.93
σ^2_ξ	0.04	0.08	0.08	0.006	0.0132	0.0129
$\eta^2_{\tau'\xi}$	−0.0006	−0.0001	−0.003	−0.004	−0.007	−0.01
$-l$	561.7	298.1	263.6	980.3	493.7	484.4
df	196	99	94	386	181	202

We observe that the log-transformation of the data diminished the distance of the estimated value of ξ between seroreversion and survival phenomena. This is a result of the log-transformation which shortened the distance of the regions of age where both phenomena occur. Using the relation between τ' and the half-age τ given by expression (19), the estimates of the half-age for seroreversion and survival are (in months), respectively, $\tau = 11.49$ and $\tau = 83.92$ for both genders, $\tau = 11.54$ and $\tau = 68.36$ for females, and $\tau = 11.42$ and $\tau = 103.93$ for males.

Comparing the half-ages provided by Table 4 with Tables 2 and 3, we observe that the log-transformed data present values closer to those in Table 2 than with the raw data. In relation to seroreversion, the log-transformed data fitting is always better than the raw data, but the improvement is very little. However, as for survival, only for the fitting related to males showed a little improvement. Note that the variance obtained from the log-transformed data is lower than that obtained from the raw data. We observe that the parametrized surviving half-age estimate is smaller than the nonparametric estimate, while in relation to the seroreversion half-age, it is larger.

Second, we will comment the Tables 3 and 4 accompanied by figures obtained from the fittings.

With respect to the seroreversion phenomenon, the estimated age at which the seroreverting probability is equal to one half is only slightly smaller for males (11.87) than females (12.07), contrary to the Kaplan-Meier estimates. The phenomenon-gauge for seroreversion are (in months) $\beta = 4.10$ for both genders, $\beta = 4.31$ for females and $\beta = 3.88$ for males. There is no statistical evidence of difference between males and females (for both phenomena and for both parameters, the p-values are larger than 0.50). Since the estimates are very close and the Kaplan-Meyer estimates for male and female are also very close, we will treat them as a unique group.

Figure 2 shows the probability of keeping the initial (seropositive) state, which shows a very good fitting.

This figure shows the estimated probability function with and without the log-transformation of the data. In relation to the raw data, in the earlier ages, the 'shoulder' is less prominent than that observed in the nonparametrized estimate. On the other hand, the logarithm transformation, which privileges during the estimation of the lower ages, the shoulder is more pronounced in this estimate. The fitting provided by the log-transformed data is practically coincident with that provided by the Kaplan-Meier estimation.

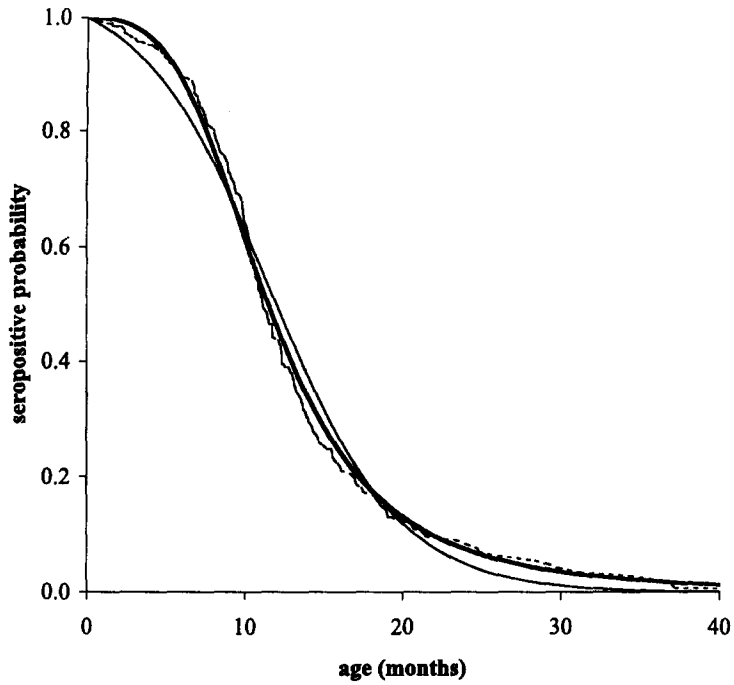


Figure 2. The parametrized estimation of the serorevertion phenomenon, using the raw data (thin curve) and the log-transformed, $\ln(a+1)$, data (thick curve). The Kaplan-Meier estimation (dotted curve), and the curves are irrespective of gender.

