MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume 4, Number 4, October 2007

http://www.mbejournal.org/

pp. 595–607

ON THE BASIC REPRODUCTION NUMBER R_0 IN SEXUAL ACTIVITY MODELS FOR HIV/AIDS EPIDEMICS: EXAMPLE FROM YUNNAN, CHINA

Nicolas Bacaër

Institut de Recherche pour le Développement (I.R.D.)
32 avenue Henri Varagnat, 93143 Bondy cedex, France

Xamxinur Abdurahman

College of Mathematics and System Sciences, Xinjiang University
14 Shengli Road, Urumqi, 830046, China

Jianli Ye

National Center for Women and Children's Health, Department of Information Management
13 Dong Tu Cheng Street, Chang Yang District, Beijing, 100013, China

Pierre Auger

Institut de Recherche pour le Développement (I.R.D.)
32 avenue Henri Varagnat, 93143 Bondy cedex, France

(Communicated by Abba Gumel)

Abstract. Heterogeneity in sexual behavior is known to play an important role in the spread of HIV. In 1986, a mathematical model based on ordinary differential equations was introduced to take into account the distribution of sexual activity. Assuming proportionate mixing, it was shown that the basic reproduction number R_0 determining the epidemic threshold was proportional to $M + V/M$, where M is the mean and V the variance of the distribution. In the present paper, we notice that this theoretical distribution is different from the one obtained in behavioral surveys for the number of sexual partnerships over a period of length τ. The latter is a "mixed Poisson distribution" whose mean m and variance v are such that $M = m/\tau$ and $V = (v - m)/\tau^2$. So $M + V/M = (m + v/m - 1)/\tau$. This way, we improve the link between theory and data for sexual activity models of HIV/AIDS epidemics. As an example, we consider data concerning sex workers and their clients in Yunnan, China, and find an upper bound for the geometric mean of the transmission probabilities per partnership in this context.

1. Introduction.

In 2002, behavioral surveys were undertaken among eight hundred adult males (ages 18 to 50) and eight hundred female sex workers in Yunnan and Sichuan, two provinces in southwest China. These surveys were part of the China-UK HIV/AIDS Prevention and Care Project. Among many other questions, the adult males were asked about the number of commercial sex partners during the last 12 months [11, Table 148-149]. Female sex workers were asked about the number of clients during the week before the interview [12, Table 78]. The detailed

2000 *Mathematics Subject Classification.* 92D30.
Key words and phrases. HIV, basic reproduction number, mixed Poisson distribution.

results for the 407 adult males from Yunnan are shown in Table 1. The samples from different cities have been weighted according to the population of these cities [11, Table 201], so Table 1 does not show integer fractions of 407. The corresponding mean and standard deviation (the square root of the variance) are shown in the first line of Table 2. For the 403 female sex workers from Yunnan who participated in the study, only the mean, the standard deviation, the minimum and the maximum have been published (second line of Table 2).

Table 1. Distribution of the number of commercial sex partners in the last 12 months for 407 adult males in Yunnan. From [11, Table 149].

	0	1	2	3	4	5	6-9	10-14	> 14
%	78.6	1.9	3.2	2.9	1.7	1.8	3.5	4.3	2.1

Table 2. Distribution of the number of commercial sex partners in the last 12 months for 407 adult males and distribution of the number of clients during one week for 403 female sex workers in Yunnan. From [11, Table 148] and [12, Table 78].

	mean	standard deviation	minimum	maximum
adult males	1.0	2.2	0	not available
sex workers	3.0	4.1	0	40

In Kunming, the provincial capital of Yunnan, the HIV epidemic became very serious among injecting drug users in 1996-1997 (Fig. 1a). Since then, it seems that HIV prevalence among this high risk group has stabilized somewhere between 20% and 30%, a level probably determined by the percentage of injecting drug users sharing needles. Sexual transmission of HIV seems limited compared to what occured some years ago in other areas of Southeast Asia such as Thailand. HIV prevalence has remained relatively low among female sex workers and their clients (Fig. 1b, c). Notice that the small increase after 1996 is probably due to the fact that some female sex workers and some clients are also injecting drug users. But there is no evidence of an exponential increase in sexual transmission. The situation in other parts of Yunnan province has been reviewed recently [18].

The absence of an exponential increase in HIV prevalence among sex workers suggests that the basic reproduction number R_0 for sexual transmission of HIV between sex workers and clients is below 1. This is somewhat surprising. Is it possible to understand this observation from the data concerning sexual behavior?

In this paper we will use a mathematical model for HIV epidemics, which takes into account the distribution of sexual activity. The first models of this kind focused on male homosexuals and divided the population according to the number k of sexual partners per year ([1, 2, 15] and [4, Chap. 11]): the compartment k was called the "sexual activity group k" (k being a nonnegative integer). Assuming proportionate mixing, it was shown that the epidemic threshold did not depend only on the mean (call it M) but also on the variance V of the distribution. More precisely, the basic reproduction number R_0 was shown to be proportional to $M + V/M$ with an epidemic occuring if and only if $R_0 > 1$. For a homogeneous population, the

