THE STAGES OF HIV INFECTION: WAITING TIMES AND INFECTION TRANSMISSION PROBABILITIES

Ira M. Longini, Jr. W. Scott Clark, Michael Haber
Dept. of Epidemiology and Biostatistics
Emory University
Atlanta, GA 30322

Robert Horsburgh, Jr.
AIDS Program, Centers
for Disease Control
Atlanta, GA 30333

Abstract

We use stochastic models to estimate the waiting time distributions for progressive stages of HIV infection and to estimate stage-specific infection transmission probabilities between exposed and infected persons. We partition the infection period into four progressive stages: 1. infected but antibody negative; 2. antibody positive but asymptomatic; 3. pre-AIDS symptoms and/or abnormal hematologic indicator; and 4. clinical AIDS. We also define a fifth stage, death due to AIDS. A time-dependent Markov model is fitted to data on 45 persons with known time of HIV exposure to estimate the waiting time in stage 1, i.e., the pre-HIV-antibody period. The mean pre-HIV-antibody period is estimated to be 2.6 ± 0.2 months. A time-homogeneous Markov model was previously used to estimate the waiting time distributions for stages 2-3 from data on a sample of 513 homosexual and bisexual men from San Francisco. By combining this model with the model for stage 1, we estimate the mean AIDS incubation period, i.e., the waiting time in stages 1-3, as 9.8 ± 0.7 years. The estimated mean HIV infectious period, i.e., the waiting time in stages 1-4, is 11.8 ± 0.8 years. The estimated AIDS incubation period distribution is combined with a stochastic transmission model to estimate the stage-specific infection transmission rates from a sample of 45 heterosexual sex partners of persons with AIDS. The probability that an exposed person will be infected by a single sexual contact with an

C. Castillo-Chavez (ed.), *Mathematical and Statistical Approaches to AIDS Epidemiology*

infected person in stage 4 is estimated to be 0.0057 ± 0.0016. For stage 3, this probability is estimated to be 0.0007 ± 0.0002, and it is estimated to be very near to 0 for stage 2. Thus, an exposed person is about eight times more likely to be infected by a person who has AIDS than by one who has pre-AIDS symptoms
($p < 0.001$). The stochastic models that we developed can be extended to assess the risk of HIV transmission for different levels of important risk factors when more detailed data become available.

1. Introduction

The natural history of HIV in the human host can be visualized as a progression of stages from initial infection to death. In addition, the infectiousness of infected persons may vary considerably with respect to the stage of infection. These two important characteristics of HIV, coupled with the complicated social patterns of human contacts, largely determine the spread of HIV. The staging process and infectious contact patterns have been incorporated into dynamic mathematical models of HIV transmission to better understand HIV epidemiology and control (Jacquez et al. 1988; Koopman et al. 1988; Hethcote 1988). This paper presents some of the progress that has been made by using stochastic models of the staging process to estimate key parameters that govern the natural history and transmissibility of HIV.

Using the work of Longini et al. (1989), we model the natural history of HIV as a five-stage Markov process. The period of infection is partitioned into four transient states (stages) which are as follows: 1) infected but antibody negative; 2) antibody positive but asymptomatic; 3) pre-AIDS symptoms and/or abnormal hematologic indicator; and 4) clinical AIDS. The fifth stage is the absorbing state, death. Infected persons are modeled to flow irreversibly through the stages as shown in Figure 1. There are several waiting time distributions embedded in Figure 1 which are epidemiologically important. The waiting time in stage 1 is the pre-HIV-antibody period, which will be referred to as the pre-antibody period in this paper. The sum of the waiting times in stages 1-3 is the AIDS incubation period while the sum of the waiting times in stages 1-4 is the HIV infectious period as well as the time from initial

infection to death (i.e., HIV survival time). The AIDS survival time is the waiting time in stage 4.

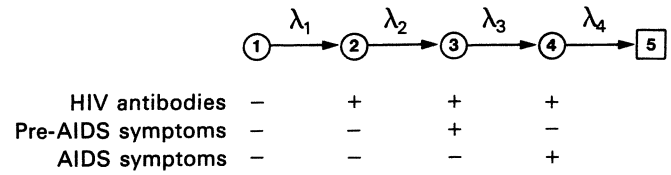

Figure 1: Flow diagram for the modeled natural history of HIV.

Section 2 of this paper deals with fitting a time-dependent, two-stage Markov model to data on newly infected persons in order to estimate the distribution of the pre-antibody period. In Section 3, this work is integrated with previous results of Longini et al. (1989) to provide estimates of the distributions of the AIDS incubation and HIV infectious periods. In Section 4, the estimated distribution of the AIDS incubation period is coupled with a stochastic model for estimating the stage-specific infectious transmission rates. We then apply the model to data collected by Fischl et al. (1987) on the sexual transmission of HIV between heterosexual partners. Finally, a discussion is given in Section 5.

2. The pre-antibody period

The pre-antibody period, i.e., waiting time in stage 1, was estimated from data on 45 persons who were infected with a known time of exposure to HIV (Horsburgh et al. 1989). The data, including the most likely route of exposure, are given below in Table 1. However, the route of exposure is not a factor in the analysis. Some of the persons were tested for infection just after the time of exposure and then again within six months. All persons were tested within six months of the time of exposure, and were followed for a relatively short period of time. Thus, they provide reliable information only about stage 1.

Table 1: Data used to estimate pre-antibody period

Person j	Time sequence τ_j months			States y_j			Route of exposure
1	0	2.53		1	2		Factor VIII
2	0	2.17		1	2		Factor VIII
3	0	3.52	7.50	1	1	2	Factor VIII
4	0	2.80	4.18	1	1	2	Factor VIII
5	0	0.79	2.86	1	1	2	Factor VIII
6	0	3.45	4.93	1	1	2	Factor VIII
7	0	2.96		1	2		Factor VIII
8	0	2.50		1	1		Factor VIII
9	0	5.10	6.91	1	1	2	Factor VIII
10	0	0.99	1.18	1	1	2	Factor VIII
11	0	0.89	3.85	1	1	2	Factor VIII
12	0	2.80		1	2		Factor VIII
13	0	0.66	2.20	1	1	2	Factor VIII
14	0	0.46	6.51	1	1	2	Factor VIII
15	0	2.73		1	2		Factor VIII
16	0	1.15	2.80	1	1	2	Factor VIII
17	0	1.22	1.45	1	1	2	Factor VIII
18	0	1.22	1.45	1	1	2	Factor VIII
19	0	0.76	1.15	1	1	2	Blood Transfusion
20	0	0.86	1.32	1	1	2	Blood Transfusion
21	0	0.29	1.48	1	1	2	Blood Transfusion
22	0	0.59	1.51	1	1	2	Blood Transfusion
23	0	0.43	3.82	1	1	2	Blood Transfusion
24	0	0.92	3.22	1	1	2	Blood Transfusion
25	0	2.37	3.45	1	1	2	Blood Transfusion
26	0	0.30	6.05	1	1	2	Needlestick
27	0	1.22	3.82	1	1	2	Needlestick
28	0	0.03	3.98	1	1	2	Needlestick
29	0	0.89	1.61	1	1	2	Needlestick
30	0	1.91	5.92	1	1	2	Needlestick
31	0	0.43	2.24	1	1	2	Needlestick
32	0	2.96	5.92	1	1	2	Needlestick
33	0	1.64	2.47	1	1	2	Organ Transplant
34	0	0.99	2.63	1	1	2	Organ Transplant
35	0	1.32	1.84	1	1	2	Organ Transplant
36	0	1.64		1	2		Organ Transplant
37	0	1.38		1	2		Organ Transplant
38	0	1.32	7.99	1	1	2	Organ Transplant
39	0	1.32	7.99	1	1	2	Organ Transplant
40	0	1.84	9.05	1	1	2	Cutaneous
41	0	1.38	4.01	1	1	2	Cutaneous
42	0	0.69	1.41	1	1	2	Cutaneous
43	0	0.53	1.55	1	1	2	Sexual
44	0	1.61	2.07	1	1	2	Sexual
45	0	1.05	2.96	1	1	2	IV drug use

