

Statistical Studies of Age - Specific HIV - Prevalence Data

Estudios estadísticos sobre datos de prevalencia del VIH
según grupos de edad

HELMUT KNOLLE^a

Abstract

The infectivity function is a function giving a measure of how infectious a given individual is t time units after becoming infected. Today, no feasible and ethically acceptable study design is known, which would lead to estimates of HIV-infection probability within steady heterosexual partnerships, using standard statistical methodology. In this paper a transmission model is used as a link between the infectivity function and data sets which already exist or can be generated with standard methods and moderate expenses. This model suggests that the distribution of HIV-infections by age and sex depends on the infectivity function as well as on age-dependent patterns of sexual partner choice. Application of the model requires population-based data of age-specific HIV-incidences in men and women of the general heterosexual population. At present, the only known data set suitable for this purpose is a set of HIV-test results from a sample of 8690 Colombian women in pregnancy who attended prenatal care. The prevalence of HIV was 0.33% in the group of 12-24 years, but only 0.16% in the group of 25-34 years. The model can explain this strange result. A data set of age-specific HIV-prevalences in heterosexual Colombian men would be useful, but is not known. Therefore, further research and data collecting is required in order to arrive at well founded conclusions.

Key words: AIDS, Infectivity, Health risk, Branching process, Threshold, Age distribution, Colombia.

Resumen

La función de infectividad es una función que dice qué tan infecciosa es una persona transcurrido un tiempo t después de haberse infectado. Hoy en día no se conoce ningún diseño de estudio, que sea factible y éticamente aceptable y que conduzca a estimados de la infectividad del VIH entre uniones heterosexuales estables, usando los métodos estadísticos corrientes. Por eso,

^aInvestigador independiente. E-mail: hknolle@gmx.ch

aquí se utiliza un modelo de transmisión como puente entre la función de infectividad y datos que ya existen o pueden ser generados con pocos gastos. El modelo sugiere que la distribución según sexo y edad de la infección por VIH en la población heterosexual de bajo riesgo depende de la función de infectividad así como de los patrones de mezcla entre diferentes grupos de edad. La aplicación del modelo requiere datos poblacionales sobre la incidencia de la infección en los grupos de edad de hombres y mujeres heterosexuales. Actualmente el único conjunto de datos adecuado consiste en los resultados de pruebas por VIH en 8690 mujeres colombianas en embarazo que asistieron a consulta de control prenatal. La prevalencia del VIH era 0.33% en el grupo de 12-24 años, y 0.16% en el grupo de 25-34 años. El modelo puede explicar esta diferencia. Pero no existen datos sobre prevalencias del VIH en hombres heterosexuales de diferentes edades que se necesitan en este contexto. Por eso, se requieren otras investigaciones y recolecciones de datos para llegar a conclusiones seguras.

Palabras clave: sida, infectividad, riesgo de salud, proceso ramificado, umbral, distribución de edad, Colombia.

1 Introduction

Infectious disease data have two features that distinguish them from other data. They are highly dependent and the infection process is only partially observable. A consequence of these features is that the analysis of data is usually most effective when it is based on a model that describes aspects of the infection process, i.e. on a transmission model (Becker & Britton 1999). The core of any transmission model is the infectivity function. This is a function $b(t)$ giving a measure of how infectious a given individual is t time units after becoming infected. Knowledge about the infectivity function is of great public health interest. Remember that the most used method of disease control in the 19th century and before has been the isolation of infected persons. Still a few years ago, isolation has been used for the control of the new lung disease SARS. But when should isolation begin, and when should it end?. The answer depends on the infectivity function. The spread of an infectious disease is at once a biological and a social phenomenon. Therefore, any mathematical model describing this process must take into account the types of social behavior that are related to the transmission of the infectious agent. This is especially true in the case of HIV, the causative agent of AIDS. The sexual transmission of HIV in groups with different behavior requires different models with different parameters.

During the early stage of mathematical modeling of the AIDS epidemic, until 1988 say, it was assumed that the infectivity function of sexually transmitted HIV is constant during all the incubation period of AIDS and falls to zero at its end. The event of transmitting HIV from an infective to a susceptible partner by c contacts was conceived as the first success in a series of c independent Bernoulli trials with parameter b . So, the probability of being infected after c contacts with an infected partner would be $P(c) = 1 - (1 - b)^c$. This is equivalent to $P(t) = 1 - e^{-gt}$ when

infection is considered as the first event of a Poisson process with intensity g . Early deterministic models in continuous time, e.g. the model in Dietz & Hadelers (1988), have usually assumed a Poisson process with constant intensity. For heterosexual pairs it has been admitted that b might be different depending on whether the infective partner is male or female. In any case, P would be a strictly increasing function of c or t , respectively.