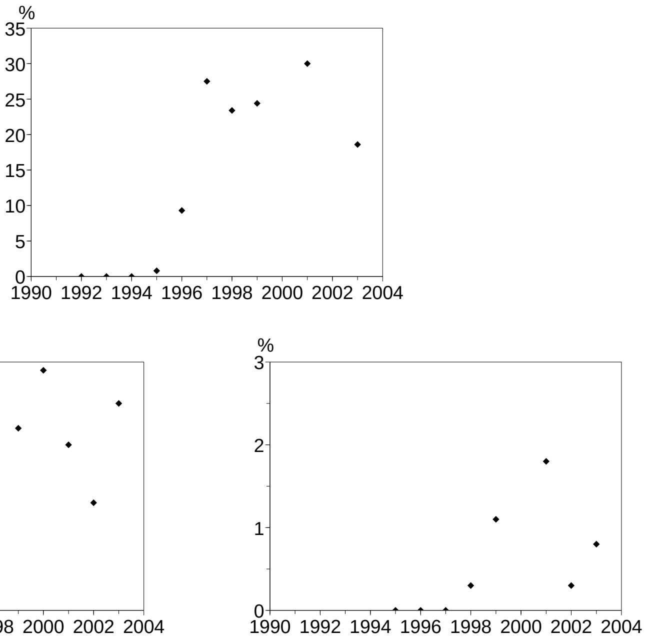

FIGURE 1. Top: HIV prevalence among drug users in Kunming. Data from [19, 20, 21, 22]. Bottom: HIV prevalence among sex workers (left) and among clients (right). Data from [13]. Notice the different scales and see [5] for more references and data.

variance V is 0 and R_0 is just proportional to the mean sexual activity M. This model can be modified to take into account continuous distributions instead of discrete distributions for the rate of acquiring new sexual partners [1, 2]. The basic reproduction number R_0 is still proportional to $M + V/M$ [6, p. 81].

The model has also been extended to heterosexual populations ([15], [4, §11.3.9]): if M and V are the mean and variance in sexual activity for men, if $\widehat{M}$ and $\widehat{V}$ are the mean and variance for women, then the basic reproduction number R_0 is proportional to

$$\sqrt{(M + V/M)(\widehat{M} + \widehat{V}/\widehat{M})}\,.$$

The variant with continuous distributions instead of discrete distributions for the rate of acquiring new sexual partners gives the same result for R_0 [6, p. 83] (notice in this reference that the formula for R_0 is obtained by considering directly the next-generation operator, that is without writing down the equations of the model).

At first, it may seem that the mean and variance from Table 2 can be used directly in the formula for R_0. In this paper we argue that this is not the case. Assuming for example a continuous distribution for the rate of acquiring new sexual partners, the idea is that people in the sexual activity group x have sex with a new partner during an infinitesimally small interval of time dt with a probability $x\,dt$ (x is a real nonnegative number). So if we ask some people belonging to the sexual activity group x how many partners they had during a period of length τ preceding the interview, we expect to get as a result a Poisson distribution with mean $x\,\tau$.

Therefore, the distribution in the population of the number of sexual partners over a given period of time is a "mixed Poisson distribution" [8, Chap. 2] (distributions of this kind have been used by insurance companies to model the number of claims of a certain type, e.g., those filed after a car accident or sickness, over a given period of time [8, Chap. 9]). If M and V are the mean and variance of the distribution in sexual activity (which is a probability distribution on the real half-line $x \geq 0$) and if $m(\tau)$ and $v(\tau)$ are the reported mean and variance of the distribution of the number of sexual partners over a period of length τ from a behavioral survey (notice that this probability distribution is on the set of nonnegative integers), one can show that $M = m(\tau)/\tau$ and $V = (v(\tau) - m(\tau))/\tau^2$, so that

$$M + \frac{V}{M} = \frac{1}{\tau}\Big(m(\tau) + \frac{v(\tau)}{m(\tau)} - 1\Big)\,. \tag{1}$$

Notice in particular that $V \leq v(\tau)/\tau^2$. The variance in the survey is always bigger than the variance of the sexual activity distribution.

For a homogeneous population, e.g., a population consisting only of a single sexual activity group x, the reported distribution would be a Poisson distribution; in such a case, $v(\tau) = m(\tau)$. Then the right side of (1) is equal to $m(\tau)/\tau$ and R_0 is just proportional to the reported mean number of partners, as one would expect.

If $\tau \to +\infty$, we have $1/\tau \to 0$ and

$$\frac{1}{\tau}\Big(m(\tau) + \frac{v(\tau)}{m(\tau)} - 1\Big) \simeq \frac{1}{\tau}\Big(m(\tau) + \frac{v(\tau)}{m(\tau)}\Big)\,.$$

So the "-1" term in (1) is important only when τ is small. But in a real survey like the one made in Yunnan, τ has to be small because the person interviewed has to be able to count (i.e., remember) all his/her sexual partners over a time interval of length τ. This question becomes harder to answer as τ increases. The designers of the survey in Yunnan estimated that $\tau = 1$ year was a long enough time frame for male clients and $\tau = 1$ week for sex workers. This is the practical way of solving one difficulty: estimate directly from a client or from a sex worker his/her sexual activity group x and construct the distribution of sexual activity in a population.

Finally, notice that an expression similar to the $m(\tau)+v(\tau)/m(\tau)-1$ on the right side of equation (1) appears also in the theory of random graphs when studying the largest connected component ([16] and [6, p. 170]).

