Modeling HIV Epidemic Under Contact Tracing — The Cuban Case

H. DE ARAZOZA¹*, R. LOUNES¹‡, T. HOANG³ and Y. INTERIAN³

¹Departamento Ecuaciones Diferenciales, Facultad Mat-Comp, Universidad de la Habana, San Lázaro y L, Habana 4 Cuba; ²Université René Descartes, Laboratoire de Statistique Médicale, UPRES A CNRS 8071, 45 rue des Saints-Pères 75006 Paris-France

(Received 22 November 1998; In final form 9 August 1999)

A nonlinear model is developed for an epidemic with contact tracing, and its dynamic is studied. A linear version of the model is presented in both deterministic and stochastic versions. We present the data for the Cuban HIV/AIDS epidemic and fit the linear model to the data, we obtain estimates for the size of the Cuban HIV epidemic.

Keywords: HIV/AIDS epidemic, contact tracing, epidemic control, deterministic path, birth and death process

1 INTRODUCTION

Contact tracing has been used as a method to control endemic contagious diseases [9,10]. While there is still a debate about contact tracing for the HIV infection [4,15] the resurgence of infectious tuberculosis and outbreaks of drug-resistant tuberculosis secondary to HIV induced immunodepression is forcing many public health departments to reexamine this policy [2,7]. A model of the HIV epidemic allowing for contact tracing would help evaluate the effect of this method of control on the size of the HIV epidemic. The HIV/AIDS data for Cuba were used to develop this model. The model provides a unique test for validating such model.

In Cuba the first AIDS case was diagnosed in April 1986. Some HIV seropositives had been detected at the end of 1985. Earlier the Cuban Government had started preventive measures in order to contain the possible outbreak of the epidemic. One of the measures was a total ban on the import of blood and blood byproducts. Once the first cases were confirmed, a program based on the experience with other sexually transmitted diseases was initiated. This program had among other actions the tracing of sexual contacts of known HIV seropositives to prevent the spreading of the virus. Contact tracing was carried out by the Epidemiology Departments at all levels of the Cuban Health System through partner notification and interviews. With this case in mind we present models for the HIV epidemic in Cuba, that take into account detection of HIV positives long before they become AIDS cases. We estimate the size of the HIV epidemic, the percentage of known

*Corresponding Author: E-mail: arazoza@matcom.uh.cu
‡E-mail: lounes@citi2.fr and hoang@citi2.fr
HIV positives, and discuss the effect of contact tracing on the size of the epidemic. We develop a four class nonlinear deterministic model, a linear version of this nonlinear model, and a stochastic version of the linear model. We present the data for the Cuban HIV/AIDS epidemic and fit the models to the data. All data presented is for the period 1986–1997.

2 THE CUBAN HIV-AIDS DATA

There are 707 diagnosed AIDS cases in Cuba, 527 males and 180 females. Of the males 74% are homosexuals (we consider the group of homosexuals to be formed by the homosexuals and the bisexuals). There have been 481 deaths due to AIDS. Through the program a total of 1831 HIV positives have been found, 477 female and 1354 males. Of the males 75% are homosexuals.

Contact tracing allowed to find 35.1% of the HIV positives. Table I gives the new HIV/AIDS cases detected by years.

The epidemic is a small one, with a population of around 11 millions the cumulative incidence rate for AIDS is 65 per million. The epidemic is mainly localized in 5 provinces: the City of Havana with the neighboring Havana Province account for 51.5% of the HIV positives, Pinar del Rio 7.4%, Villa Clara 14.5% and Sancti Spiritus 5%. The first three are in the western part of the island, the others in the centre of the island. The epidemic is mostly autonomous with 12% infected outside Cuba. The most affected age groups is the 15–35 years of age.

The epidemic started as a mainly heterosexual one but the homosexual component has increased in time. Vertical transmission has been very reduced (6 cases), as all pregnant women are tested at the beginning of their pregnancy and given the option of an abortion. There are only 8 cases of infection by transfusions, and 2 hemophiliacs. These are the result of the early measures taken by the Cuban Government.

A Kaplan-Meier estimation for the incubation period, taking 899 HIV positives with a known probable date of infection, gives a mean at 9.2 years with a standard deviation of 0.273 and the median at 8.42 years, with 95% confidence interval [7.5,8.83]. Some patients receive medication that helps to increase the incubation period, but this does not change the incubation period significantly.