Because the time of exposure is known for these 45 persons, the transition intensity (i.e., hazard function) for stage 1, $h_1(t)$ can be modeled as a time-dependent function, where $h_1(t)\Delta t + o(\Delta t)$ is the probability that a person who has been in stage 1 for t units of time will make the transition to stage 2 during the time interval $[t, t+\Delta t]$, and where $\lim o(\Delta t)/\Delta t = 0$ as $\Delta t \longrightarrow 0$.

We selected a three-parameter Weibull distribution to model the waiting time in stage 1. This distribution can have an increasing, decreasing, or constant hazard function depending upon the value of the shape parameter. Thus, the Weibull distribution should be sufficiently flexible to model the waiting time distribution in stage 1. In addition, the exponential distribution, which we previously used to model the waiting time distribution for stage 1 (see Longini et al. 1989), is a special case of the Weibull distribution. The probability density function (pdf) and the transition intensity (hazard) function, respectively, of the three-parameter Weibull distribution are

$$f_1(t) = \begin{cases} 0 & \text{if } t \leq \delta \ , \\ \lambda_1^{\alpha}\, \alpha[(t-\delta)^{\alpha-1}] \exp\{-[\lambda_1(t-\delta)]^{\alpha}\} & \text{if } t > \delta \ , \end{cases} \tag{1}$$

$$h_1(t) = \begin{cases} 0 & \text{if } t \leq \delta \ , \\ \lambda_1^{\alpha}\, \alpha(t-\delta)^{\alpha-1} & \text{if } t > \delta \ , \end{cases}$$

where $\lambda_1 > 0$, $\alpha > 0$ and $\delta \geq 0$. The time delay δ is needed since a finite amount of time is required for a primary antibody response (i.e., it is biologically impossible for a person to seroconvert immediately after infection). When $\alpha = 1$, then (1) reduces to a two-parameter exponential distribution.

Let j (j = 1,2,...,45) be the index for each of the persons in the cohort, and let m_j be the number of times person j was observed. Then the array $\boldsymbol{\tau}_j = (\tau_{j0}, \tau_{j1}, \ldots, \tau_{jm_j})$ represents the times at which person j was observed to be in the stages given by the array $\mathbf{y}_j = (y_{j0}, y_{j1}, \ldots, y_{jm_j})$, where $\tau_{j0} = 0$ and $y_{j0} = 1$ for all j. Since the mean waiting time in stage 1 (about 2.5 months) is much shorter than that in stage 2 (about 53

months--see Section 3), and the period of observation for each person is short, we modeled the system as simply two stages with stage 2 being closed. Given the above assumptions, we are interested in the probability, $p_j(\boldsymbol{\tau}_j;\mathbf{y}_j)$, of observing the array $\mathbf{y}_j$ at times $\boldsymbol{\tau}_j$. Table 1 presents the data for the arrays $(\boldsymbol{\tau}_j,\mathbf{y}_j)$, where the time units are months.

The following array patterns were observed in the data: $(0,\tau_{j1};1,1)$, $(0,\tau_{j1};1,2)$ and $(0,\tau_{j1},\tau_{j2};1,1,2)$. The probabilities of these patterns are given below in terms of $h_1(t)$ which is defined in (1):

$$p_j(0,\tau_{j1};1,1) = \begin{cases} 1 & , \quad \text{if } \tau_{j1} \leq \delta, \\ \exp[-h_1(\tau_{j1}-\delta)], & \text{if } \tau_{j1} > \delta, \end{cases} \tag{2}$$

$$p_j(0,\tau_{j1};1,2) = 1 - p_j(0,\tau_{j1};1,1) \ , \tag{3}$$

$$p_j(0,\tau_{j1},\tau_{j2};1,1,2) = \begin{cases} 0 \quad , & \text{if } \tau_{j1} \leq \delta, \ \tau_{j2} \leq \delta, \\ 1-\exp[-h_1(\tau_{j2}-\delta)], & \text{if } \tau_{j1} \leq \delta, \ \tau_{j2} > \delta, \\ \exp[-h_1(\tau_{j1}-\delta)] - \exp[-h_1(\tau_{j2}-\delta)] \ , & \\ & \text{if } \tau_{j1} > \delta, \ \tau_{j2} > \delta. \end{cases} \tag{4}$$

The likelihood function for estimating the parameters $\boldsymbol{\theta} = (\delta,\lambda,\alpha)$ was found by taking the product of the functions (2-4) for all 45 persons. It is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{45} p_j(\boldsymbol{\tau}_j;\ \mathbf{y}_j) \ . \tag{5}$$

Maximum likelihood estimates (MLEs) of the parameters, $\boldsymbol{\theta}$, were found by numerically maximizing the natural logarithm of (5) for λ_1 and α at fixed values of δ. This was accomplished with the derivative-free pseudo-Gauss-Newton algorithm in the BMDP statistical package (Ralston 1985). This algorithm also provided the asymptotic variance-covariance matrix of the conditional MLE's $\hat{\lambda}_1$ and $\hat{\alpha}$. Hypothesis tests were conducted using the likelihood ratio test (see Chapter 10 of Hogg and Craig, 1970).

The log-likelihood function was maximized at preset values of δ in the interval [0.00,1.14] using a step size of 0.01. Selected results are given in Table 2. When no delay is modeled, i.e., $\delta = 0.00$, the Weibull

distribution provides a significantly better fit to the data than does the exponential distribution (i.e., $\chi^2_{(1)} \cong 23$, $p < 0.001$, by the likelihood ratio test). However, from Table 2, the likelihood function is greatest for both the Weibull and exponential distributions at $\delta = 0.99$, but the log-likelihood functions were relatively flat around $\delta = 1$. Thus, we will use this value as our estimate. At this value of δ, the Weibull does not provide a significantly better fit to the data than does the exponential distribution (i.e., $\chi^2_{(1)} \cong 0.6$, $p \cong 0.55$, by the likelihood ratio test). The estimated value of α when $\delta = 1$ is 0.876 ± 0.152, which provides another check on the conjecture that the waiting-time distribution in stage 1 is a two-parameter exponential distribution. Thus, we select the two-parameter exponential distribution as the most parsimonious model for the data.

Table 2: Estimating the pre-antibody period: Maximized log-likelihood function values at preset values of δ

Delay δ Months	Maximized value of $\ell n\,[L(\boldsymbol{\theta})]$	
	Exponential $\alpha = 1$	Weibull $\alpha \neq 1$
0.00	-74.73	-63.16
0.40	-65.94	-60.79
0.80	-57.90	-57.54
0.90	-56.67	-56.67
0.98	-56.29	-56.10
0.99	-56.27	-55.99
1.00	-56.33	-55.99
1.01	-56.39	-55.99
1.02	-56.47	-56.00
1.10	-57.85	-56.33
1.14*	-60.15	-57.31

*For $\delta \geq \min\{\tau_{j2}\} = 1.15$ the value of $\ell n\,[L(\boldsymbol{\theta})] \longrightarrow -\infty$.