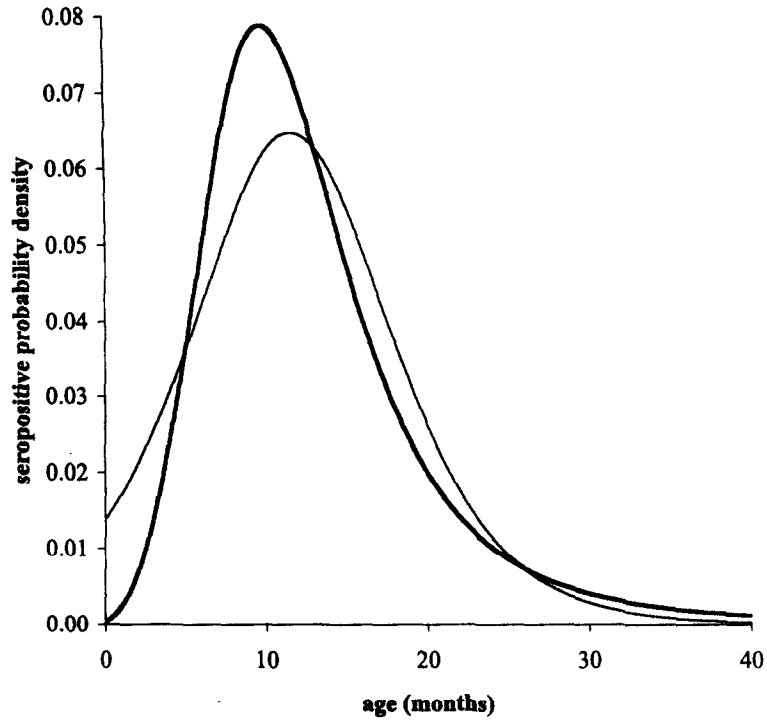
Figure 3 shows the curves of the probability density of transition and the hazard function related to the serorevertion.

Figure 3a shows the probability density of transition related to the serorevertion. We note that the mode is around the half-age with respect to the raw and the log-transformed data, the half-age being lower for the last case. Figure 3b shows the hazard function related to the serorevertion. Observe that the hazard increases monotonically with age when dealing with the raw data, however, the hazard function decreases for higher ages when dealing with the log-transformed data. In this case, the mode is around 15 months. At age zero, we have near zero value for the probability density of transition and zero value for the hazard function.

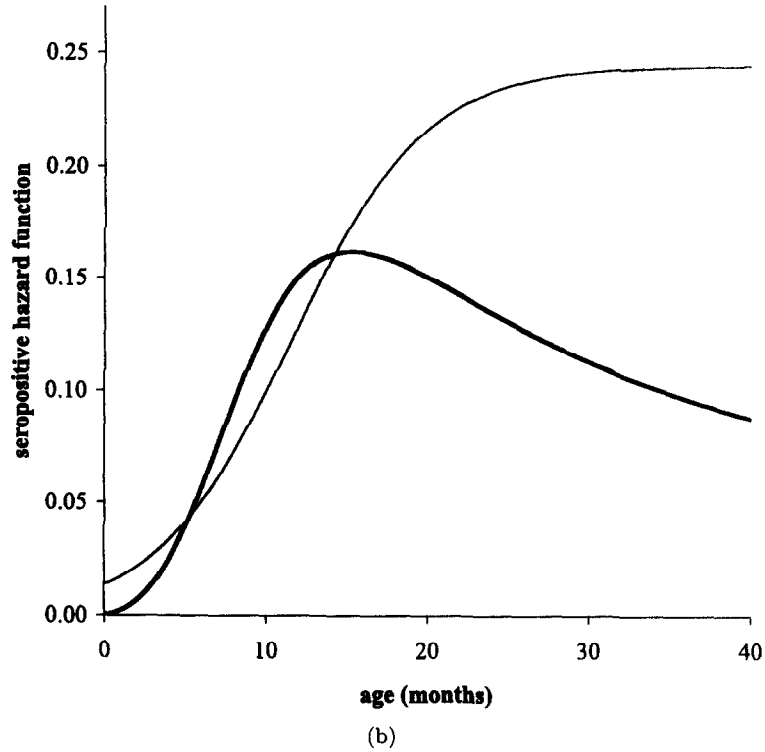
Comparing the fitting obtained from the raw and the log-transformed data with respect to the derived functions given by Figure 3, we observe that the log-transformed data give better estimation. The first improvement is in the probability density of transition, where a very low value is assigned at birth (Figure 3a). The value of transition density around 0.015 assigned by the raw data to the newborns is biologically unreasonable. The other improvement is in the hazard function, where we observe a peak at around 15 months [18], and further decreasing. Biologically, since we are treating the degradation of antibodies against HIV, it is reasonable to find lower hazard values for higher ages. On the other hand, we can use 15 months (the mode) as the censoring age, instead of 18 months [20]. Hence, a child can be considered uninfected by HIV if antibodies against HIV is absent at this time of censorship.