This way of thinking has changed after the publication of a study by Peterman et al. (1988). Since this study involved only heterosexual pairs in which one partner had become infected by blood transfusion, the number of acts of intercourse with the infected partner could be estimated rather accurately. The surprising result of this study was that pairs in which transmission had not occurred reported about twice as many sex acts than pairs in which the spouse of the index case was HIV-positive at the end of the study.

Two hypotheses that could explain these data have been proposed. One hypothesis assumes that HIV-transmission in a partnership with n sexual contacts can still be conceived as a possible outcome of n independent Bernoulli trials with probability of success b , but that b varies according to a high-low-high pattern during the incubation period of AIDS. Formally, this can be written as:

$$b(t) = \begin{cases} b_0 & 0 < t < t_1 \\ \beta - \epsilon & t_1 < t < t_2 \\ \beta + \epsilon & t_2 < t < t_3 \end{cases} \quad 0 < \epsilon < \beta$$

where t_1 is between 4 and 8 weeks, and t_2 is greater than 5 years.

The papers of Hadelers (1989) y Dietz et al. (1993) are based on this hypothesis which may be called the modified Bernoulli model. The other hypothesis implies that the infection probability per partnership varies between pairs (and perhaps also in time) and that an uninfected partner becomes infected around the time of the first sexual contact in a partnership or never (soon-or-never model). This is assumed by Watts & May (1992).

Today, studies such as that of Peterman et al. (1988) are no longer possible because HIV-infections by blood transfusion are extremely rare. No feasible and ethically acceptable study design is known, which would lead to estimates of HIV-infection probability within steady heterosexual partnerships, using standard statistical methodology. Therefore the idea of this paper is to use a transmission model as a link between the infectivity function and data sets which already exist or can be generated with standard methods and moderate expenses. A model for the spread of HIV via steady heterosexual partnerships was proposed by Knolle (2004a). This model can be used to predict the age-distribution of HIV-infections in a population (sometimes called the general population) of heterosexual men and women who have sexual contacts only within a steady monogamous partnership and do not have any other risk of infection with HIV. Let's first look at some data sets of age-specific HIV-prevalences in different countries. After that the transmission model will be presented, and it will be explained how the model and the data can be used in order to infer properties of the infectivity function.

2 Data from Colombia, Burundi and the United Kingdom

A major difficulty in obtaining HIV-data from the general population is to separate HIV-tests of people with high sexual mobility from tests of people in steady partnerships. Perhaps the best way out is to look for data of pregnant women in prenatal care and of men who make an HIV-test as prerequisite for a life insurance contract, because a high percentage of these people is likely to live in a steady partnership. Surely, a selection bias is always possible, if the test is voluntary.

An unlinked anonymous HIV-test is a test that is made for purely epidemiological reasons, such that the test result cannot be linked to personal identifiers. Unlinked anonymous testing of pregnant women who attend prenatal care provides estimates of HIV-prevalence in the general female population which are almost free of selection bias. Colombia is one of the very few countries in which this epidemiological tool has been applied. The following data are taken from a bulletin published by the Instituto Nacional de Salud (2000). During 8 weeks in 1999 all women who attended prenatal care in selected health cares at 12 different locations were tested, giving a total of 8690 tests among which 21 were positive (Table 1).

TABLE 1: Prevalences of HIV-infection in Pregnant Women in Colombia (1999).

Age	Number Tested	Number of HIV+	%HIV+
< 12	3	0	
12-24	3851	13	0.33
25-34	3614	6	0.16
35-44	1096	2	0.18
> 44	50	0	
Unknown	76	0	
Total	8690	21	

Source: Instituto Nacional de Salud (2000).

During the same campaign, all men and women who attended general medical in selected health cares, and all men and women who attended consultation for sexually transmitted diseases (STD) were tested. Tables 2 and 3 show data for Colombian men and women who consulted for STD, respectively.