The estimate of λ_1 is 0.625 ± 0.081, conditional on $\delta = 1$. The mean waiting time in stage 1, i.e., pre-antibody period, is estimated to be $\hat{\mu}_1 = 1 + 1/\hat{\lambda}_1 = 2.6$ months. The estimated standard error of $\hat{\mu}_1$ is obtained using the method of statistical differentials (see pages 69-72 in Elandt-Johnson and Johnson, 1980). It is $\widehat{s.e.}(\hat{\mu}_1) \cong \widehat{s.e.}(\hat{\lambda}_1)/\hat{\lambda}_1^2 = 0.2$ months when δ is assumed constant. Thus, an approximate 95% confidence interval on the mean pre-antibody period, μ_1, is [2.2,3.0] months. The

median length of the pre-antibody period is estimated to be $\hat{\tilde{\mu}}_1 = 1 + (\ell n\ 2)/\hat{\lambda}_1 = 2.1$ months. The approximate 95% confidence interval on $\tilde{\mu}_1$ is [1.8,2.4] months. The estimated time at which 95% of the infected persons could be expected to seroconvert is $\hat{\tau}_{.95} = 1 - \ell n(.05)/\hat{\lambda}_1 = 5.8$ months, with an approximate standard error of 0.6 months. Then the approximate 95% confidence interval on $\tau_{.95}$ is [4.6,7.0] months.

3. The AIDS incubation period

The AIDS incubation period, i.e., waiting time in stages 1-3, was estimated by combining the results from Section 2 with data on a sample of 513 seropositive men who were from a larger cohort of homosexual and bisexual men from San Francisco (Jaffe et al. 1985). These men were periodically bled for serostatus and examined for pre-AIDS and AIDS indicators and symptoms. While the men in this study were presumably infected by sexual contact, the time of exposure was unknown. The time of seroconversion was observed to an interval for 90% of the men, while the remaining 10% entered the sample as seropositive for HIV. Thus, this cohort yielded information about the waiting times in stages 2-4, but not about stage 1. Of the 513 men included in the analysis, 130 (25%) had developed AIDS and 76 of these 130 (58%) had died, during the seven to eight years of follow-up.

A time-homogeneous, staged Markov model was fitted to these data by Longini et al. (1989) to estimate the waiting times in stages 1-4. In that analysis, they estimated the mean pre-antibody period to be 2.2 ± 0.7 months from a cohort of 90 persons with transfusion and factor VIII-associated infections. We now propose to use the model described in Section 2 to model stage 1 with the somewhat improved data given in table 1. For this model, the mean pre-antibody period is estimated to be 2.6 ± 0.2 months, as described above.

The progression of an infected person through the stages of infection and ultimately to death is modeled as a time-homogeneous Markov process with a time delay δ in the first stage. The transition intensities for stages 2-4 are the constants $\lambda_i > 0$, $i = 2,3,4$, and the transition intensity for stage 1 is zero, if $t \leq \delta$, and it is $\lambda_1 > 0$, if $t > \delta$. The probability that a person who is in stage i at time τ will be in stage $k \geq$

i at time $t > \tau$ is defined as $p_{ik}(\tau,t)$. Explicit formulae for this probability are easily found using standard methods for Markov processes (see Chapter 11.7 in Chiang 1980). If all the transition intensities are distinct, i.e., $\lambda_i \neq \lambda_j$ for all $i \neq j$, then the transition probabilities from stage 1, starting at time 0, to the other states are

$$p_{1k}(0,t) = \begin{cases} 0 & , \text{ if } t \leq \delta, \\ (-1)^{k-1}\lambda_1 \ldots \lambda_{k-1} \sum_{j=1}^{k} \exp[-\lambda_j (t-\delta)] \Big/ \prod_{\substack{\ell=1 \\ \ell \neq j}}^{k} (\lambda_j - \lambda_\ell) & , \text{ if } t > \delta, \end{cases} \tag{6}$$

$$k = 1,2,3,4,$$

$$p_{15}(0,t) = \begin{cases} 0 & , \text{ if } t \leq \delta \\ -\lambda_1 \ldots \lambda_4 \sum_{j=1}^{4} \{1-\exp[-\lambda_j (t-\delta)]\} / [\lambda_j \prod_{\substack{\ell=1 \\ \ell \neq j}}^{4} (\lambda_j - \lambda_\ell)], & \text{if } t > \delta. \end{cases} \tag{7}$$

Since the transitions among stages 2-4 are assumed to be time homogeneous, the transition probabilities among these stages for all $\tau > \delta$ and $t > \tau$ are

$$p_{ik}(\tau,t) = (-1)^{k-i} \lambda_i \ldots \lambda_{k-1} \sum_{j=i}^{k} \exp[-\lambda_j (t-\tau)] \Big/ \prod_{\substack{\ell=i \\ \ell \neq j}} (\lambda_j - \lambda_\ell), \tag{8}$$

$$i = 2,3,4;\ i \leq k \leq 4,$$

$$p_{i5}(\tau,t) = (-1)^{4-i} \lambda_i \ldots \lambda_4 \sum_{j=i}^{4} \{1-\exp[-\lambda_j (t-\tau)]\} / [\lambda_j \prod_{\substack{\ell=i \\ \ell \neq j}}^{4} (\lambda_j - \lambda_\ell)], \tag{9}$$

$$i = 2,3,4.$$

We define T_I as the random variable for the AIDS incubation period, i.e., the waiting time in stages 1-3. The probability that a person who was infected at time 0 is in stage 3 at time t, and then makes the transition to stage 4 at time t + dt is $\lambda_3 p_{13}(0,t)dt$, where $p_{13}(0,t)$ is given in (6). Thus, the pdf for T_I is $f_I(t) = \lambda_3 p_{13}(0,t)$. Note that $f_I(t)$ is a special case of the three-parameter general gamma distribution (see page 222 in Johnson and Kotz 1970). The hazard function, $h_I(t)$, for the AIDS incubation period is

$$h_I(t) = f_I(t)/[1 - F_I(t)], \tag{10}$$

where $F_I(t) = \int_0^t f_I(\omega)d\omega$ is the cumulative distribution function (cdf) of T_I.