The paper is organized as follows. In Section 2, we present a heterosexual ODE model with a continuous sexual activity structure that takes into account the turnover of males and of female sex workers (the average "working" time in Yunnan is 2.5 years [12, Table 63]). Because the data from Yunnan contain two time units (the year for the males and the week for the sex workers), a model with continuous sexual activity structure is much more appropriate than a model with discrete sexual activity structure. We give some attention to the necessary balance between the number of partnerships counted either from the point of view of males or from the point of view of sex workers. The basic reproduction number R_0 for the present model is found using an aggregation method similar to the one used in [1, 2, 15] and [4, Chap. 11]. In Section 3, we show formula (1) and use it for the data from Table 2. This way, we can get an upper bound for the geometric mean of the transmission probabilities per partnership in the context of Kunming. In Section 4, we emphasize the dependence of the reported mean $m(\tau)$ and of the reported variance $v(\tau)$ as a function of the time τ over which sexual partners are counted. Our analysis sheds some doubt on the power law found in [3], which is supposed to connect $m(\tau)$ and $v(\tau)$ independently of τ. The conclusion summarizes the aims of this paper.

2. **The model.**

The compartments of the model used in this paper are as follows (x is a real positive number):

- $S_0(t)$: number of susceptible males at time t that never have sex with sex workers;
- $S(t,x)$: density of susceptible male clients at time t such that the probability of having a new commercial sex partner during an infinitesimally small interval of time dt is $x\,dt$; these males are said to belong to the "sexual activity group x";
- $I(t,x)$: density of infected male clients in the sexual activity group x;
- $\widehat{S}(t,x)$: density of susceptible female sex workers in the sexual activity group x;
- $\widehat{I}(t,x)$: density of infected female sex workers in the sexual activity group x.

We use the term "density" because e.g., $\int_{x_1}^{x_2} S(t,x)\,dx$ is the number of susceptible male clients whose sexual activity is between x_1 and x_2. Let us write

$$N(t,x) = S(t,x) + I(t,x)\,, \quad \widehat{N}(t,x) = \widehat{S}(t,x) + \widehat{I}(t,x)$$

for the total density of male clients and for the total density of female sex workers. The parameters of the model are:

- A, $\widehat{A}(t)$: number of new males and new female sex workers respectively, entering the adult population per unit of time; notice that the recruitment of sex workers is time dependent;
- ε ($0 < \varepsilon < 1$): fraction of males that never have sex with sex workers;
- $F(x)$, $\widehat{F}(x)$: probability density that males and female sex workers respectively, enter the sexual activity group x; we assume that $\int_{0^+}^{\infty} F(x)\,dx = 1-\varepsilon$ and that $\int_0^{\infty} \widehat{F}(x)\,dx = 1$; the sexual activity distribution of adult males is therefore composed of a Dirac delta function at $x=0$ of mass ε, the remaining fraction $1-\varepsilon$ being distributed over the half-line $x>0$; all female sex workers are active ($x>0$) by definition;
- B, $\widehat{B}$: HIV transmission probability per partnership from females to males and from males to females respectively; they are averages ignoring heterogeneity in viral load and in the presence or absence of other sexually transmitted infections (these two aspects are not considered in the model because the necessary data is not available); people with different levels of sexual activity may also have different probabilities of using condoms (B and $\widehat{B}$ would then also be functions of x) but there is only limited data published on this issue in the report concerning Yunnan [11, Table 184];
- C: rate at which adult males quit the population under study because of aging;
- $\widehat{C}$: rate at which female sex workers stop selling sex;
- $D = \widehat{D}$: rate at which infected people develop AIDS; people with AIDS are considered as removed from the population.

We introduce some notations. Let $G(x)$ be the cumulative probability distribution function for the sexual activity of males:

$$G(x) = \begin{cases} 0 & \text{if} \quad x<0, \\ \varepsilon + \int_0^x F(y)\,dy & \text{if} \quad x \geq 0. \end{cases}$$

Let M and V be the corresponding mean and variance. Then

$$M = \int_0^\infty x\, dG(x) = \int_0^\infty x\, F(x)\, dx, \tag{2}$$

$$V = \int_0^\infty x^2\, dG(x) - M^2 = \int_0^\infty x^2\, F(x)\, dx - M^2\,. \tag{3}$$

Notice that the parameter ε does not appear explicitly in these formulas but it is implicit since $\int_0^\infty F(x)\, dx = 1 - \varepsilon$. We define in a similar way the mean $\widehat{M}$ and the variance $\widehat{V}$ for the sexual activity of female sex workers (without the discussion about $x = 0$). The model's equations are

$$\frac{dS_0}{dt}(t) = A\,\varepsilon - C\, S_0(t) \tag{4}$$

for the males who do not have sex with female sex workers,

$$\frac{\partial S}{\partial t}(t,x) = A\, F(x) - C\, S(t,x) - B\, x\, S(t,x)\, \frac{\int_0^\infty y\, \widehat{I}(t,y)\, dy}{\int_0^\infty y\, \widehat{N}(t,y)\, dy}\,, \tag{5}$$

$$\frac{\partial I}{\partial t}(t,x) = B\, x\, S(t,x)\, \frac{\int_0^\infty y\, \widehat{I}(t,y)\, dy}{\int_0^\infty y\, \widehat{N}(t,y)\, dy} - (C + D)\, I(t,x) \tag{6}$$