With respect to the survival phenomenon, we find some statistical evidence of the influence of gender. For the raw data, the p-values for the difference between the estimates for males and females are equal to 0.063 for τ and 0.069 for ξ, while for the log-transformed data, the p-values are 0.065 and 0.16 for τ and ξ, respectively. Comparing the results given by Tables 2–4, we can observe that the half-age is approximately 28 months less for girls than for boys and that the boys are more homogeneous than girls. The half-age estimated with the log-transformation is closer to the nonparametric Kaplan-Meyer estimates.

We show the results related to survival probability in Figure 4. Figure 4a is related to the raw data and Figure 4b is related to the log-transformed data.



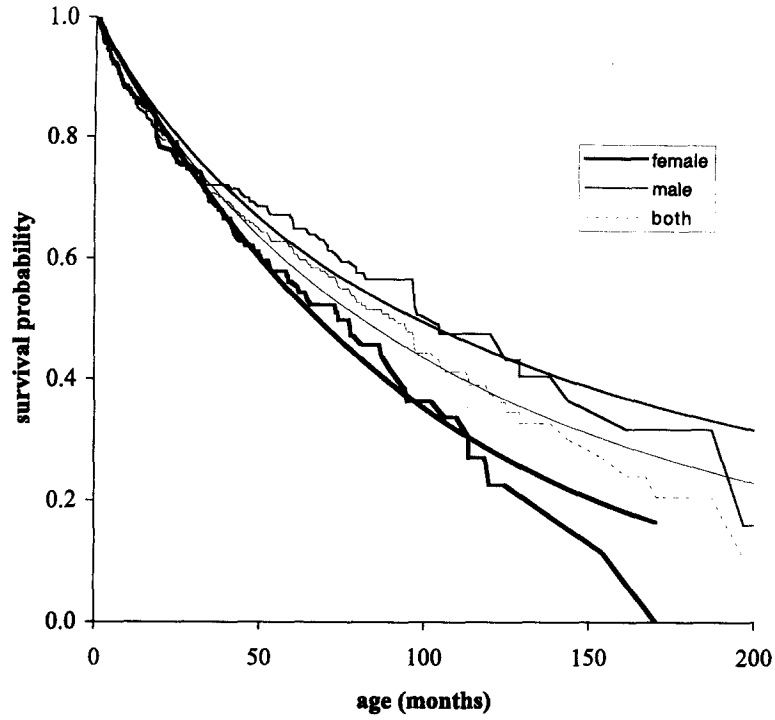
(a)



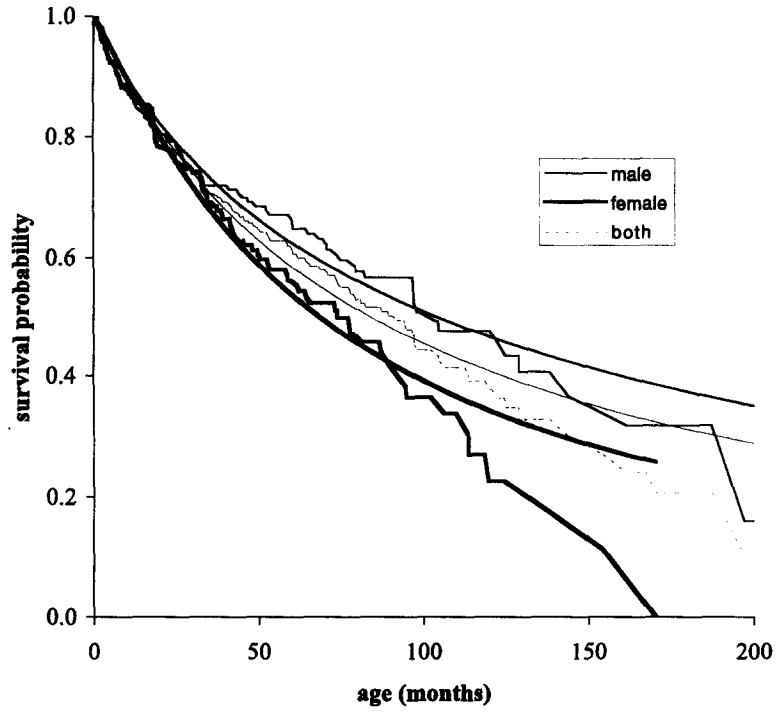
(b)

Figure 3. The probability density of transition (a) and the hazard function (b) related to Figure 2, using the raw data (thin curve) and the log-transformed, $\ln(a+1)$, data (thick curve). The curves are irrespective of gender.