The data for men will not be considered here, because the percentage of homosexual men among these is unknown, but may be high. In order to have specific data for heterosexual men which can be compared with the data for women in Table 1, one must go to a region with high HIV-prevalence in the heterosexual population, e.g. Africa.

Table 4 shows HIV-prevalence in different age groups in a sample of 1254 men in Bujumbara, the capital of Burundi (Saidel et al. 1996). All men in the sample were workers from five companies who were invited to participate on a voluntary basis in a longitudinal study to determine the annual incidence rate of HIV-infection. The choice of this data, that can be criticized in view of cultural and genetic

differences between the populations of both countries, is partly justified by the scarcity of HIV-data for heterosexual men.

In a previous paper, this author has argued that in growing populations the patterns of mixing between age-groups show typically a higher frequency of partnerships between young women and older men than in nongrowing populations (Knolle 2005). Such patterns lead to higher values of the epidemic threshold parameter of HIV and may also influence the age-distribution of HIV-prevalence (see below). In order to compare these predictions with observed trends, data such as those of Table 1, but from developed countries, would be helpful. But it seems that anonymous unlinked testing of representative samples of the general population is rarely carried out in developed countries. As a surrogate, Table 5 shows the numbers by age of all heterosexual HIV-infections in women between 15 and 44 years, diagnosed between 1999 and 2003 in the United Kingdom (UK).

Now we can make some preliminary comparisons. At the 5% level, the jump in the prevalence of HIV-infection from men younger than 28 years to older men

TABLE 2: Prevalences of HIV-infection in men who consulted for SDT in Colombia (1999).

Age	Number Tested	Number of HIV+	%HIV+
< 12	4	0	
12 – 24	329	4	1.21
25 – 34	486	8	1.65
35 – 44	311	7	2.25
> 44	225	1	0.44
<i>Unknown</i>	2	0	
<i>Total</i>	1357	20	

Source: Instituto Nacional de Salud (2000).

TABLE 3: Prevalence of HIV-infection in women who consulted for SDT in Colombia (1999).

Age	Number Tested	Number of HIV+	%HIV+
12 – 24	967	8	0.83
25 – 34	1090	7	0.64
35 – 44	659	3	0.45
> 44	272	0	
<i>Unknown</i>	5	0	
<i>Total</i>	2993	18	

Source: Instituto Nacional de Salud (2000).

TABLE 4: Prevalences of HIV-infection in a Sample of Men in Bujumbara, Burundi.

Age	Number Tested	Number of HIV+	%HIV+	95% interval
18 – 27	500	42	8.4	(5.9, 10.9)
28 – 37	569	101	17.8	(14.5, 21.1)
38 – 47	185	33	17.8	(12.0, 23.6)

Source: Saidel et al. (1996).

TABLE 5: Infections with HIV in women, probably acquired through sexual intercourse with men, diagnosed between 1999 and 2003 in the United Kingdom.

Age	Year					
	1999	(1999)	2000	2001	2002	2003
15 – 19	15	(8.4)	30	53	65	82
20 – 24	105	(61.4)	155	225	300	328
25 – 29	207	(100)	306	468	620	681
30 – 34	223	(94.9)	337	487	571	703
35 – 39	159	(68.8)	216	278	420	440
40 – 44	66	(32.8)	89	147	180	237

The numbers in parentheses are related to 1 million women.
 Source: Health Protection Agency (2004, Quarterly Tables 65, 04/4, Table 11a).

in Burundi is significant (Table 4). In order to compare among age-groups in (Table 5), it is necessary to calculate the number of HIV diagnoses per 1 million women in each group. This can be done easily for the data of 1999, because for that year general census data are available (*UN Demographic Yearbook* 2002). The results are shown in parentheses. Similarly to men in Burundi, HIV incidence in women in the UK younger than 25 is significantly lower (at 5%) than in women between 25 and 29 years, the confidence limits being 47 and 77 for 61, and 80 and 120 for 100 (Sachs 1984, pp. 267).

In sharp contrast to these results, in pregnant women in Colombia the estimated prevalence is highest in the age-group from 12 to 24 years (Table 1). But greater samples are needed, because with the present data the hypothesis $p_1 = p_2 = p_3$ cannot be rejected at the 5% level (the chi-square is 2.5 with 1 d.f.). If the trend shown in Table 1 would be confirmed by the results of the study carried out in 2004 (in spite of several attempts these data could not be obtained from the Instituto Nacional de Salud), the age-distribution shown in Table 1 could be a feature typical for HIV-infections in monogamous women in growing populations. Support for this thesis will be found in Section 4.