It follows from the modeling assumptions that the mean waiting time in stage 1 is $\mu_1 = \delta + 1/\lambda_1$, and the mean waiting time in the other stages is $\mu_i = 1/\lambda_i$; i = 2,3,4. The median waiting time in stage 1 is $\tilde{\mu}_1 = \delta + \mu_1 \ell n2$, while it is $\tilde{\mu}_i = \mu_i \ell n2$ for i = 2,3,4. Then the expected length of the AIDS incubation period is

$$E(T_I) = \int_o^\infty \omega\, f_I(\omega)\, d\omega = \mu_1 + \mu_2 + \mu_3 \ , \tag{11}$$

and the variance of T_I is

$$Var(T_I) = 1/\lambda_1^2 + 1/\lambda_2^2 + 1/\lambda_3^2 \ . \tag{12}$$

We define T as the random variable for the length of the HIV infectious period, i.e., the waiting time in stages 1-4. (Note that T is also the time from infection to death.) If we use the same argument as the one above, then the pdf of T is $\lambda_4 p_{14}(0,t)$. This is a special case of the four-parameter general gamma distribution. The cdf of T is

$$F(t) = \int_0^t \lambda_4 p_{14}(0,\omega)\, d\omega, \tag{13}$$

and the expected length of the HIV infectious period is

$$E(T) = \mu_1 + \mu_2 + \mu_3 + \mu_4. \tag{14}$$

The parameters λ_i, i = 2,3,4, were estimated by Longini et al. (1989) by ML methods for Markov processes using the data on the 513 homosexual and bisexual men from the San Francisco study. The parameters δ and λ_1 were estimated as described in Section 2. The estimated parameters and mean and median waiting times are given in Table 3. The estimated mean

AIDS incubation period from (11) and Table 3 is 9.8 years (117.7 months) with a 95% confidence interval of [8.4,11.2] years. The asymptotic standard error of the estimated AIDS incubation period was obtained by the use of the method of statistical differentials (Elandt-Johnson and Johnson, 1980). Based on the model, the estimated median AIDS incubation period is 8.3 years (99.4 months). The estimated hazard function for progression to AIDS is monotonically increasing in t (see Fig. 2), which agrees with the form of the hazard functions used to model the AIDS incubation period by other investigators (see Longini et al. 1989 for a discussion of this point). Based on the estimated density function, the probability that a newly infected person will have developed AIDS within five years of infection is 0.26.

Table 3: Estimated parameters, mean and median waiting times in each stage of infection based on the staged Markov model with $\delta = 1$

Stage	Parameter estimate	Waiting time	
		Mean	Median
i	$\hat{\lambda}_i \pm$ one s.e.	$\hat{\mu}_i$	$\hat{\tilde{\mu}}_i$
	$mos.^{-1}$	mos.(yrs.)	mos.(yrs.)
1	0.625 ± 0.081	2.6 (0.2)	2.1 (0.2)
2	0.019 ± 0.002	52.6 (4.4)	36.5 (3.0)
3	0.016 ± 0.002	62.5 (5.2)	43.3 (3.6)
4	0.042 ± 0.004	23.8 (2.0)	16.5 (1.4)

The estimated parameters for stages 2, 3, and 4 are from Longini et al. (1989).

If we assume that persons remain infectious up to their time of death, the estimated mean HIV infectious period from (14) and Table 3 is 11.8 years (141.5 months) with a 95% confidence interval of [10.3,13.3] years. Based on the model, the estimated median HIV infectious period is 10.3 years (124.0 months). The estimated probability density and hazard functions have the same shape as those shown in Fig. 2. As mentioned above, the HIV infectious period is also the time to death from the beginning of stage 1, i.e., time of initial infection. Statistics concerning the survival time from each stage of infection are easily derived from the staged Markov process and the parameter estimates (see

Longini et al. 1989 for detailed estimates).

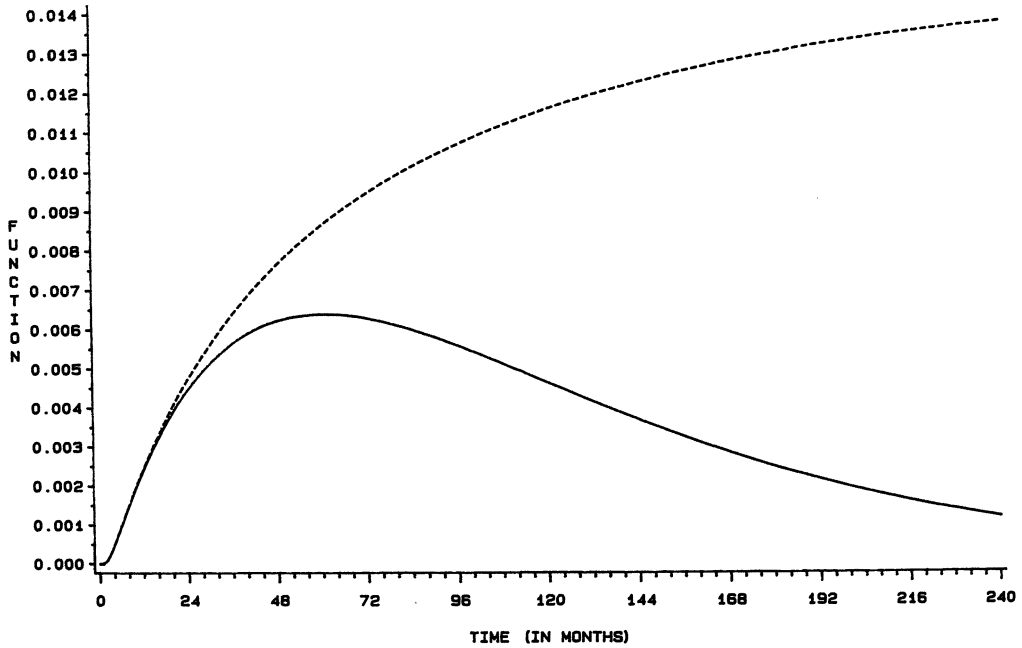

Figure 2: The estimated pdf $\hat{f}_I(t) = \hat{\lambda}_3 \hat{p}_{13}(0,t)$, solid line, and hazard function $\hat{h}_I(t)$ (see (10)), dashed line, of the AIDS incubation period, where the estimated parameters come from Table 3.

4. Stage-specific infection transmission probabilities

As shown in the above sections, HIV has a long, variable infectious period, and the level of infectiousness is thought to vary considerably as the disease progresses (Lange et al. 1986; Hyman and Stanley 1988; and Anderson and May 1988). A number of host factors may affect HIV

transmissibility, including concomitant venereal infections and T-helper cell numbers. However, it has been proposed that infectiousness increases with the stage of infection, i.e., progression (Burke and Redfield 1988; Goedert et al. 1987; Laga et al. 1988). With regard to the four stages of infection described above, it is possible that persons are highly infectious, on the average, during stages 1 and 4, not very infectious during stage 2, and infectious at an intermediate level, on the average, during stage 3 (Seage et al. 1989). In this section, we test this hypothesis by developing a statistical model for estimating the probability that a susceptible person is infected given that the susceptible is exposed to an infected person who is in a particular stage of infection. Such probabilities can best be estimated from prospective cohort studies where susceptibles are exposed to known infectives for specified lengths of time.

We will develop the model for the following study design which was employed by Fischl et al. (1987) to study the heterosexual transmission of HIV from persons diagnosed with AIDS to their sex partners. "Monogamous" sexual partners are ascertained when one partner (the index case) is diagnosed with AIDS. We will label this time of diagnosis as time 0. The exposed partner (to the index case) is then bled periodically at times $0, t_1, \ldots, t_k$, and the samples are examined for HIV. (We assume that the intervals $[t_j, t_{j+1}]$, and $j = 0, \ldots, k-1$ are wide enough so that if a person is infected during an interval, then seroconversion will probably occur during that same interval.) In addition, the amount of time that the partners had sexual contact prior to diagnosis of the index case is known and labeled as variable $u \geq 0$. Then, the period of exposure prior to AIDS diagnosis of the index is $\min\{T_I, u\}$, where T_I is the random variable for the length of the AIDS incubation period, with pdf $f_I(t)$ given in Section 3. In addition, the amount of time that the partners continue to have sexual relations after time 0 is known and labeled as w. Then the total period of exposure to the index case, when the infection status of the exposed person is known, is $\min\{T_I, u\} + w$. Fischl et al. (1987) assumed that the exposed sex partner in the pair did not have exposure to infected persons outside of the relationship, although there was evidence of limited outside exposure for some exposed partners. We use this assumption in our analysis, although the statistical model can be easily modified to incorporate exposure from outside of the relationship (see Longini et al, 1988).