for the other males, and

$$\frac{\partial \widehat{S}}{\partial t}(t,x) = \widehat{A}(t)\, \widehat{F}(x) - \widehat{C}\, \widehat{S}(t,x) - \widehat{B}\, x\, \widehat{S}(t,x)\, \frac{\int_0^\infty y\, I(t,y)\, dy}{\int_0^\infty y\, N(t,y)\, dy}\,, \tag{7}$$

$$\frac{\partial \widehat{I}}{\partial t}(t,x) = \widehat{B}\, x\, \widehat{S}(t,x)\, \frac{\int_0^\infty y\, I(t,y)\, dy}{\int_0^\infty y\, N(t,y)\, dy} - (\widehat{C} + \widehat{D})\, \widehat{I}(t,x)\,, \tag{8}$$

for the female sex workers. We assume that initially

$$S_0(0) = A\,\varepsilon/C\,, \quad N(0,x) = A\, F(x)/C\,, \quad \widehat{N}(0,x) = A\, M\, \widehat{F}(x)/(C\, \widehat{M})\,, \tag{9}$$

$0 \le S(0,x) \le N(0,x)$ and $0 \le \widehat{S}(0,x) \le \widehat{N}(0,x)$ for all $x > 0$. From (4), one can see that $S_0(t) = A\,\varepsilon/C$ for all t. Since the number of sexual partnerships counted by male clients should be equal to the number of sexual partnerships counted by female sex workers, we should have

$$\int_0^\infty x\, N(t,x)\, dx = \int_0^\infty x\, \widehat{N}(t,x)\, dx \tag{10}$$

for all t. In order to satisfy this constraint, we assume somewhat artificially from now on that the recruitment of sex workers $\widehat{A}(t)$ is given by the formula

$$\begin{aligned}\widehat{A}(t)\, \widehat{M} = A\, M + (\widehat{C} - C) \int_0^\infty x\, S(t,x)\, dx + (\widehat{C} - C - D) \int_0^\infty x\, I(t,x)\, dx \\ + \widehat{D} \int_0^\infty x\, \widehat{I}(t,x)\, dx\,,\end{aligned} \tag{11}$$

meaning that the recruiting people can balance supply and demand exactly at any time. Indeed, it follows from equations (5)-(8) and from (11) that

$$\begin{aligned}\frac{d}{dt}\int_0^\infty x\Big(N(t,x)-\widehat{N}(t,x)\Big)\,dx &= \int_0^\infty x\Big(\frac{\partial N}{\partial t}(t,x)-\frac{\partial \widehat{N}}{\partial t}(t,x)\Big)\,dx\\ &= A\,M - C\int_0^\infty x\,N(t,x)\,dx - D\int_0^\infty x\,I(t,x)\,dx\\ &\quad -\Big[\widehat{A}(t)\,\widehat{M}-\widehat{C}\int_0^\infty x\,\widehat{N}(t,x)\,dx-\widehat{D}\int_0^\infty x\,\widehat{I}(t,x)\,dx\Big]\\ &= -\widehat{C}\int_0^\infty x\Big(N(t,x)-\widehat{N}(t,x)\Big)\,dx\,. \end{aligned}\tag{12}$$

Given the choice of initial conditions (9), we have

$$\int_0^\infty x\Big(N(0,x)-\widehat{N}(0,x)\Big)\,dx = 0\,. \tag{13}$$

It follows from (12) and (13) that the balance equation (10) is satisfied for all t.

If we assume also that $\widehat{C} > C + D$, then system (5)-(8) with $\widehat{A}(t)$ defined by (11) can be shown to be well-posed with a nonnegative $\widehat{A}(t)$ for all t. The meaning is that the turnover of sex workers should be fast enough for the recruitment to adapt to variations in offer and demand caused by AIDS mortality. We will see in the next section that the data for Yunnan satisfy the condition $\widehat{C} > C + D$. Let us emphasize some of the main assumptions of the present model.

- Duration of partnerships is not taken into account. This is a common assumption in mathematical models of commercial sex. To make it more precise, we consider the data for the "mean frequency of having sexual intercourse with a commercial partner over the last 30 days" for those men who had commercial sex in the past year (i.e., 21.4% of all men according to Table 1 above): the mean frequency is 2.4 for one month [11, Table 174]. Given that the mean number of commercial partners for all the males is 1.0 over one year (Table 2 above), we can estimate the mean number of commercial sex partners over one year (call it μ) for those men who had commercial sex in the past year because $1.0 = 0\times 78.6\% + \mu\times 21.4\%$. We get $\mu = 1.0/0.214 \simeq 4.67$. The mean number of commercial sexual relationships for these clients over one year can be estimated by multiplying the monthly number of sexual relationships by the number of months in one year: $2.4 \times 12 = 28.8$. So we get an average of $28.8/4.67 \simeq 6.2$ sexual relationships with the same sex worker and an average duration of "partnership" equal to $6.2/2.4 \simeq 2.6$ months (with an average of 2.4 relationships per month). This time scale is much smaller than, for example, the average "working" time $1/C'$ of sex workers (2.5 years according to [12, Table 63]), so the initial assumption is somewhat justified.
- Mixing between sexual activity groups is assumed to follow proportionate mixing because we have no better information from the behavioral survey.
- As the epidemic develops, it is assumed that the recruitment of sex workers is adapted to keep the balance between offer and demand. There are several other possibilities. For example, men can start visiting sex workers less often. These assumptions do not enter into the computation of what is the main focus of the present paper, namely the basic reproduction number R_0, which depends only on the very beginning of the epidemic.