The fitting gives higher probability of surviving for male children than females. We observe that, in comparison with the Kaplan-Meier estimation, at lower ages the log-transformed data fits better than the raw data, which is not true at higher ages. The different behaviour with gender can be better observed in the probability density of transition and the hazard function.



(a)

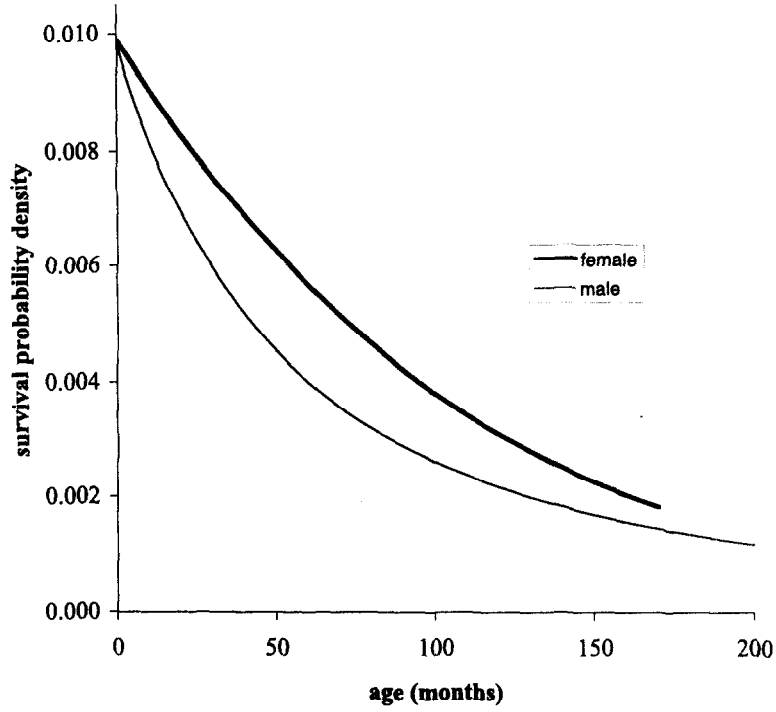


(b)

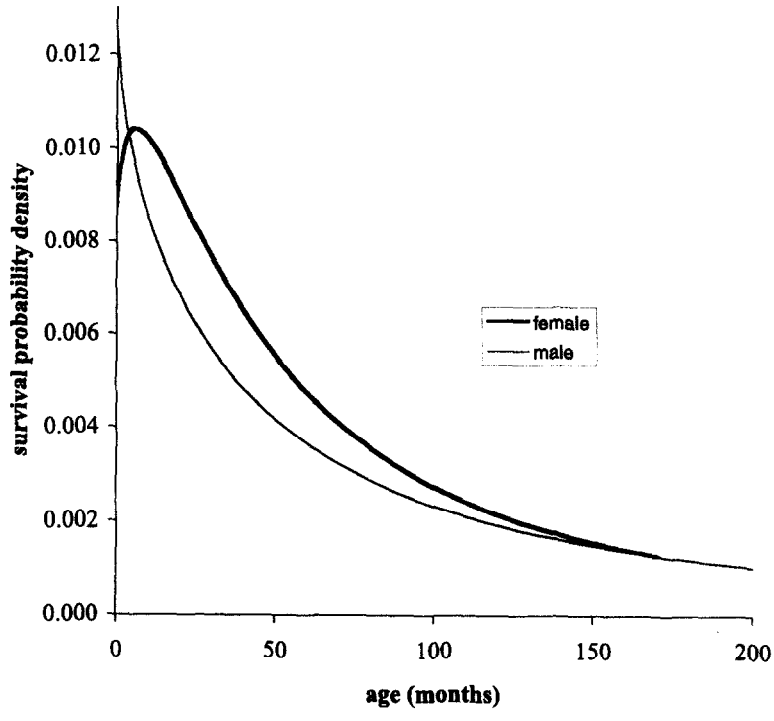
Figure 4. The parametrized estimation of the seroreversion phenomenon, using the raw data (a) and the log-transformed, $\ln(a+1)$, data (b). We also show the Kaplan-Meier estimation.