Kaplan-Meier estimation for the survival probability, gives an average of 2.28 years with a standard deviation of 0.128 and the median at 1.08 years with 95% confidence interval [1,1.33].

3 THE NONLINEAR MODEL WITH FOUR CLASSES

In Cuba when a person is detected as HIV positive she/he is informed of the results of her/his test and follows counseling so that she/he is well aware of his condition as a possible source of infection. Depending on a first evaluation she/he continues her/his regular life or enters a sanatorium with relaxed access and visitation for further counseling evaluation and treatment [14]. Consequently to this provision of health care, the HIV positives do not, in general, infect other susceptibles.

We consider the following variables

N the population of sexually active persons or susceptibles,
X the number of HIV positives that do not know they are infected,
Y the number of HIV positives that know they are infected,
MODELING HIV EPIDEMIC

Z the number of AIDS cases, with transitions between the variables given by the following diagram where

\( \alpha \) is the rate at which susceptibles become unknown HIV positives as result of their infection by the individuals in class \( X \),

\( k \) the rate at which the HIV positives are detected,

\( \beta \) the rate at which the HIV positives develop AIDS,

\( \mu \) the death rate of the sexually active population,

\( \mu' \) the death rate due to AIDS,

\( \delta \) the recruitment into the class of susceptibles,

\( \nu \) the immigration of unknown HIV positives,

\( \rho \) the immigration of AIDS cases.

The diagram reflects the fact that the infection is carried out only by the HIV positives that are not aware of their infectiousness. The differential system that describes the model is:

\[
\begin{align*}
\frac{dN}{dt} &= -\alpha NX - \mu N + \delta, \\
\frac{dX}{dt} &= \alpha NX - (k + \mu + \beta)X + \nu, \\
\frac{dY}{dt} &= kX - (\mu + \beta)Y, \\
\frac{dZ}{dt} &= \beta X + \beta Y - \mu' Z + \rho.
\end{align*}
\]

System (1) is a nonlinear system. Since the immigration constants will have no influence on the behavior of the equilibria but only on the values, we will study the following system

\[
\begin{align*}
\frac{dN}{dt} &= -\alpha NX - \mu N + \delta, \\
\frac{dX}{dt} &= \alpha NX - (k + \mu + \beta)X, \\
\frac{dY}{dt} &= kX - (\mu + \beta)Y, \\
\frac{dZ}{dt} &= \beta X + \beta Y - \mu' Z + \rho.
\end{align*}
\]

with two equilibria at

\[
\left( \frac{\delta}{\mu}, 0, 0, 0 \right),
\]

which is the no epidemic case, and

\((N_0, X_0, Y_0, Z_0)\)

where

\[
N_0 = \frac{k + \mu + \beta}{\alpha}, \quad X_0 = \frac{\delta}{k + \mu + \beta} - \frac{\mu}{\alpha}, \quad Y_0 = \frac{k}{\mu + \beta} X_0, \quad Z_0 = \frac{\beta}{\mu' \mu} \left( \frac{1}{\mu} + \frac{k}{\mu + \beta} \right) X_0.
\]

The basic reproduction number \([3]\) of the model is:

\[
R_0 = \frac{\alpha \delta}{\mu' (k + \mu + \beta)}.
\]

The techniques used in the study of the stability of the equilibria are similar to those used in other cases in the literature so we will only give the results with some comments. We will give the basic result on the global stability of the equilibria of system (2), restricted to the space

\( \mathcal{N} = \{(N, X, Y, Z), N \geq 0, X \geq 0, Y \geq 0, Z \geq 0\} \).

Notice that in (2), the first two equations function as an independent two-dimensional system for the variables \( N \) and \( X \).

**Theorem 1** Consider the system (2) restricted to the region \( \mathcal{N} \), then

1. If \( R_0 < 1 \) then there is only one equilibrium point in \( \mathcal{N} \) and it is globally asymptotically stable. This equilibrium point corresponds to the epidemic free equilibrium.

2. If \( R_0 > 1 \) then there are two equilibria in \( \mathcal{N} \) and one is globally asymptotically stable, the other is unstable. The stable one corresponds to an endemic level of the epidemic, and the unstable one is the epidemic free one.

The proof of the theorem is based on the stability of the corresponding equilibria for the system formed by the first two equations of (2) in the restriction of \( \mathcal{N} \), and the fact that you can integrate the last two equations of (2), obtaining expressions in terms of \( X(t) \), and then find the limits of the trajectories in the full space.