The disease-free steady state of system (5)-(8) and (11) is given by

$$S_*(x) = A\,F(x)/C, \quad \widehat{S}_*(x) = A\,M\,\widehat{F}(x)/(C\,\widehat{M}), \tag{14}$$

$I(x) = 0$, $\widehat{I}(x) = 0$ for all $x > 0$. Linearizing equations (6) and (8) near this disease-free steady state and changing to lower case letters to avoid confusion, we obtain

$$\frac{\partial i}{\partial t}(t,x) = B\,x\,S_*(x)\,\frac{\int_0^\infty y\,\widehat{i}(t,y)\,dy}{\int_0^\infty y\,\widehat{S}_*(y)\,dy} - (C+D)\,i(t,x) \tag{15}$$

$$\frac{\partial \widehat{i}}{\partial t}(t,x) = \widehat{B}\,x\,\widehat{S}_*(x)\,\frac{\int_0^\infty y\,i(t,y)\,dy}{\int_0^\infty y\,S_*(y)\,dy} - (\widehat{C}+\widehat{D})\,\widehat{i}(t,x)\,. \tag{16}$$

Using the same method as the one used in [4, §11.3.9] for a model with discrete distributions of sexual activity, we introduce the aggregated variables

$$J(t) = \int_0^\infty x\,i(t,x)\,dx, \quad \widehat{J}(t) = \int_0^\infty x\,\widehat{i}(t,x)\,dx\,.$$

It follows from (15)-(16) that

$$\frac{dJ}{dt}(t) = B\frac{\int_0^\infty x^2\,S_*(x)\,dx}{\int_0^\infty x\,\widehat{S}_*(x)\,dx}\,\widehat{J}(t) - (C+D)\,J(t)$$

$$\frac{d\widehat{J}}{dt}(t) = \widehat{B}\frac{\int_0^\infty x^2\,\widehat{S}_*(x)\,dx}{\int_0^\infty x\,S_*(x)\,dx}\,J(t) - (\widehat{C}+\widehat{D})\,\widehat{J}(t)\,.$$

Given (2)-(3) and (14), this system can be rewritten as

$$\frac{dJ}{dt}(t) = B\,\frac{V+M^2}{M}\,\widehat{J}(t) - (C+D)\,J(t) \tag{17}$$

$$\frac{d\widehat{J}}{dt}(t) = \widehat{B}\,\frac{\widehat{V}+\widehat{M}^2}{\widehat{M}}\,J(t) - (\widehat{C}+\widehat{D})\,\widehat{J}(t)\,. \tag{18}$$

One can then easily show that the zero steady state of this last system is linearly stable if and only if

$$R_0 = \sqrt{\frac{B\,\widehat{B}(M+V/M)(\widehat{M}+\widehat{V}/\widehat{M})}{(C+D)\,(\widehat{C}+\widehat{D})}} < 1\,. \tag{19}$$

Let us add one comment about this formula. One could have arrived at the same result without including in the model the males who never have sex with sex workers, i.e., by restricting our attention to the clients of sex workers. Indeed, the distribution of sexual activity for these clients is $\widetilde{F}(x) = F(x)/(1-\varepsilon)$ and the mean sexual activity is $\widetilde{M} = M/(1-\varepsilon)$. Notice that $\widetilde{M} > M$. But if $\widetilde{V}$ is the corresponding variance, then

$$M + \frac{V}{M} = \frac{\int_0^\infty x^2 F(x)\,dx}{\int_0^\infty x\,F(x)\,dx} = \frac{\int_0^\infty x^2 \widetilde{F}(x)\,dx}{\int_0^\infty x\,\widetilde{F}(x)\,dx} = \widetilde{M} + \frac{\widetilde{V}}{\widetilde{M}}\,.$$

So we get exactly the same R_0.

3. **Link with the data.** People in the sexual activity group x have a number of partners over a period of length τ which is distributed according to a Poisson distribution with a mean equal to $x\,\tau$ [7, §XVII.2]. So the probability of these people having j partners over a period of length τ is

$$e^{-x\tau}\,(x\tau)^j/j!$$

It follows that the fraction of males who declare j partners over a period of length τ in a behavioral survey should be

$$f_j(\tau)=\int_0^\infty e^{-x\tau}\,\frac{(x\tau)^j}{j!}\,dG(x)=\begin{cases}\varepsilon+\int_{0+}^\infty e^{-x\tau}\,F(x)\,dx & \text{if } j=0,\\ \int_0^\infty e^{-x\tau}\,\frac{(x\tau)^j}{j!}\,F(x)\,dx & \text{if } j\geq 1.\end{cases}\qquad(20)$$

In other words, $(f_j(\tau))_{j\geq 0}$ is a "mixed Poisson distribution" [8, Chap. 2]. The reported mean $m(\tau)$ and variance $v(\tau)$ are given by

$$m(\tau)=\sum_{j\geq 1} j\,f_j(\tau)\,,\quad v(\tau)=\sum_{j\geq 1} j^2\,f_j(\tau)-m(\tau)^2\,.\qquad(21)$$

It follows from (2), (20), and (21) that the reported mean $m(\tau)$ is equal to τM (see also [8, Prop. 2.1(i)]):

$$m(\tau)=\int_0^\infty e^{-x\tau}\sum_{j\geq 1}\frac{(x\tau)^j}{(j-1)!}\,dG(x)=\tau\int_0^\infty x\,dG(x)=\tau M\,.\qquad(22)$$