In Figure 5, we present the probability density of transition related to survival. Figure 5a is related to the raw data and Figure 5b is related to the log-transformed data.

Figures 5a and 5b show monotonically decreasing functions, except for the female in Figure 5b, which shows an initial increasing. Also we observe, in general, that female children are under higher probability density to death due to AIDS than male children.



(a)

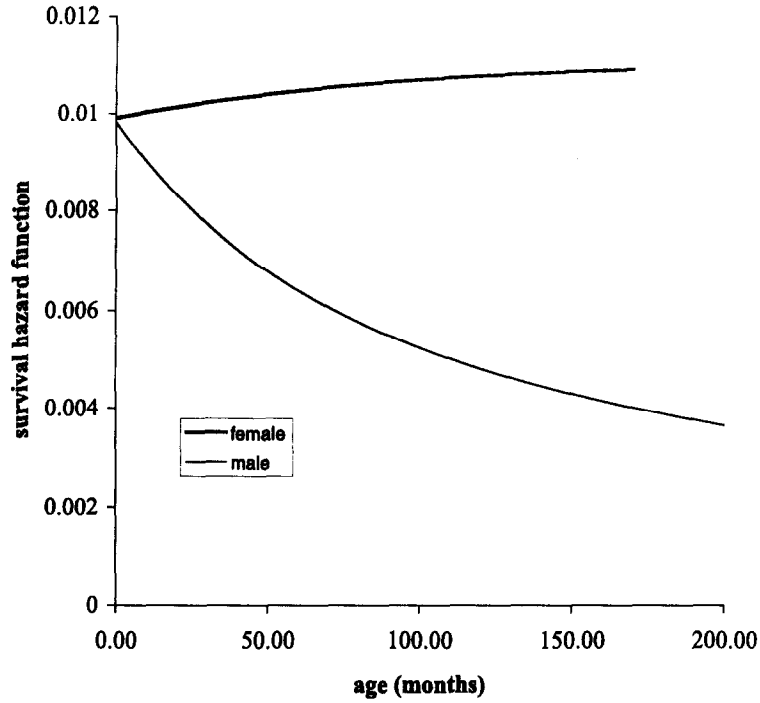


(b)

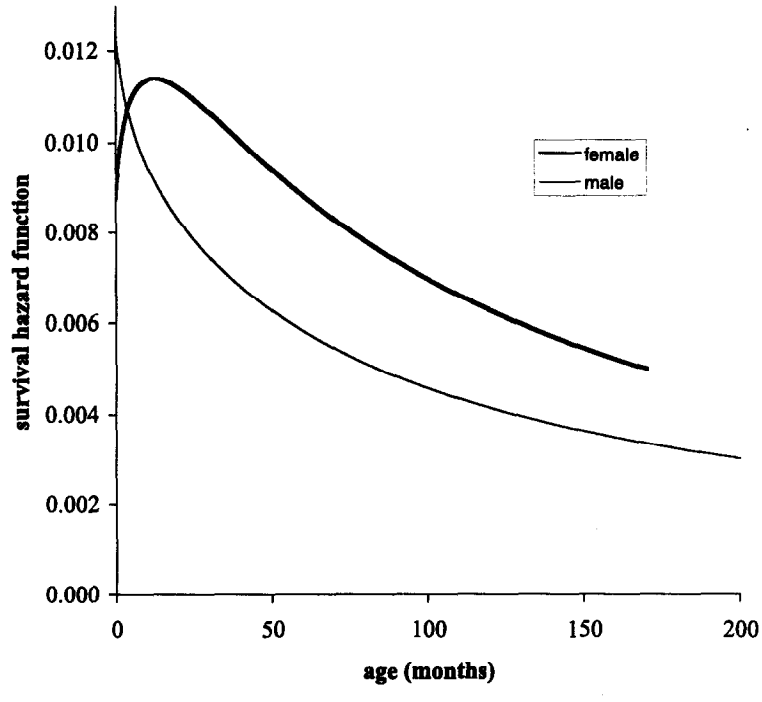
Figure 5. The probability density of transition related to Figure 4, using the raw data (a) and the log-transformed, $\ln(a+1)$, data (b).

In Figure 6, we present the hazard function related to survival. Figure 6a is related to the raw data and Figure 6b is related to the log-transformed data.