The infection transmission rate is modeled to vary over the length of the infectious period. For exposure prior to time 0, the infectious transmission rate is $\beta(\tau,x)$, so that $\beta(\tau,x)d\tau + o(d\tau)$ is the probability that a susceptible will be infected during the interval of time $(\tau,\tau+d\tau)$, where τ is measured backwards from time 0, given exposure to an infected person who has an AIDS incubation period of length $T_I = x$. Since T_I is a random variable, it is unlikely that we would know precisely when the index was infected. Thus, we need to incorporate the distribution of T_I into our analysis. Then the cumulative infection transmission rate for exposure prior to time 0 is

$$\beta(\tau) = \int_0^\infty \beta(\tau,x)\ f_I(x)dx, \quad 0 \le \tau \le u. \tag{15}$$

For exposure after time 0, we let $\gamma(\tau)$ be the infection transmission rate for exposure τ time units after time 0. Then, the cumulative hazard function (for being infected) for an exposed person who has been exposed to the index for u time units before time 0 and w time units afterwards is

$$\Lambda(u,w) = \int_0^u \beta(\tau)d\tau + \int_0^w \gamma(\tau)d\tau. \tag{16}$$

It follows directly that the probability that an exposed person is still not infected at time w is $\exp[-\Lambda(u,w)]$. Additional risk due to exposure to infectives outside the relationship could be modeled by adding the appropriate terms to the cumulative hazard function (16).

The functions $\beta(\tau,x)$ and $\gamma(\tau)$ are assumed to have particular parametric forms, with parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. These parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\gamma})$, can be estimated using ML methods as long as they are identifiable. With the study design described above, three types of observations can be seen in the data. Each type is listed below with the appropriate contribution to the likelihood function:

1. If the exposed person is positive (for infection) at time 0, then the contribution is
   $$L_1(\boldsymbol{\theta}) = 1 - \exp[-\Lambda(u,0)]. \tag{17}$$
2. If the exposed person is not positive at time $t_{k-1} < w$ and is positive at time t_k, then the contribution is
   $$L_2(\boldsymbol{\theta}) = \exp[-\Lambda(u,t_{k-1})] - \exp[-\Lambda(u,s_k)], \tag{18}$$

where $s_k = \min(w, t_k)$.

3. If the exposed person is negative at the time of the last observation, t_k, then the contribution is

$$L_3(\theta) = \exp[-\Lambda(u, s_k)]. \tag{19}$$

The total likelihood function is the product of the appropriate contributions (17-19) for all the exposed persons in the data set. The total likelihood function is maximized using standard methods as described in Section 2.

Since the infectiousness of a person is thought to vary approximately with respect to the stages of infection described in Section 3, the function $\beta(\tau, x)$ is modeled as a step function with each step corresponding roughly to a stage of infection. The function is

$$\beta(\tau, x) = \begin{cases} \beta_3, & \text{if } 0 \le \tau \le x(1 - \delta_2), \\ \beta_2, & \text{if } x(1 - \delta_2) < \tau \le x(1 - \delta_1), \\ \beta_1, & \text{if } x(1 - \delta_1) < \tau \le x, \\ 0, & \text{if } \tau > x, \end{cases} \tag{20}$$

$\beta_1 \ge 0$, $\beta_2 \ge 0$, $\beta_3 \ge 0$,

where δ_1 is the fraction of the mean waiting time in stage 1 relative to the mean AIDS incubation period. Similarly, δ_2 is the fraction corresponding to the mean waiting times in stages 1 plus 2. We make no *a priori* assumptions about the ordering of the β_i's. From the data in Table 3, we estimate $\delta_1 = 2.6/117.7 = 0.022$ and $\delta_2 = (2.6 + 52.6)/117.7 = 0.469$. Thus, β_1, β_2, and β_3 correspond to the infection transmission rates with an infected person who is in stages 1, 2, and 3 of infection respectively. Since a person is in stage 4 of infection after AIDS diagnosis by definition, then the function $\gamma(\tau)$ is

$$\gamma(\tau) = \beta_4, \quad \beta_4 \ge 0, \text{ for } \tau \ge 0. \tag{21}$$

Thus, β_4 is the transmission rate for a person in stage 4 of infection.

Once the above parameters are estimated, two infection transmission probabilities of epidemiologic interest can be calculated. The first is the probability that a susceptible will be infected given t time units of exposure to an infective who is in stage i of infection, which is

$$P_i(t) = 1 - \exp(-\beta_i t), \qquad i = 1,2,3,4. \tag{22}$$

The second alternative is the probability that a susceptible will be infected in a single sex act with an infective who is in stage i of infection. In order to calculate this probability, we will assume that all couples have an average rate of r sexual contacts per unit of time. This probability is

$$p_i = 1 - \exp(-\beta_i/r), \qquad i = 1,2,3,4. \tag{23}$$

The model described by equations (15-19) was fitted to data on 45 heterosexual couples that were followed by Fischl et al.(1987), where the index case had AIDS. The exposed person was bled every six months up to 18 months after the AIDS diagnosis of the index case. Infection was identified by seropositivity to HIV. The complete data set is given in Table 1 of Fischl et al. (1987) and will not be given here. Of the 45 exposed persons, 13 were seropositive at time 0 (i.e., the time of AIDS diagnosis of the index case). Their contribution to the likelihood function is modeled by (17). In addition, 13 persons seroconverted after time 0, and their contribution to the likelihood function is modeled by (18). Finally, 19 exposed persons remained seronegative, and their contribution to the likelihood function is modeled by (19). Thus, 58% of the exposed persons were infected. Because stage 1 represents such a small proportion of the AIDS incubation period, 2.2%, we were unable to estimate β_1. Three different distributions were used to model the AIDS incubation period: 1) The three-parameter general gamma distribution used in Section 3, i.e., $f_I(t) = \lambda_3 p_{13}(0,t)$, 2) the two-parameter Weibull distribution, which has been used by a number of other researchers (see Longini et al. 1989), and 3) the degenerate (i.e., constant) distribution for the sake of comparison. The parameters were set so that the mean was the same for all three distributions, $E(T_I) = 118$ months. The parameter estimates were similar for all three distributions. Thus, only the results when the general gamma distribution is used are presented in Table 4.

The first row of Table 4 shows the results for a model that ignores the stages of infection and assumes a uniform transmission rate, β, through the entire infectious period. In this case, the probability of infection per sexual contact would be 0.0010. This probability was

calculated using a constant rate of r = 11 (mostly) unprotected sexual contacts per month, the average of the rate reported by Fischl et al. (1987). The value of the maximized log-likelihood function is -73.5. The next three rows in table 4 show the estimated parameters for the model with stage-specific transmission rates (20). The value of the maximized log-likelihood function for this model is -56.6. Thus, we reject the null hypothesis that $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$ ($p < 0.001$) when we use the likelihood ratio test (with 3 degrees of freedom), and we conclude that there is a significant difference in the infectiousness of an infected person in stages 2, 3, and 4. From Table 4, the infection transmission rate for exposure to an infected person in stage 2, i.e., the asymptomatic state, is estimated to be 0. However, this occurs because the estimate of β_2 hit the lower bound of 0 at the maximum of the likelihood function. Thus, the standard error for β_2 cannot be computed. We can only state that the transmission rate for stage 2 is probably close to 0 for this particular data set.