Besides, it follows from (20) that

$$\sum_{j\geq 1} j(j-1)f_j(\tau)=\int_0^\infty e^{-x\tau}\sum_{j\geq 2}\frac{(x\tau)^j}{(j-2)!}\,dG(x)=\tau^2\int_0^\infty x^2\,dG(x)\,.\qquad(23)$$

Combining the expression (3) of the variance V with (21)-(22)-(23), we see that

$$\begin{aligned}V&=\frac{1}{\tau^2}\sum_{j\geq 1} j(j-1)f_j(\tau)-\Big(\frac{m(\tau)}{\tau}\Big)^2\\&=\frac{1}{\tau^2}\sum_{j\geq 1} j^2\,f_j(\tau)-\frac{1}{\tau^2}\sum_{j\geq 1} j\,f_j(\tau)-\frac{m(\tau)^2}{\tau^2}=\frac{v(\tau)-m(\tau)}{\tau^2}\end{aligned}\qquad(24)$$

(see also [8, Prop. 2.1(ii)]). Notice here that since $m(\tau)\geq 0$ and $V\geq 0$, any "mixed Poisson distribution" has two important properties: $v(\tau)\geq m(\tau)$ (the inequality being strict if $V>0$) and $v(\tau)\geq \tau^2\,V$ (the inequality being strict if $m(\tau)>0$, i.e., if $M>0$). Finally, (22) and (24) yield

$$M+\frac{V}{M}=\frac{m(\tau)}{\tau}+\frac{(v(\tau)-m(\tau))/\tau^2}{m(\tau)/\tau}=\frac{1}{\tau}\Big(m(\tau)+\frac{v(\tau)}{m(\tau)}-1\Big)\,.\qquad(25)$$

Similarly, we obtain for the sex worker report extending over a period τ' the formula

$$\widehat{M}+\frac{\widehat{V}}{\widehat{M}}=\frac{1}{\widehat{\tau}}\Big(\widehat{m}(\widehat{\tau})+\frac{\widehat{v}(\widehat{\tau})}{\widehat{m}(\widehat{\tau})}-1\Big)\,.\qquad(26)$$

Remarks:

- From the expressions (2)-(3) for M and V, we see that the "physical" dimension (or unit) of x is $[\text{time}]^{-1}$, that of M is $[\text{time}]^{-1}$, that of V is $[\text{time}]^{-2}$, that of τ is [time], but m and v are dimensionless (they are reported numbers). One can check that both the right side and the left side of formula (25) have the same dimension.

- If $\widehat{F}(x)$ is a Gamma distribution as in [1, 2], then it can be shown that the associated discrete distribution $(\widehat{f}_j(\widehat{\tau}))_{j\geq 0}$, given by

$$\widehat{f}_j(\widehat{\tau}) = \int_0^\infty e^{-x\widehat{\tau}}\, \frac{(x\widehat{\tau})^j}{j!}\, \widehat{F}(x)\, dx\,,$$

  is a negative binomial distribution [8, p. 17]. This point, already shown by Greenwood and Yule [10], does not seem to have been noticed before in the literature on sexual activity models for HIV/AIDS.

Now, we would like to apply formulas (19), (25), and (26) to the data from Table 2 to obtain an estimate of R_0. The problem is that some parameters are not well known. On one side, we know the rate $\widehat{C}$ at which sex workers stop selling sex: since the average "working" time is 2.5 years [12, Table 63], we can take $\widehat{C} = 1/2.5 = 0.4$ per year. The percentage of adult males who had sex with a female sex worker during the past year is relatively constant as a function of age (Table 3). So for the parameter C, we will keep the mean aging rate between the ages of 18 and 50, i.e., $C = 1/(50-18) = 1/32$ per year. For the progression rate from HIV infection to AIDS, we take for simplicity a constant $D = \widehat{D} = 1/10$ per year, giving an average incubation period of 10 years. Notice that $\widehat{C} > C + D$, as required in the previous section.

Table 3. Percentage of men who had sex with a sex worker during the past 12 months (from [11, Table 173]).

age	18-30	30-34	35-39	40-44	45-50
had sex	25.6%	25.2%	20.2%	19.1%	8.8%

However, problems arise from the uncertainty surrounding the transmission probabilities per partnership B and $\widehat{B}$. We estimated in section 2 that partnership in the present context of commercial sex means on average 6 sexual contacts. A study of HIV transmission between sex workers and clients in Thailand [14] suggested that the transmission probability during one sexual contact from female to male could be as high as 3% because of the high prevalence of other sexually transmitted infections among sex workers. Other studies have found much lower transmission probabilities per contact, e.g. 0.11% for HIV-discordant couples in Uganda [9]. It is difficult to say which estimate between these two extreme values would be appropriate for Kunming. In addition, we remark that condom use is included by some kind of averaging procedure inside the parameters B and $\widehat{B}$.