Figure 6a shows a monotonically increasing hazard function to female children when considering raw data. Biologically, this increasing hazard of dying at higher ages is very reasonable. Again, we observe that the female children are under an increased hazard of dying than male children.



(a)



(b)

Figure 6. The hazard function related to Figure 4, using the raw data (a) and the log-transformed, $\ln(a+)$, data (b).

Using the same probability function, we estimated two distinct phenomena. In relation to serorevertion, the results provided by parametrized estimation are very close to those provided by the Kaplan-Meier estimation. This is not observed with respect to survival phenomenon. The parametrized estimation comes close to the Kaplan-Meier estimation up to 120 months, afterwards we have a great discrepancy. Note that the similarity of both estimations with respect to male children increases until 135 months. We explain the difference between the Kaplan-Meier

and the parametrized estimations taking into account the sample size. The sample size related to male comprises 194 children under 135 months, and above this age, 11 children. In relation to female, the sample size is 179 children under 120 months, and above this age, five children.

With respect to the estimation, we observed the same feature given by the Kaplan-Meier estimation, that is, the results provided by the parametric estimation show that the seroreverion is not influenced by gender, however, the survival is strongly dependent on gender. Biologically, the degradation of the antibodies appears to be independent of the characteristics of the individuals, which is not true with respect to survival. For this reason, gender influences strongly the survival, but does not matter in relation to the seroreverion.

4. DISCUSSION

We developed a statistical model to deal with the data from a retrospective collection of followed-up children. The data consisted of observing the age at which children seroreverted or died due to HIV infection. The model was developed based on generally unobserved biological aspects, such as the threshold level of anti-HIV antibodies and the minimum amount of T-lymphocyte cells, that marks the exact moment of the transition from the initial to final states. The resulting model was structured in two parameters, the half-age and the phenomenon-gauge. Both parameters of the model were of easy interpretation, according to the biological findings.

Initially, both phenomena, the seroreverion and the survival, were treated in terms of the Kaplan-Meier estimation method. The results provided by this method were taken as the gold standard values.

The model, even though being very simple with only two parameters, fitted nicely to the data set of two distinct phenomena: seroreverion and survival. The fitted curves (reversed sigmoidal and exponentially decaying shapes for, respectively, seroreverion and survival phenomena) showed the appropriateness of the model to describe both phenomena. The different shapes were obtained due to the scaling factor, called the parameter of phenomenon-gauge. This survival distribution function seems to produce the same shape with respect to Gamma distribution or log-normal distribution. However, from this model, we can do a direct comparison with the Fermi-Dirac distribution, which provided the interpretation for the phenomenon-gauge regarded to intra-phenomenon.

As for the estimated parameters, female children, in comparison with males, seem to show a slower decay of maternally derived antibodies, but less resistance against HIV infection, which leads quickly to death due to HIV/AIDS. Also, we observed that the male children are much more heterogeneous than females with respect to survival. However, male children behave slightly more homogeneously than females with respect to seroreverion. If we take into account the broadly dispersed age of the followed-up children and use the log-transformation, we observe quite the same set of estimated parameters obtained with the raw data.

As for the variance of the estimated parameters, we observed very large values for the survival study, while for the seroreverion study, they are very small (see Table 3). These results are biologically reasonable. For the seroreverion, we have the degradation of the maternally derived anti-HIV antibodies circulating in the bloodstream of the child. Therefore, this degradation must be very insensitive to the different physiological conditions of the children. However, the interaction between HIV and the immunological system of the children must vary broadly, depending upon many features. Some of them must be the nutritional status, the genetic dependency of the immune system, and the genetic constitution of each individual.

By comparing the parametrized estimation with those provided by the nonparametrized method, we observe that the Kaplan-Meier estimates are generally larger than the estimates given by our model. A possible reason could be the censored data in the Kaplan-Meier method, which, in the maximum likelihood method, were assigned by a probability density function.

With respect to the log-transformation of collected data, we observed very slight improvement of the estimation, in relation to the raw data. However, this transformation "homogenized" the large range of occurrence of the phenomenon, and for this reason, the variance related to half-age was greatly diminished.

APPENDIX

The likelihood estimation method [21,22] uses the logarithm of the likelihood function, given by equation (12). To perform this estimation, we need the initial guess. This can be provided by a Kaplan-Meier estimator.