3 The Threshold Phenomenon

A central theorem of the stochastic theory of infectious diseases is the epidemic threshold theorem. It states that in large populations, there will be either minor epidemics or major epidemics with hardly any epidemics of a size between these two extremes. The epidemic threshold parameter R_0 has the property that a major epidemic is an event of positive probability if and only if $R_0 > 1$. For a homogeneous population, R_0 is the mean number of infections caused by one infective person who is introduced into the totally susceptible population. Its actual value depends both on the biology of the infection process and on social behavior of the host population. Figure 1 shows the probability distribution of the final size of an epidemic due to an air-borne viral infection in a school class of 16 children when $R_0 = 2.56$ and the number of contacts is Poisson-distributed. The calcula-

4 The Transmission Model

A branching process which describes HIV-transmission in the monogamous heterosexual population has been defined in Knolle (2004a):

Definition 1. Suppose that men have at most k and women have at most m partnerships during their life. The partnership-infection process is a Galton-Watson process with $k + m$ types which are defined as follows:

- for $1 \leq i \leq k$ type i is a man who acquired infection in his i -th partnership
- for $k + 1 \leq i \leq k + m$ type i is a woman who acquired infection in her $(i-k)$ -th partnership.

The generating functions of this process are calculated on the basis of the following probabilities:

- c_{ij} = probability that a man in his i -th partnership has a partner who is in her j -th partnership ($i = 1, \dots, k; j = 1, \dots, m$);
- d_{ij} = probability that a woman in her i -th partnership has a partner who is in his j -th partnership ($i = 1, \dots, m; j = 1, \dots, k$);
- u_{il} = probability that a man infected in his i -th partnership will infect his l -th partner ($i = 1, \dots, k - 1; l = i + 1, \dots, k$)
- v_{il} = probability that a woman infected in her i -th partnership will infect her l -th partner ($i = 1, \dots, m - 1; l = i + 1, \dots, m$).

The epidemic threshold parameter R_0 is given by the largest eigenvalue of the mean matrix of this Galton-Watson process. In the present case, however, the process is not positively regular, but it is still true that there is a positive probability of a major epidemic if and only if $R_0 > 1$ (Knolle 2004a).

The conditions of this model allow also to include episodes of infidelity. If a stable partnership with number i is interrupted by an episode of infidelity, this episode must be numbered $i + 1$ and the continuation of the partnership must be numbered $i + 2$.

We call $C = (c_{ij})$ and $D = (d_{ij})$ the mixing matrices of the process. The problem of assigning realistic values to the entries of these matrices has not yet been solved completely. The major problem is that in most countries the only available data refer to marriages, whereas data on informal partnerships do not exist. Some issues of the *Demographic Yearbook* of the United Nations contain the special topic "Nuptiality", for example the issue for the year 1982 published in 1984. In that issue, Table 29 is a cross-table for age of both spouses, and Table 30 is a cross-table for previous marital status of both spouses. If one compares data from countries with rapidly growing and with slowly growing or constant populations, then two typical patterns can be observed. Everywhere the mean age at first marriage is some years higher in men than in women. The difference is in

the correlation. Indeed, the correlation between age of groom and age of bride is weaker in rapidly growing populations, because unions between elderly men and very young women are frequent (Table 6).

TABLE 6: Age-Distribution at marriage of wives who marry a man of age 35-39 years in Ecuador (1976) and in Switzerland (1981)

Age	Ecuador	Switzerland
< 20	11.9	0.2
20 – 24	22.3	15.6
25 – 29	24.1	31.9
30 – 34	20.5	28.5
35 – 39	12.5	15.8
> 39	8.7	7.3
Total	100(= 1763)	100(= 2731)
Grow Rate	3.4%	0.2%

A similar trend is seen in cross-tables of marital status before marriage. In constant populations many divorced women marry again, and most divorced men marry a woman who is also divorced. In rapidly growing populations most divorced men marry a woman who never married before, and most divorced women do not marry again.

The following assumptions are partly justified by the marriage patterns in rapidly growing populations. It is assumed that marriage may be preceded by two informal partnerships.