Given these difficulties, we turn the question the other way around: knowing that we do not yet see so far an exponential increase in sexual transmission of HIV, what can we infer about the unknown parameters of the model? From $R_0 < 1$ and from (19), (25), and (26), we obtain an upper bound for the geometric mean of these averaged transmission probabilities per partnership:

$$\begin{aligned}\sqrt{B\,\widehat{B}} &< \sqrt{\frac{\tau\,\widehat{\tau}\,(C+D)\,(\widehat{C}+\widehat{D})}{(m+\frac{v}{m}-1)(\widehat{m}+\frac{\widehat{v}}{\widehat{m}}-1)}} \\ &= \sqrt{\frac{1\times(7/365)\times(0.4+0.1)\times(1/32+0.1)}{(1.0+\frac{2.2^2}{1.0}-1)(3.0+\frac{4.1^2}{3.0}-1)}} \simeq 0.58\%\,.\end{aligned}$$

Under the extra simplifying assumption $B = \widehat{B}$ and considering (see section 2) that a partnership represents an average of 6 sexual contacts, we can get an upper bound for the transmission probability per sexual contact (call it b): $b < 0.1\%$. Indeed, $1-(1-0.1\%)^6 \simeq 0.6\%$. So our result is closer to the lower estimates for the transmission probability per sexual contact, but this may be due to a high level of condom use. In the behavioral survey made in 2002 in Yunnan, 73.5% of clients said they used a condom the last time they had sex with a sex worker [11, Table 177].

If we could observe an exponential increase in sexual transmission, we would be able to get an estimate of the geometric mean $\sqrt{B\,\widehat{B}}$ since the growth rate λ of the model is the largest eigenvalue of the matrix on the right side of system (17)-(18)

$$\lambda = \frac{-(C+\widehat{C}+D+\widehat{D})+\sqrt{(C-\widehat{C})^2+4B\widehat{B}(M+V/M)(\widehat{M}+\widehat{V}/\widehat{M})}}{2}$$

and is related to the doubling time T by $T = \log 2/\lambda$. The main point here is that M and V cannot be read directly from the data but that instead we must use $m(\tau)$ and $v(\tau)$ with the extra "-1" term from formulas (25)-(26).

4. **Relationship between mean and variance.** In [3], it was argued that

> a wide range of surveys of different populations employing different sampling methods and various time intervals for recall, reveal a remarkably consistent trend in the relationship between the mean and variance in the rate of acquisition of new partners. The summary statistics are related by a power law, $v = a\,m^b$, where a and b are constants.

The constants found were $a = 0.555$ and $b = 3.231$. We notice that in the framework of our model, this power law cannot hold for "various time intervals for recall [...] ranging from the past month to lifetime". Indeed, the time interval for recall is τ in the present notations. Since equations (22) and (24) imply that $m(\tau) = \tau\,M$ and $v(\tau) = \tau^2\,V + \tau\,M$, and since $v(\tau)$ is not a homogeneous function of τ (there is no α such that $v(s\tau) = s^\alpha\,v(\tau)$ for all $s > 0$), the mixed Poisson distribution from the present model seems incompatible with the simple scaling law of [3]. In other words, even if the point representing the reported mean and reported variance over a fixed time of recall lies very close to the power-law curve (as is the case for the sex worker data from Yunnan, see Fig. 2), the point representing the mean and variance for the same population but for a different time of recall may be far from the power-law curve. This is illustrated in Fig. 2. We refer to [3, Fig. 2a] for the original cloud of points, which was supposed to justify the power law. For an easier comparison, the scale of Fig. 2 is the same as the one used in [3, Fig. 2a]. With the model used in the present paper, we notice that $m(\tau) = \tau\,M$ and $v(\tau) = \tau^2\,V + \tau\,M$ imply that $v(\tau)/m(\tau)^2$ is approximately constant (i.e. independent of τ) for large values of τ and that $v(\tau)/m(\tau)$ is approximately constant for small values of τ.

5. **Conclusion.** The influence of mathematical models has been analyzed in [17] in the following way[1]:

> The results of several modeling efforts, especially those of Roy Anderson and colleagues [...] have shown that the rate of partner change is one of the key factors influencing the speed and size of the epidemic [...]

[1]Notice that the author of [17] is vice president of Futures Group. This may explain why the survey made in Yunnan and Sichuan by Futures Group Europe reports both the mean and the variance for the distribution of the number of sexual partners.

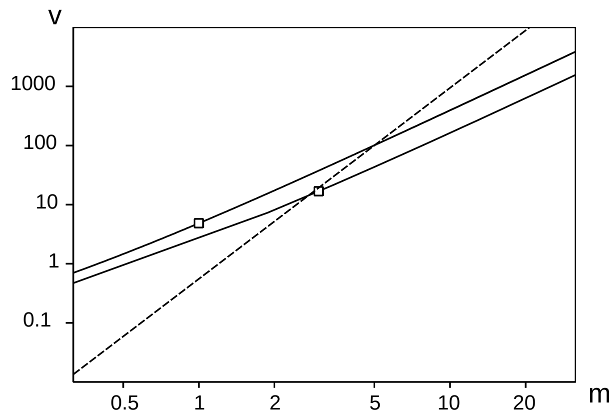

FIGURE 2. Reported variance $v(\tau)$ as a function of the reported mean number of partners $m(\tau)$ (logarithmic scale). Solid lines: $(m(\tau) = \tau\, M,\ v(\tau) = \tau^2\, V + \tau\, M)$ with (M, V) fixed and a varying interval for recall τ. Dots: data from Table 2 (males on the left, female sex workers on the right). Dashed line: power law $v = a\, m^b$ from [3].