**4 DETERMINISTIC AND STOCHASTIC LINEAR MODEL**

The nonlinear system is very hard to fit taking into account that the system is in general "stiff" and that
the number of points available is small. Also as the epidemic is still at an early stage, we can consider that the nonlinear term of the equations can be linearized by taking \( N \) as constant and substituting \( aN \) with \( A \) in (1) to obtain a new linear system that we will fit to the Cuban HIV/AIDS data.

Consider the same variables as in the previous section and the following system of equations

\[
\begin{align*}
\frac{dX}{dt} &= (\lambda - k - \mu - \beta)X + \nu, \\
\frac{dY}{dt} &= kX - (\mu + \beta)Y, \\
\frac{dZ}{dt} &= \beta X + \beta Y - \mu'Z + \rho, 
\end{align*}
\]

where \( k, \beta, \mu', \nu \) and \( \rho \) are the same as in the previous section, \( \lambda \) is the rate at which new HIV positives are produced by the individuals in class \( X \) and \( \mu \) is the death rate of the HIV positive population. Let \( \sigma = \lambda - k - \mu - \beta \) and \( \gamma = \mu + \beta \).

The solutions to this system with initial conditions

\[
X(0) = x_0, \quad Y(0) = y_0, \quad Z(0) = z_0
\]

are given by

\[
\begin{align*}
X(t) &= \frac{\sigma x_0 + \nu}{\sigma} e^{\sigma t} - \frac{\nu}{\sigma}, \\
Y(t) &= \frac{k(x_0 + \nu)}{\sigma(\lambda - k)} e^{(\lambda - k) t} + \frac{k \nu}{\sigma \gamma} e^{-\gamma t} - \frac{k \nu}{\sigma \gamma} e^{-\gamma t}, \\
Z(t) &= \frac{\beta \lambda(x_0 + \nu)}{\sigma(\sigma + \mu')} e^{\sigma t} + \frac{\beta}{\mu'} \Omega e^{-\gamma t} \\
&\quad + \left\{ z_0 + \frac{k \beta (k + \gamma)}{\mu' \sigma \gamma} - \frac{\beta}{\mu'} - \frac{\beta \lambda(x_0 + \nu)}{\sigma(\sigma + \mu')} \Omega e^{-\gamma t} \right\} e^{-\mu' t} \\
&\quad + \frac{\rho}{\mu'} - \frac{\nu \beta (k + \gamma)}{\mu' \sigma \gamma}.
\end{align*}
\]

where

\[
\Omega = \left[ y_0 + \frac{k \nu}{\sigma \gamma} - \frac{k(x_0 + \nu)}{\sigma(\lambda - k)} \right].
\]

The equilibrium point is

\[
\left( \frac{-\nu}{\sigma}, \frac{k \nu}{\sigma \gamma}, \frac{\rho}{\mu'}, \frac{\nu \beta (k + \gamma)}{\mu' \sigma \gamma} \right).
\]

It follows from (4) that the eigenvalues are \( \sigma, -\gamma \) and \( -\mu' \). If \( \sigma \) is negative the equilibrium point is stable and the solutions tend to an endemic level given by the values of the variables at the equilibrium point.

We will now present a stochastic version of this linear model (4). Let \( (X(t), Y(t), Z(t), : t \geq 0) \) be a process such that conditional to \( X(t) = x, Y(t) = y \) and \( Z(t) = z \), the infinitesimal transition probabilities are:

\[
\begin{align*}
P[X(t + dt) = x + 1, Y(t + dt) = y, Z(t + dt) = z] &= (\lambda x + \nu) dt + o(dt) \\
P[X(t + dt) = x - 1, Y(t + dt) = y, Z(t + dt) = z] &= k x dt + o(dt) \\
P[X(t + dt) = x, Y(t + dt) = y + 1, Z(t + dt) = z] &= \beta x dt + o(dt) \\
P[X(t + dt) = x, Y(t + dt) = y, Z(t + dt) = z + 1] &= \mu x dt + o(dt) \\
P[X(t + dt) = x, Y(t + dt) = y - 1, Z(t + dt) = z + 1] &= \beta y dt + o(dt) \\
P[X(t + dt) = x, Y(t + dt) = y - 1, Z(t + dt) = z] &= \mu y dt + o(dt) \\
P[X(t + dt) = x, Y(t + dt) = y, Z(t + dt) = z + 1] &= \rho dt + o(dt) \\
P[X(t + dt) = x, Y(t + dt) = y, Z(t + dt) = z - 1] &= \mu' z dt + o(dt) \\
P[X(t + dt) = x, Y(t + dt) = y, Z(t + dt) = z] &= 1 - (\lambda + \gamma + k) x + \mu' z + \gamma y + \nu + \rho dt + o(dt).
\end{align*}
\]