The estimator $\hat{\mathbf{\Omega}} = [\hat{\tau}\hat{\xi}]^{\top}$ provided by the Kaplan-Meier is used as the initial guess in the maximum likelihood estimation method with the logarithm of the likelihood function (12). This expression maximizes at

$$\mathbf{y}(\mathbf{\Omega}) = \frac{\partial}{\partial \mathbf{\Omega}} l(\mathbf{\Omega}) = \sum_t \frac{1}{f_A(a_t, \mathbf{\Omega})} \frac{\partial}{\partial \mathbf{\Omega}} f_A(a_t, \mathbf{\Omega}) + \sum_c \frac{1}{S_A(a_c, \mathbf{\Omega})} \frac{\partial}{\partial \mathbf{\Omega}} S_A(a_i, \mathbf{\Omega}) = \mathbf{0}, \tag{13}$$

because the inverse of the covariance matrix, neglecting the second derivatives of the likelihood function in relation to the parameters given by

$$\mathbf{\Sigma}^{-2}(\mathbf{\Omega}) = -\frac{\partial^2}{\partial \mathbf{\Omega}^2} l(\mathbf{\Omega}) \sim \sum_t \left[\frac{\frac{\partial}{\partial \mathbf{\Omega}} f_A(a_t, \mathbf{\Omega})}{f_A(a_t, \mathbf{\Omega})} \right]^2 + \sum_c \left[\frac{\frac{\partial}{\partial \mathbf{\Omega}} S_A(a_i, \mathbf{\Omega})}{S_A(a_c, \mathbf{\Omega})} \right]^2 \tag{14}$$

has negative value. The estimator that obeys (13), $\hat{\mathbf{\Omega}}$, is the value searched.

Due to the approximation in the second derivative, the likelihood estimation is obtained by the Levenberg-Marquardt nonlinear fitting method. This method is the modified Newton-Raphson method, where the increments in the new set of parameters are given by

$$\mathbf{\Sigma}_{LM}^{-2}(\mathbf{c}) = \begin{cases} \boldsymbol{\sigma}^{-2}(\mathbf{\Omega})(1+\varepsilon), & \text{on the diagonal,} \\ \boldsymbol{\eta}^{-2}(\mathbf{\Omega}), & \text{off the diagonal,} \end{cases} \tag{15}$$

where $\boldsymbol{\sigma}^2$ and $\boldsymbol{\eta}^2$ are, respectively, the variance and covariance of the matrix (14), and ε is an auxiliary parameter.

The first derivative of the likelihood function l with respect to the parameters are given by

$$\begin{aligned} \frac{\partial}{\partial \tau} l &= \sum_t \frac{1}{f_A(a_t)} \frac{\partial}{\partial \tau} f_A(a_t) + \sum_c \frac{1}{S_A(a_c)} \frac{\partial}{\partial \tau} S_A(a_t), \\ \frac{\partial}{\partial \xi} l &= \sum_t \frac{1}{f_A(a_t)} \frac{\partial}{\partial \xi} f_A(a_t) + \sum_c \frac{1}{S_A(a_c)} \frac{\partial}{\partial \xi} S_A(a_t), \end{aligned} \tag{16}$$

where

$$\begin{aligned} \frac{\partial}{\partial \tau} S_A(a) &= \frac{\xi e^{\xi(a-\tau)} \left(1 - e^{-\xi a}\right)}{\left(1 - e^{-\xi\tau}\right)^2} S_A^2(a), \\ \frac{\partial}{\partial \xi} S_A(a) &= \frac{\left(a - \tau - a e^{-\xi\tau}\right) e^{\xi(a-\tau)} + \tau e^{-\xi\tau}}{\left(1 - e^{-\xi\tau}\right)^2} S_A^2(a), \\ \frac{\partial}{\partial \tau} f_A(a) &= -\frac{\xi e^{\xi(a-\tau)} S_A(a)}{\left(1 - e^{-\xi\tau}\right)} \left[\frac{\xi S_A(a)}{1 - e^{-\xi\tau}} - 2 \frac{\partial}{\partial \tau} S_A(a) \right], \\ \frac{\partial}{\partial \xi} f_A(a) &= \frac{e^{\xi(a-\tau)} S_A(a)}{\left(1 - e^{-\xi\tau}\right)} \left[\frac{\left[1 + \xi(a-\tau) - (1 + \xi a) e^{-\xi\tau}\right] S_A(a)}{1 - e^{-\xi\tau}} + 2\xi \frac{\partial}{\partial \tau} S_A(a) \right]. \end{aligned} \tag{17}$$

Observe that when we use the logarithm of age, we have the change on the function $f_A(a')$, given by

$$f_A[\ln a'] = \frac{\xi e^{\xi \ln(a')}}{a' \left(e^{\xi\tau'} - 1\right) \left[1 + \left(e^{\xi \ln(a')} - 1\right) / \left(e^{\xi\tau'} - 1\right)\right]^2}, \tag{18}$$

where $a' = a + 1$, and we have the relation

$$\tau = e^{\tau'} - 1 \tag{19}$$

between τ' and τ.