Table 4: Estimated infection transmission rates and probabilities

Stage	Transmission rate	Transmisson probability	
	$mo.^{-1}$	Per 12 mos.	Per sexual contact
i	$\hat{\beta}_i$	$\hat{P}_i(12)^*$	$\hat{p}_i^{**}$
no stages[†]	0.011 ± 0.002	0.124 ± 0.021	0.0010 ± 0.0002
2	0[††]	0[††]	0[††]
3	0.008 ± 0.002	0.093 ± 0.024	0.0007 ± 0.0002
4	0.061 ± 0.017	0.518 ± 0.101	0.0057 ± 0.0016

*Calculated from (22). **Calculated from (23) based on a constant rate of 11 sexual contacts per month (Fischl et al. 1987). [†]Fitted model with $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$. [††]Standard error not available.

From Table 4, the risk ratio per sexual contact for a susceptible when comparing exposure to an infected person in stage 4 with that in stage 3 is estimated to be $\hat{p}_4/\hat{p}_3 = 0.0057/0.0007 = 8.1$, which is significantly different from 1 ($p < 0.001$). Therefore, a person who has a single sexual contact with an infected person who has AIDS is eight times

more likely to be infected than if he or she has a single sexual contact with an infected person who has pre-AIDS symptoms. Table 4 also gives the risk of infection to a susceptible who has sex with an infected person over the period of one year. Note that the probability that a person would be infected after a year of sexual contact with a partner who has AIDS (stage 4) is estimated to be 0.518--a rather high probability. The average period of sexual contact following AIDS diagnosis was 11.2 months for the 45 couples in the data set.

5. Discussion

In this paper, we have used a staged Markov model to describe the transitions of infected persons from infection to AIDS diagnosis and ultimately, to death. The estimated transition intensity rates provide statistical estimates of the length of the pre-antibody period, AIDS incubation period, HIV infectious period, and the time to death (i.e., survival time) for infected persons. In addition, by combining the staging distribution with a non-homogeneous Poisson process for the risk of infection, we were able to estimate the stage-specific infection transmission rates for exposed susceptible persons. Thus, we were able to use the general concept of staging for HIV infection to construct a relatively complete statistical picture of the progression to disease for infected persons and the risk that such persons may pose to others through heterosexual contact.

The estimates of the transition intensities, $\lambda_1,\ldots,\lambda_4$, given in Table 3 and those of the infection transmission rates, β_3 and β_4, given in Table 4, have been used by Koopman et al. (1988) and Hethcote et al. (1988) in dynamic models of HIV transmission. These models have been used to investigate questions of HIV epidemiology and control as well as to make long-term projections for the course of the HIV epidemic in specific populations.

As pointed out in Section 1, the pre-antibody period (i.e., waiting time in stage 1) is important because persons can have infective virus in their blood during this period (Ward et al. 1988). Thus, they may be highly infectious to others at a time when they are antibody negative and difficult to detect as infected. Our estimate of 2.1 months for the median of the pre-antibody period is close to that estimated by Horsburgh,

et al. (1989) of 2.4 months from experimental data. That analysis used a staged Markov model for estimation similar to the model described in Section 2, and was based on data from pre- and post-seroconversion blood samples from 39 infected persons whose infection status was also assessed by detection of HIV DNA using the polymerase chain reaction technique. Both the above estimates are in general agreement with direct observations on newly infected persons (see Horsburgh et al. 1989 for details). Our estimate that 95% of infected persons would seroconvert within 5.7 months of exposure is consistent with the observations of Nekwei et al. (1988). They reported that 96% of the newly infected persons that they studied seroverted within six months.

A potential bias exists in the data in Table 1 because subjects in which HIV infection has persisted for several years without the detection of antibody may not yet have been reported. However, this is unlikely since infection without antibody detection has not been reported to persist for longer than 14 months, and no delays in the recognition of infection have been reported from prospective studies of exposed persons who have been followed for at least 24 months (see Horsburgh et al. 1989 for details).

With respect to the estimation procedure used in Section 2, there is no MLE for the time delay, δ, for right interval-censored data. In the case of no interval censoring, the MLE for δ is the minimum observed waiting time from infection to seroconversion (see page 245 in Kalbfleisch and Prentice 1980). Thus, it was necessary to find the MLEs for λ and α, conditioned on the preset values of δ. The estimated standard errors of $\hat{\lambda}$ and $\hat{\alpha}$ are probably underestimated since they do not reflect variation in the estimate of δ.

The interval censoring causes some inflation in the estimated standard error of $\hat{\lambda}_1$. In the case of no interval censoring, then we have $\widehat{\text{var}}(\hat{\lambda}_1) \cong \lambda_1 / \sum_{j=1}^{n} W_j$, where W_j is the time spent by person j in stage 1 (see chapter 11 in Chiang, 1980). The estimates of λ_1 and W_j from the interval-censored data in Table 1 are $\hat{\lambda}_1 = 0.625$ and $\hat{W}_j \cong 2.6$, $j = 1,\ldots,45$. If we assume that the data were not interval-censored, then $\widehat{\text{var}}(\hat{\lambda}_1) \cong (0.625)/[(45)(2.6)] = 0.0053$, and we have $\widehat{\text{s.e.}}(\hat{\lambda}_1) \cong \sqrt{0.0053} = 0.073$. This estimatated standard error is only slightly smaller than the estimate of 0.081 that was found for the

interval-censored data.

The most parsimonious model for the pre-antibody period is the two-parameter exponential distribution, which is time homogeneous except for the time delay, δ. This distribution provides as good an explanation of the data as the time-dependent, three-parameter Weibull distribution. The implication is that the hazard rate of transition from the antibody-negative to the antibody-positive stage of infection is probably constant over time following an initial time delay of about one month (i.e., $\delta = 1$). We also used a nonparametric method (see Turnbull 1976) to estimate the cdf of the waiting-time distribution in stage 1 from the data in Table 1. However, the censoring pattern in the data resulted in a number of gaps in the empirical cdf, and it is difficult to compare the two curves. The nonparametric estimate of the time delay is the shortest observed seropositive time from Table 1, which is $\delta = 1.15$. It is clear that the actual time delay must be somewhat less than 1.15 months, and our estimate of $\delta = 1$ seems reasonable.

Our estimate of the mean AIDS incubation period of 9.8 years is consistent with estimates obtained by others using different statistical methods and data sets (see the discussion in Longini et al. 1989). Because 75% of the observations used to estimate the distribution of the AIDS incubation period were right-censored for the development of AIDS, our estimates of the mean and median AIDS incubation periods are highly distribution dependent. These estimates are extrapolations based on limited data and the selected model. Another possible approach to data analysis would be to use a semi-Markov model (Lagakos et al. 1978). This approach relies on nonparametric estimation of the waiting-time distribution in each stage. Unfortunately, the interval-censoring in our data is too severe for us to employ this approach. Nonetheless, we believe that our analysis provides the most efficient use of the available data.