> Although there is little evidence that this understanding has influenced program design to any great extent, it has certainly influenced research and evaluation efforts. Several of the key prevention indicators developed by GPA[2], UNAIDS, and USAID are designed to measure rates of partner change and concurrent partnerships. If these indicators are seriously applied, they will eventually influence program decisions by showing which interventions improve these indicators and which do not affect them.

In this paper, we have tried to "apply seriously" the indicator $M + V/M$ to real data from Yunnan. We have found simple relations between the theoretical parameters (M, V) on one side, the reported mean $m(\tau)$ and the reported variance $v(\tau)$ for a period of time τ over which sexual partners are reported on the other side. It turns out that the expression for the indicator $M + V/M$ in terms of $m(\tau)$ and $v(\tau)$ has a form similar to an important parameter in the theory of random graphs. Our formulation has also shed some doubt on the mysterious power law found in [3].

Acknowledgements. Authors were supported by the "Programme de Recherches Avancées de Coopérations Franco-Chinoises" (PRA SI05-01, "Modélisation de l'épidémie de VIH/SIDA dans les provinces du Yunnan et du Xinjiang").

REFERENCES

[1] R. M. Anderson and R. M. May, *The invasion, persistence and spread of infectious diseases within animal and plant communities*, Philos. Trans. R. Soc. Lond. B, **314** (1986), 533–570.

[2]GPA: Global Program on AIDS.

[2] R. M. Anderson, G. F. Medley, R. M. May, and A. M. Johnson, *A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS*, IMA J. Math. Appl. Med. Biol., **3** (1986), 229–263.

[3] R. M. Anderson and R. M. May, *Epidemiological parameters of HIV transmission*, Nature, **333** (1988), 514–519.

[4] R. M. Anderson and R. M. May, "Infectious Diseases of Humans: Dynamics and Control," Oxford University Press, 1991.

[5] N. Bacaër, X. Abdurahman, and J. Ye, *Modeling the HIV/AIDS epidemic among injecting drug users and sex workers in Kunming, China*, Bull. Math. Biol., **68** (2006), 525–550.

[6] O. Diekmann and J. A. P. Heesterbeek, "Mathematical Epidemiology of Infectious Diseases," John Wiley & Sons, New York, 2000.

[7] W. Feller, "An Introduction to Probability Theory and its Applications, vol. I, 3rd ed.," John Wiley & Sons, New York, 1968.

[8] J. Grandell, "Mixed Poisson Processes," Chapman & Hall, London, 1997.

[9] R. H. Gray, M. J. Wawer, R. Brookmeyer et al., *Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda*, Lancet, **357** (2001), 1149-1153.

[10] M. Greenwood and G. U. Yule, *An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurence of multiple attacks of disease or of repeated accidents*, J. Roy. Statist. Soc., **83** (1920), 255–279.

[11] Horizon Market Research and Futures Group Europe, "2001 Behavioural Surveillance Survey in Yunnan and Sichuan: Adult Male Report," December 2002, http://www.futuresgroup.com/Documents/2001BSSadultmale.pdf.

[12] Horizon Market Research and Futures Group Europe, "2001 Behavioural Surveillance Survey in Yunnan and Sichuan: Sex Worker Report," December 2002, http://www.futuresgroup.com/Documents/2001BSSsexworker.pdf.

[13] L. Lu, M. Jia, X. Zhang et al., *Analysis for epidemic trend of acquired immunodeficiency syndrome in Yunnan Province of China*, Chin. J. Prev. Med., **38** (2004), 309–312.

[14] T. D. Mastro, G. A. Satten, T. Nopkesorn et al., *Probability of female-to-male transmission of HIV-1 in Thailand*, Lancet, **434** (1994), 204–207.

[15] R. M. May and R. M. Anderson, *Transmission dynamics of HIV infection*, Nature, **326** (1987), 137–142.

[16] M. Molloy and B. Reed, *The size of the giant component of a random graph with a given degree sequence*, Combin. Probab. Comput., **7** (1998), 295–305.

[17] J. M. Stover, *Influence of mathematical modeling of HIV and AIDS on policies and programs in the developing world*, Sex. Transm. Dis., **27** (2000), 572–578.

[18] Y. Xiao, S. Kristensen, J. Sun et al., *Expansion of HIV/AIDS in China: Lessons from Yunnan Province*, Soc. Sci. Med., **64** (2007), 665–675.

[19] H. Yu, X. An, M. Jia et al., *Report on HIV/AIDS surveillance in Yunnan province in 1999*, J. Chin. AIDS/STD Prev. Cont., **7** (2001), 74–76.

[20] J. Zhang, H. Cheng, M. Jia et al., *Ten years of experience on AIDS control in Yunnan (1989∼1998)*, Chin. J. Epidemiol., **20** (1999), 377–380.

[21] X. Zhang, Y. Ma, H. Yu et al., *Analysis on the surveillance result of HIV/AIDS in Yunnan in 2001*, Ji Bing Jian Ce, **17** (2002), 327–330.

[22] X. Zhang, J. Lu, L. Fu et al., *Analysis on survey results of HIV/AIDS in Yunnan Province in 2003*, Ji Bing Jian Ce, **19** (2004), 409–412.

Received on April 6, 2007. Revised on August 1, 2007.

E-mail address: bacaer@bondy.ird.fr
E-mail address: axamxi@xju.edu.cn
E-mail address: yejianli@chinawch.org.cn
E-mail address: auger@bondy.ird.fr