REFERENCES

1. R.M. Anderson and R.M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford Science Publications, Oxford, (1992).
2. E.M. Connor, A.B. Minnefor and J.M. Oleske, Human immunodeficiency virus infection in infants and children, In *Current Topics in AIDS*, (Edited by M.S. Gottlieb *et al.*), pp. 185–210, John Wiley & Sons, New York, (1987).
3. Centers for Disease Control, Recommendations of the use of The U.S. Public Health Service Task Force on the use of zidovudine to reduce perinatal transmission of HIV, *MMWR* **43** (RR-11) (1994).
4. L.M. Mofenson, Epidemiology and determinants of vertical HIV transmission, *Seminars in Pediatric Infectious Disease* **5** (4), 252–265 (1994).
5. M.F. Rogers, L.M. Mofenson and R.R. Moseley, Reducing the risk of perinatal HIV transmission through zidovudine therapy: Treatment recommendations and implications, *JAMWA* **50** (3/4) (1995).
6. J.J. Goedert, A.M. Duliège, C.I. Amos, S. Felton, R.J. Biggar and The International Registry of HIV-Exposed Twins, High Risk of HIV-1 infection for first-born twins, *Lancet* **338**, 1471–1475 (1991).
7. Ministério da Saúde, Brazil, *Boletim Epidemiológico AIDS* **VIII** (3) (1995).
8. L.A. Guay, P. Musoke, T. Fleming, D. Bagenda, M. Allen, C. Nakabiito, J. Sherman, *et al.*, Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial, *Lancet* **354**, 795–802 (1999).
9. R.S. Azevedo Neto, A.S.B. Silveira, D.J. Nokes, H.M. Yang, S.D. Passos, M.R.A. Cardozo and E Massad, Rubella seroepidemiology in a non-immunized population of São Paulo State, Brazil, *Epidemiol. Infect.* **113** (1), 161–173 (1994).
10. R.M. Anderson and R.M. May, Herd immunity to Helminth infection and implications for parasite control, *Nature* **315** (6), 493–496 (1985).
11. A.D. Barbour, Schistosomiasis, In *Population Dynamics of Infectious Diseases*, (Edited by R.M. Anderson), Chapman & Hall, London, (1982).
12. G.P. Garnett and R.M. Anderson, Factors controlling the spread of HIV in heterosexual communities in developing countries: Patterns of mixing between different age and sexual activity classes, *Phil. Trans. R. Soc. London* **B 342**, 137–159 (1993).
13. D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes*, Chapman & Hall, London, (1992).
14. C.B. Williams, The use of logarithms in the interpretation of certain entomological problems, *Ann. Appl. Biol.* **24**, 404–414 (1937).
15. A. Stuart and J.K. Ord, *Kendall's Advanced Theory of Statistics, Volume 1*, Fifth Edition, Oxford University Press, New York, (1989).
16. D.R. Cox and D. Oakes, Analysis of survival data, In *Monographs on Statistics and Applied Probability, Volume 21*, (Edited by D.R. Cox *et al.*), Chapman & Hall, London, (1990).
17. A. Messiah, *Quantum Mechanics*, North-Holland, Amsterdam, (1958).
18. W.Y. Tsai, J.J. Goedert, J. Orazem, S.H. Landesman, *et al.*, A nonparametric analysis of the transmission rate of human immunodeficiency virus from mother to infant, *Biometrics* **50**, 1015–1028 (1994).
19. H.A. Kahn and C.T. Sempos, Statistical methods in epidemiology, In *Monog. Epidem. Biostat., Volume 12*, Oxford University Press, New York, (1987).
20. J.J. Goedert, H. Mendez and J.E. Drummond, Mother-to-infant transmission of human immunodeficiency virus Type 1: Association with prematurity or low anti-gp, *Lancet 334* **120**, 1351–1354 (1989).
21. H.M. Yang, Directly transmitted infections modeling considering age-structured contact rate—Epidemiological analysis, *Mathl. Comput. Modelling* **29** (7), 11–30 (1999).
22. H.M. Yang and F.A.B. Coutinho, Acquired immunity on a schistosomiasis transmission model—Analysis of the stabilizing effects, *J. Theoret. Biol.* **196**, 473–482 (1999).