Number of partnerships:

$$k = 4, m = 3$$

Mixing matrices:

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ g_1 & g_2 & g_3 \end{bmatrix}$$

$$D = \begin{bmatrix} 1 - h_1 & 0 & 0 & h_1 \\ 0 & 1 - h_2 & 0 & h_2 \\ 0 & 0 & 1 - h_3 & h_3 \end{bmatrix}$$

This means that a woman in her i -th partnership ($i = 1, 2, 3$) with probability $1 - h_i$ has a partner who also is in his i -th partnership, but with probability h_i has a partner who is in his fourth partnership. We suppose that condoms are not used in steady partnerships. About the infection process the following assumptions are made:

1. Nobody knows its own or the infection status of his (her) partner.
2. If an infectious and a susceptible person form a pair, transmission occurs soon or never.

3. The infectivity function has three phases with different levels.
4. Any partnership is longer than the first phase.
5. Any partnership is shorter than the second phase.
6. The first three partnerships together are longer than the incubation period of AIDS.

Assumption number 4 implies that the level of the first phase does not appear in the model. The other assumptions can be formalized with the following values of the infection probabilities:

$$\begin{aligned}
 u_{12} &= u_{23} = u_{34} = a - \delta, \\
 v_{12} &= v_{23} = b - \varepsilon, \\
 u_{13} &= u_{24} = a + \delta, \\
 v_{13} &= b + \varepsilon, \\
 u_{14} &= 0
 \end{aligned}
 \quad (-a \leq \delta \leq a, -b \leq \varepsilon \leq b)$$

The mean matrix of the Galton-Watson process is:

$$M = \begin{bmatrix} 0 & A \\ B & 0 \end{bmatrix}$$

where:

$$A = \begin{bmatrix} 0 & a - \delta & a + \delta \\ h_1(a + \delta) & h_2(a + \delta) & h_3(a + \delta) + (a - \delta) \\ h_1(a - \delta) & h_2(a - \delta) & h_3(a - \delta) \\ 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & (b - \varepsilon) & (b + \varepsilon) & 0 \\ 0 & 0 & (b - \varepsilon) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The zero rows and columns are due to the fact that this process is not positively regular. The positive square root of

$$R_0^2 = 2h_1(ab - \delta\varepsilon) + h_2(a - \delta)(b - \varepsilon) \quad (1)$$

is the largest eigenvalue of M , and R_0^2 is an eigenvalue of M^2 with linear independent eigenvectors

$$\begin{aligned}
 V_1 &= (0, h_1(b - \varepsilon), h_1(b + \varepsilon) + h_2(b - \varepsilon), 0, 0, 0, 0) \\
 V_2 &= (0, 0, 0, 0, h_1, h_2, 0)
 \end{aligned}$$

Equation (1) shows the strong influence of the mixing pattern (h_1 and h_2) and of variable infectivity (δ and ε) on R_0 . Data such as that in Table 6 (age < 20)

indicate that h_1 is surely greater in Ecuador than in Switzerland, but a formal estimation procedure is not yet available. In non-growing populations such as that of Switzerland and other European countries, the patterns of partner choice tend to symmetry, i.e. women usually have as many partners during life as men, and the mixing matrices are identical and symmetric. The general model of the partnership-infection-process can be applied to this case, too. For example:

let $k = m = 3, u_{ij}$ and v_{ij} as before, and for $0 < h < \frac{1}{2}$ let

$$C = D = \begin{bmatrix} 1 - 2h & h & h \\ h & 1 - 2h & h \\ h & h & 1 - 2h \end{bmatrix}$$

In this case we have:

$$A = \begin{bmatrix} 2ah & (a - \delta)(1 - 2h) + (a + \delta)h & (a - \delta)2h + (a + \delta)(1 - 2h) \\ (a - \delta)h & (a - \delta)h & (a - \delta)(1 - 2h) \\ 0 & 0 & 0 \end{bmatrix}$$

and the same expression for B , when a is replaced by b and δ by ε . In the case $h = \frac{1}{3}$ (uniform mixing) we find, according to equation (6.2) in Knolle (2004a):

$$R_0^2 = \left(a - \frac{\delta}{3}\right) \left(b - \frac{\varepsilon}{3}\right) \tag{2}$$

The corresponding eigenvectors of M^2 are:

$$V_1 = [1, 1, 1, 0, 0, 0]$$

$$V_2 = [0, 0, 0, 1, 1, 1]$$

This means that for uniform mixing between age groups, the risk of HIV-infection would be independent of age.