Brookmeyer et al. (1987) have identified several sources of bias that can arise in the analysis of cohort data of HIV infected persons. The estimate of the AIDS incubation period can be affected by length-biased sampling if persons were selected because they developed AIDS. Such a bias results in an underestimate of the mean and median AIDS incubation period. Our data are not subject to this source of bias because persons in our samples were selected only because they were known to be infected. In addition, Brookmeyer et al. (1987) have described other forms of bias

that arise in prevalent HIV cohorts (i.e., cohorts with some or all of the persons being seropositive upon entry into the cohort) if the hazard function of the AIDS incubation period distribution is time dependent. Since the hazard function of the general gamma distribution that we employ is monotonically increasing, there is potential for such biasing. However, this problem is minimized in our analysis because the staged Markov model that we use partitions the AIDS incubation period into three stages, each with a constant hazard function.

A final caveat with respect to our estimate of the AIDS incubation period concerns how the data were combined. The waiting-time distribution for stage 1 was estimated largely from persons with nonsexual sources of infection (see Table 1), while the distributions for stages 2 and 3 were estimated from persons who probably were sexually infected. Thus, we make the assumption that the waiting-time-distribution for stage 1 does not vary with the source of infection.

Our estimates of the stage-specific infectious transmission rates appear to be the first such estimates made. Anderson and May (1988) review 23 published reports where the probability of transmission per sexual partnership was estimated. None of these reports considered stage-specific infection rates and the general trend was to calculate the probability of transmission per partnership rather than per sexual contact. Transmission probabilities that are calculated in the former fashion are difficult to compare since the duration of partnerships and rate of sexual contact per partnership may vary considerably. For male-to-female transmission (16 reports), the calculated probability ranged from 0.03 to 0.73 per sexual partnership. For female-to-male transmission (5 reports) it ranged from 0.08 to 0.71, while for male-to-male (2 reports) it varied from 0.10 to 0.60. The above reports included that of Fischl et al. (1987), which involved both male-to-female and female-to-male transmission, although Fischl et al. report no significant difference in the probability of infection by sex of the index case. As pointed out in Section 4, the overall probability of infection per partnership was 0.58 for that study.

Recently, Wiley et al. (1989) estimated the average probability of infection per male-to-female sexual contact (unprotected penile-vaginal intercourse) to be 0.001 from two cohorts of heterosexual partners where the man was the index infected person (see Padian et al. 1987 and Peterman et al. 1988). Their mean estimate is equal to ours when we assume that the

transmission rate is a constant across the stages of infection (see Table 4). However, they note that there is considerable heterogeneity in the per-contact infection transmission probability, and they speculate that this is due to considerable variation in infectiousness and/or susceptibility. DeGruttola et al. (1989) have estimated the transmission probability to be in the range of 0.005 to 0.010 per unprotected receptive anal contact for homosexually active men. They also report substantial heterogeneity in the transmission probability. Our results strongly support the hypothesis that the infection transmission rate exhibits considerable variation over the HIV infectious period. This is evident since the model with a constant transmission rate was rejected when compared to one with variation.

We found that the estimated probability of infection per sexual contact is eight times as high with an infected person with AIDS (stage 4) as with a person with pre-AIDS symptoms. Thus, it appears that persons with AIDS are much more infectious than those with pre-AIDS symptoms. This finding is supported by the study of Osmond et al. (1988) where 117 homosexual men, who were regular sexual partners of men with AIDS, were tested for HIV antibody. They found that receptive anal intercourse was a strong risk factor for infection for those men who had such contact beyond the date of AIDS diagnosis in the index case, but not if receptive anal intercourse ceased before the date of AIDS diagnosis.

There are several important points that need to be addressed when interpreting the estimated transmission rates and probabilities in table 4. First, we used a constant rate of 11 sexual contacts per month to calculate the probability of infection per sexual contact, p_i, because Fischl et al. (1987) did not report the number of sexual contacts per couple per month. If such data had been available, they could have been included in (15) to provide a direct estimate of p_i.

Second, for those couples in which both persons were found to be seropositive at the time of AIDS diagnosis of the "index" case, it may not be clear who was infected by whom. In fact, the person labeled "exposed" may have infected the "index," or both could have been infected from outside of the relationship (see Longini et al. 1982). If the former is true, our estimates would not be affected since the two persons are indistinguishable in terms of risk factors, and they are exposed to one another for the same period of time. If the latter is true, then information on exposure outside of the relationship would have to be

incorporated into the analysis to provide accurate estimates of the parameters (see Longini et al. 1982; Longini et al. 1988; and Haber et al. 1988).

Third, nine of the index cases were IV drug users. If their partners were also IV drug users, then they may have shared needles, and there would be an additional route of transmission among these couples. This would lead to an additional risk of transmission, which should be modeled and included in the analysis.

Fourth, all the index cases under study had AIDS. This could be a source of bias since our estimates of the infectious contact rates for the pre-AIDS stages are conditioned on exposure to persons who developed AIDS in a relatively short period of time. Such estimates could be different for exposure to persons who may take a very long period of time to develop AIDS.

Fifth, our estimate of the infectious contact rate during stage 2 turned out to be 0. This is because the parameter β_2 hit the lower bound of 0 at the maximized value of the likelihood function. This is a particular result from the data and the mean length of the AIDS incubation period that we used. When we used a mean AIDS incubation period of shorter length, the estimate for β_2 was a very small number, but not 0. We feel that this result indicates that the risk of infection given exposure to an asymptomatic, antibody-positive person is low in general, but such a person could be quite infectious if the right cofactors were present.

Finally, we were unable to obtain any estimate for the infectious contact rate in stage 1, β_1. This is because of the retrospective nature of the data prior to AIDS diagnosis of the index case and the extremely short duration of stage 1 with respect to the rest of the AIDS incubation period. The parameter β_1 could be estimated from a study that followed couples prospectively from the time of early detection of infection (or at least seroconversion) of the index case. However, such a study would be difficult to perform.

The transmission model presented here (15-21) was used to examine the risk of transmission by stage of infection. The data of Fischl et al. (1987) contained risk factor information that could have been examined, such as the type of index case (e.g., IV drug user, bisexual), the use of barrier contraceptives, the presence of concomitant venereal infections, and the direction of sexual transmission (i.e., male-to-female or

female-to-male); but numbers in each risk category were small. The effect of these risk factors on the infectious contact rates could be examined by formulating risk-specific rates. Longini et al. (1988) and Haber et al. (1988) have formulated such models for the transmission of viral diseases such as influenza. In the case of HIV, the infectious contact rate may be β_{ri}, which is the contact rate between a susceptible of risk category r with an infective who is in stage i of infection. Then the parameters could be estimated and the appropriate risk ratios examined. For example, let r = 1 if the susceptible has genital ulcers and r = 0 if not. Then the risk ratio $RR_i = \beta_{1i}/\beta_{0i}$ would measure the increased risk due to genital ulcers given exposure to an infected person in the i-th stage of infection. In addition, the index i could indicate additional categories for infectiousness such as use of condoms by infected males. Longini et al. (1988) have shown that risk ratios based on infectious contact rates provide a more accurate and less confounded measure of risk than do those based on infection attack rates. Such analyses should become feasible as more data become available.

Acknowledgements

This research was partially supported by contracts 200-07-0515 and 88060702 from the CDC and by NIH Grant 1-R01-AI22877.