In this case we consider the same interpretation of the variables. In our case \( Y(t) \) is the class of the known HIV seropositives, as some of the new cases detected are found through contact tracing (35%), it can be argued that the infected persons are not observed independently of each other. The coefficient \( k \) in the equations is the rate at which new seropositives are found through the program, this includes contact...
tracing and screening programs, as for example, the testing of all pregnant women and blood donors. Taking this into account we will consider our model as an approximation of a more complex situation. Parameter $\beta$ is the inverse of the incubation period for AIDS, our modeling implies that the incubation period follows an exponential distribution. This is not true in general, but in our case when we fit the Kaplan-Meier curve of the incubation period to a gamma distribution we obtain an exponential distribution, thus for the moment our hypothesis is valid.

Let $\Psi(u, v, w, t)$ be the probability generating function of the trivariate process $(X(t), Y(t), Z(t))$ with $(X(0) = x_0, Y(0) = y_0, Z(0) = z_0)$. Then $\Psi(u, v, w, t)$ verifies

$$
\frac{\partial \Psi}{\partial t} = [\lambda u^2 + kv - (\lambda + \gamma + k)u + \beta w + \mu] \frac{\partial \Psi}{\partial u}
+ [\gamma v + \beta w + \mu] \frac{\partial \Psi}{\partial v}
+ \mu'(1-w) \frac{\partial \Psi}{\partial w} + [nu - 1 + \rho(w - 1)]\Psi (8)
$$

and $\Psi(u, v, w, 0) = u^{x_0} v^{y_0} w^{z_0}$. The general solution of (8) is complicated by the presence of stage $Z(t)$. Our main interest is in the process $(X(t), Y(t), : t \geq 0)$, by a similar method as in [12] and [13] we can find the expression for $\Psi$ for the first two components of the process $(X(t), Y(t), Z(t), : t \geq 0)$:

$$
\Psi(u, v, t) = [1 - (1 - v)e^{-(\lambda+\gamma)t}]u^{x_0} \Phi(u, v, t)^{y_0} \times \exp \left\{ \int_0^t \left[ \Phi(u, v, t - \tau) - 1 \right] d\tau \right\}
$$

where $\Phi$ satisfies

$$
\frac{\partial \Phi}{\partial t} = [\lambda u^2 + kv - (\lambda + \gamma + k)u + \gamma] \frac{\partial \Phi}{\partial u}
+ (1 - v)\gamma \frac{\partial \Phi}{\partial v}
$$

and $\Phi(u, v, 0) = u$.

As in [13], we can find the expression of $\Phi$ and then that of the two marginals of $\Phi$. We can also study the behaviour of $(X(t), t \geq 0)$ and $(Y(t), t \geq 0)$ for the cases $t \to +\infty, \lambda \to 0, \gamma \to 0, v \to 0$. We do not give the expressions because this is not the scope of this paper. For the trivariate process $(X(t), Y(t), Z(t), : t \geq 0)$, we will consider diffusion approximation for the model in a future paper.

Let $K(a, b, c, t)$ be the cumulant-generating function of the trivariate process. Expanding $K(a, b, c, t)$ in powers of $a, b, c$ and using (8) we obtain

$$
\frac{\partial K}{\partial t} = [\lambda e^a + ke^{(b-a)} + \beta e^{(c-a)} + \mu(t)e^{-a} - \lambda - \gamma - k]
\times \frac{\partial K}{\partial a} + \mu'(e^{-a} - 1) \frac{\partial K}{\partial c}
+ [\mu e^{-b} + \beta e^{(c-b)} - \gamma] \frac{\partial K}{\partial b}
+ \nu(e^{-a} - 1) + \rho(e^{-1})
$$

with initial condition $K(a, b, c, 0) = x_0a + y_0b + z_0c$, hence the system for the first two moments