For other values of h , an explicit formula for R_0 is not available, but it can be shown numerically that R_0 is an increasing function of h .

For $a = b = 0.6, \delta = \varepsilon = 0.4, h = 0.2$, one obtains $R_0 = 0.291$, whereas $h = \frac{1}{3}$ implies $R_0 = 0.467$, according to equation (2). The nonzero components of the eigenvectors are 0.2, 0.255, 0.545. This means that the incidence of HIV-infection increases with age.

5 Towards Parameter Estimation

The HIV-epidemic in Colombia is of a type that the World Health Organization calls a concentrated epidemic. This means a situation in which a self-sustained epidemic is occurring only in some small groups at high risk, but with infection chains radiating from these groups into the general population. Therefore, the

adequate mathematical model for the spread of infection in the general population is the Galton-Watson process with immigration. Quine (1970) has proved that under fairly general conditions the distribution of types of a subcritical multitype Galton-Watson process with immigration converges to a stationary distribution. In the present case however, it is more convenient to derive the stationary distribution directly.

Consider a usual way by which an infection can be imported into the general population. Suppose a man has himself become infected by contact with a female sex worker and then infects his actual or his next steady partner. In this case, the ancestor of the partnership-infection process is a woman in her first, second or third partnership, i.e. $Z_0 = e_5, e_6$, or e_7 , where e_i denotes the i -th unit vector ($i = 1, \dots, k + m$), and $E(Z_1)$ is, respectively,

$$\begin{aligned} e_5 M &= (0, b - \varepsilon, b + \varepsilon, 0, 0, 0, 0) \\ e_6 M &= (0, 0, b - \varepsilon, 0, 0, 0, 0) \\ e_7 M &= 0 \end{aligned}$$

For $E(Z_2)$ we find:

$$\begin{aligned} e_5 M^2 &= 2(ab - \delta\varepsilon)(0, 0, 0, 0, h_1, h_2, x) \\ e_6 M^2 &= (a - \delta)(b - \varepsilon)(0, 0, 0, 0, h_1, h_2, y) \end{aligned}$$

where the values of x and y have no impact on further generations, since $e_7 M = 0$.

We continue and find

$$\begin{aligned} e_5 M^3 &= 2(ab - \delta\varepsilon)(0, h_1(b - \varepsilon), h_1(b + \varepsilon) + h_2(b - \varepsilon), 0, 0, 0, 0) \\ e_6 M^3 &= (a - \delta)(b - \varepsilon)(0, h_1(b - \varepsilon), h_1(b + \varepsilon) + h_2(b - \varepsilon), 0, 0, 0, 0) \end{aligned}$$

Since these are multiples of the eigenvector V_1 of M^2 , we conclude that

$$E(Z_{2n+1}) \text{ is a multiple of } V_1 \text{ for all } n > 0$$

Furthermore, a simple calculation shows that

$$E(Z_{2n}) \text{ is of the form: } \lambda(0, 0, 0, 0, h_1, h_2, x) \text{ for all } n > 0$$

Of course, the generations Z_0, Z_1, \dots are not observable separately. What can be observed is always a superposition of different generations of several distinct branching processes.

When R_0 is < 1 , but sufficiently near to 1, the total of generations Z_2, Z_3, \dots tends to outweigh Z_0 and Z_1 .

If X^1, \dots, X^7 denote the observed numbers of infections of type 1, $\dots, 7$ in a given moment, then it can be hoped that

$$\frac{X^3}{X^2} \text{ is approximated by } \frac{b + e}{b - e} + \frac{h_2}{h_1} \text{ and } \frac{X^6}{X^5} \text{ by } \frac{h_2}{h_1}$$

Therefore $\frac{X^3}{X^2} - \frac{X^6}{X^5}$ is a statistic for a test of the hypothesis $\varepsilon > 0$ against $H_0 : \varepsilon = 0$. Its expectation for H_0 is 1, but the variance is not known to the author.

Furthermore, if $\varepsilon > 0$ then it may be concluded that X^3 must be greater than X^2 , independently of h_2 and h_1 , whereas X^6 can be expected to be smaller than X^5 whenever $h_2 < h_1$.