References

Anderson, R.M. and R.M. May. (1988). Epidemiological parameters of HIV transmission. *Nature* London 333, 514-519.

Burke, D.S. and R.R. Redfield. (1988). Letter to the editor, *New England Med.* 318, 1202-1203.

Brookmeyer, R., M.H. Gail, and B.F. Polk. (1987). The prevalent cohort study and the acquired immunodeficiency syndrome. *Am. J. Epidemiol.* 126, 14-24.

Chiang, C.L. (1980). *An Introduction to Stochastic Processes and Their Applications*, 2nd ed. Krieger, New York.

DeGruttola, V., G.R. Seage, K.H. Mayer, and C.R. Horsburgh. (1989). Infectiousness of HIV between male homosexual partners, *J. Clin.*

Epidemiol. (In Print).

Elandt-Johnson, R.C. and N.L. Johnson. (1980). *Survival Models and Data Analysis.* Wiley, New York.

Fischl, M.A., G.M. Dickinson, G.B. Scott, N. Klimas, M.A. Fletcher, and W. Parks. (1987). Evaluation of heterosexual partners, children, and household contacts of adults with AIDS. *JAMA* 257, 640-644.

Goedert, J.J., M.E. Eyster, R.J. Biggar, and W.A. Blattner. (1987). Heterosexual transmission of human immunodeficiency virus: association with severe depletion of T-helper lymphocytes in men with hemophilia. *AIDS Res. Hum. Retrov.* 3, 355-361.

Haber, M., I.M. Longini and G.A. Cotsonis. (1988). Statistical analysis of infectious disease data. *Biometrics*, 44, 163-173.

Hethcote, H.W. (1988). Mathematical and computer models for the prediction of AIDS incidence: HIV prevalence and AIDS in San Francisco, Unpublished report to the CDC.

Hogg, R.V. and A.T. Craig. (1970). *Introduction to Mathematical Statistics*, 3rd ed. Macmillan, London.

Horsburgh, C.R., Y.O. Chin, J. Jason, S.D. Holmberg, I.M. Longini, C. Schable, K.H. Mayer, A.R. Lifson, G. Schochetman, J.W. Ward, G.W. Rutherford, G.R.Seage, and H.W. Jaffe. (1989). Defining the interval between human immunodeficiency virus infection and detection of antibody to the virus. Submitted, *Ann. Int. Med.*

Hyman, J.M. and E.A. Stanley. (1988). Using mathematical models to understand the AIDS epidemic. *Math. Biosci.* 90, 415-473.

Jacquez, J.A., C.D. Simon, J.S. Koopman, L. Sattenspiel, and T. Perry (1988). Modeling and analyzing HIV transmission: the effect of contact patterns. *Math. Biosci.* 92, 119-199.

Jaffe, H.W., W.W. Darrow, D.F. Echenberg, D.F., P.M. O'Malley, J.P. Getchell, V.S. Kalyanaraman, R.H. Byers, D.P. Drennan, E.H. Braff, J.W. Curran, and D.P. Francis. (1985). The acquired immunodeficiency syndrome in a cohort of homosexual men: a six-year follow-up study. *Ann. Int. Med.* 103, 210-214.

Johnson, N.L. and S. Kotz. (1970). *Continuous Univariate Distributions - 1.*Wiley, New York. Kalbfleisch, J.D. and R.L. Prentice. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Koopman, J.S., C. Simon, J. Jacquez, J. Joseph, L. Sattenspiel., and Park, T. (1988). Sexual partner selectiveness effects on homosexual HIV transmission dynamics. *J. AIDS* 1, 486-504.

Laga, M., H. Taeman, L. Bonneux, P. Cornet, G. Vercauterent, and P. Piot. (1988). Risk factors for HIV infection in heterosexual partners of HIV infected Africans and Europeans. *IV Int. Conf. AIDS, Stockholm* (Abstr.) I, 260.

Lagakos, S.W., C.J. Summer, and M. Zelen. (1978). Semi-Markov models for partially censored data. *Biometrika* 65, 311-317.

Lange, J.M.A., D.A. Paul, H.G. Huisman, deWolf, F., Van Den Berg, H., Coutinho, R.A., Danner, S.A., Van Der Noordaa, J. and Goudsmit, J. (1986). Persistent HIV antigenaemia and decline of HIV core antibodies associated with transition to AIDS. *Brit. Med. J.* 293, 1459-1462.

Longini, I.M., W.S. Clark, R.H. Byers, G.F. Lemp, J.W. Ward, W.W. Darrow, and H.W. Hethcote. (1989). Statistical analysis of the stages of HIV infection using a Markov model. *Statist. Med.* 8, 831-843.

Longini, I.M., J.S. Koopman, M. Haber, and G.A. Cotsonis. (1988). Statistical inference on risk-specific household and community transmission parameters of infectious diseases. *Am. J. Epidemiol.* 128, 845-859.

Longini, I.M., J.S. Koopman, A.S. Monto, and J.P Fox. (1982). Estimating household and community transmission parameters for influenza. Am *J. Epidemiol.* 115, 736-751.

Nekwei, W., R.L. Colebunders, Y. Bahwe, I. Leburghe, H. Francis, and R. Ryder. (1988). Acute manifestations of HIV infection following blood transfusion. *IV Int. Conf. AIDS, Stockholm* (Abstr.) II, 350.

Osmond, D., P. Bacchetti, R.E. Chaisson, T. Kelly, R. Stempel, J. Carlson, and A.R. Moss. (1988). Time of exposure and risk of HIV infection in homosexual partners of men with AIDS. *Am. J. Pub. Health* 78, 944-948.

Padian, N., J. Wiley, and W. Winkelstein. (1987). Male-to-female transmission of human immunodeficiency virus (HIV): current results, infectivity estimates, and San Francisco population seroprevalence estimates. *III Int. Conf. AIDS, Stockholm* (Abstr.) THP. 3-48, 171.

Peterman, T.A., R.L. Stonebrunner, J.R. Allen, H.W. Jaffe, and J.W. Curran. (1988). Risk of human immunodediciency virus transmission for heterosexual adults with transfusion-associated infections. *JAMA* 259, 55-58.

Ralston, M. (1985). Derivative-free nonlinear regression, in *BMDP Stat. Software Manual*. U. Cal. Press, Berkeley, 302-329.

Seage, G.R., C.R. Horsburgh, A.M. Hardy, K.H. Mayer, M.A. Barry, J.E. Groopman, H.W. Jaffe, and G.A. Lemp (1989). Increased suppressor T cells in probable transmitters of human immunodeficiency virus infection, submitted. *Am. J. Pub. Health.*

Turnbull, B.W., (1976). The empirical distribution function with arbitrary grouped, censored, and truncated data. *J. Roy. Statist. Soc., Series B* 38, 290-295.

Ward, J.W., S.D. Holmberg, J.R. Allen, J.R., Cohn, D.O., Kritchley, S.E., Cleinman, S.H., Lenes, B.A., Ravenholt, O., Davis, J.R., Quinn, M.G. and Jaffe, H.W. (1988). Transmission of human deficiency virus (HIV) by blood transfusions screened as negative for HIV antibody. *New England Med.* 318, 473-478.

Wiley, J.A., S.J. Herschkorn, and N.S. Padian. (1989). Heterogeneity in the probability of HIV transmission per sexual contact: the case of male-to-female transmission in penile-vaginal intercourse. *Statist. Med.* 8, 93-102.