In contrast to these results, the expected age-distribution of HIV-infection is increasing with age for both men and women, if the mixing pattern is symmetric and old-with-young mixing is rare. This can be seen from a numerical calculation of the eigenvectors in the mean matrix M (see the example in section 4).

An anonymous referee has suggested to use the data of the tables in section 2 for an estimation of the R_0 of Colombia. Unfortunately, the data available at present and the theory developed so far are not sufficient for this purpose. For example, the data of Table 1 are prevalence data, but what is really needed is incidence data. Furthermore, data of heterosexual men are totally lacking. Further progress would require close collaboration between mathematical statisticians and the personnel of public health institutions.

6 Applications to Public Health

The use of condoms is the most recommended method of control of the HIV-epidemic. But many heterosexual pairs usually do not use condoms. Knowledge of the infectivity function as defined in the introduction can lead to recommendations that take this fact into account. If it is true that the infectivity of HIV varies in time according to a high-low-high pattern and that the first phase of high infectivity is shorter than two months, then an obvious recommendation would be: take a period of two months of abstinence between the last and the first sexual contact with two different partners, and make sure that any new partner follows the same rule. A second recommendation, derived from equation (1), is directed at reducing the frequency of old-with-young mixing and implies the advice for women, to be cautious with older men or, if a partnership with an old man is strongly desired, ask him first to make an HIV-test.

Recibido: abril de 2006

Aceptado: septiembre de 2006

References

- Becker, N. & Britton, T. (1999), 'Statistical Studies of Infectious Disease Incidence', *J. R. Statist. Soc.* **B61**(Part 2), 287–307.

- Becker, N. & Marschner, I. (1990), The Effect of Heterogeneity on the Spread of Disease, *in* G. Lefèvre, ed., 'Stochastic Processes in Epidemic Theory', Vol. 86, Springer.
- Dietz, K. & Hadelers, K. P. (1988), 'Epidemiological Models for Sexually Transmitted Diseases. Part 2: Effects of Variable HIV-Infectivity.', *J. Math. Biology* **26**, 1–25.
- Dietz, K., Heesterbeek, J. & Tudor, D. (1993), 'The Basic Reproduction Ratio for Sexually Transmitted Diseases. Part 2: Effects of Variable HIV-infectivity', *Math. Biosciences* **117**, 35–47.
- Hadelers, K. (1989), 'Modeling AIDS in Structured Populations', *Bull. Int. Stat. Inst.* **53**, 83–99.
- Harris, T. E. (1963), *The Theory of Branching Processes*, Springer, Berlín.
- Health Protection Agency (2004), 'HIV and other Sexually Transmitted Infections in the United Kingdom in 2002 Annual Report, November 2003'.
- Instituto Nacional de Salud (2000), Informe quincenal epidemiológico nacional, Bogotá, pp. 355–365.
- Knolle, H. (2004a), 'A Discrete Branching Process Model for the Spread of HIV Via Steady Sexual Partnerships', *Journal of Mathematical Biology* **48**, 423–443.
- Knolle, H. (2004b), 'Martingalas en la teoría de epidemias', *Revista Colombiana de Estadística* **27**(1), 67–75.
- Knolle, H. (2005), 'Un modelo demográfico y su interpretación sociológica', *Boletín de Matemáticas* **12**(2), 145–154.
- Lefèvre, C. & Picard, P. (1989), 'On the Formulation of Discrete-Time Epidemic Models.', *Math. Biosci.* **95**, 27–35.
- Peterman, T., Stoneburner, R., Allen, J. & Jaffe, H. (1988), 'Risk of Human Immunodeficiency Virus Transmission from Heterosexual Adults with Transfusion-associated Infections', *J. Amer. Med. Ass* **259**, 55–58.
- Quine, M. P. (1970), 'The Multi-type Galton-Watson Process with Immigration', *J. Appl. Prob* **7**, 411–422.
- Sachs, L. (1984), *Angewandte Statistik*, Springer, Berlín.
- Saidel, T., Sokal, D., Rice, J., Buzingo, T. & Hassig, S. (1996), 'Validation of a Method to Estimate Age-specific HIV-incidence Rates in Developing Countries Using Population-Based Seroprevalence Data', *Am. J. Epidemiol.* **144**, 214–223.
- Watts, C. H. & May, R. M. (1992), 'The Influence of Concurrent Partnerships on the Dynamics of HIV/AIDS', *Math. Biosci* **108**, 89–104.