$$
\begin{align*}
\frac{d}{dt} E(X(t)) &= \sigma EX(t) + \nu, \\
\frac{d}{dt} E(Y(t)) &= kEX(t) - \gamma EY(t), \\
\frac{d}{dt} E(Z(t)) &= \beta EY(t) + \beta EX(t) - \mu' EZ(t) + \rho, \\
\frac{d}{dt} Var(X(t)) &= 2\sigma Var(X(t)) \\
&+ (\lambda + k + \gamma)EX(t) + \nu, \\
\frac{d}{dt} Var(Y(t)) &= -2\gamma Var(Y(t)) + kEX(t) \\
&+ \gamma EY(t) + 2kCov(X(t), Y(t)), \\
\frac{d}{dt} Var(Z(t)) &= 2\beta Cov(X(t), Z(t)) + \beta EX(t) \\
&+ 2\beta Cov(Y(t), Z(t)), \\
&+ \beta EY(t) - 2\mu' VarZ(t) \\
&+ \mu' EZ(t) + \rho, \\
\end{align*}
$$

and

$$
\begin{align*}
\frac{d}{dt} Cov(X(t), Y(t)) &= (\sigma - \gamma)Cov(X(t), Y(t)) \\
&+ kVar(X(t)) - kEX(t), \\
\frac{d}{dt} Cov(X(t), Z(t)) &= (\sigma - \mu')Cov(X(t), Z(t)) \\
&+ \beta Var(X(t)) - \beta EX(t), \\
&+ \beta Cov(X(t), Y(t)), \\
\frac{d}{dt} Cov(Y(t), Z(t)) &= -(\gamma + \mu')Cov(Y(t), Z(t)) \\
&+ \beta Cov(X(t), Y(t)) \\
&+ \beta Cov(X(t), Z(t)).
\end{align*}
$$
The first three equations of (10) form an autonomous system similar to (4), with solutions that can be obtained from (5), (6) and (7) substituting $X$, $Y$ and $Z$ by their expectations $EX$, $EY$ and $EZ$; the remaining equations of (10) can be solved. Some of the expressions are cumbersome, but we are only interested in the expression for $VarX(t)$ which is given by

\[ VarX(t) = \frac{1}{\sigma} \left\{ (A + B)e^{2\sigma t} - Ae^{\sigma t} - B \right\} \] (11)

where

\[ A = (\lambda + k + \gamma) \left( \frac{\sigma \mu + \nu}{\sigma} \right), \]
\[ B = \frac{\nu}{2} \left( 1 - \frac{1}{\sigma} \right). \]

5 APPLICATION TO THE CUBAN HIV/AIDS DATA

We will use the models (4) and (10) and their solutions to fit the data for the known HIV positives in Cuba. We will use the following values for the parameters estimated from the available data:

- $X(0) = 230$, this was estimated from the number of HIV positives that were found after 1986 and were already infected in 1986.
- $Y(0) = 94$, number of detected HIV positives that were alive at the end of 1986.
- $Z(0) = 3$, number of AIDS cases that were alive at the end of 1986.
- $\mu = 0.0053$, yearly death rate for the HIV+ cases for 1991-1997. ($S.D. = 0.00254$).
- $\mu' = 0.3027$, yearly death rate for the AIDS cases for 1986-1997. ($S.D. = 0.07$).
- $\nu = 20$, number of imported HIV positives (average for 1986-1997).
- $\beta = 0.10870 = 1/9.2$, obtained from the average incubation period.
- $\rho = 0$, there are practically no imported AIDS cases.

Using least squares method, we fit formula (6) to the number of known HIV positives and we obtain the remaining parameters of the systems, i.e., $\lambda = 0.5594$ and $k = 0.4554$. (Figure 2)

Using (11) to compute an interval for $EX(t)$, we find that the number of unknown HIV positives in Cuba for 1997 is within the limits (342, 486).

We can also use the standard deviation of $\beta$, $\mu$ and $\mu'$ to get an alternative estimate for $EX$ as shown in Table II.

As we are considering a markovian process, the coefficient $k$ can be interpreted as $1/(\text{mean time to detection})$ and we can estimate that the time between infection and detection is between 2.16 and 2.23 years.

6 DISCUSSION

Parameter $k$ plays an important role in the dynamics of the epidemic, it is the inverse of the mean time from infection to detection. Other parameters in the model can not be changed in a short term. Parameter $k$ depends on the efficiency of the contact tracing system and the various screening processes of special
population groups, and can vary according to how well the different levels of the Cuban AIDS program do their job. Recall that

$$R_0 = \frac{\alpha \delta}{\mu(k + \mu + \beta)} < 1$$  \hspace{1cm} (12)$$

is the condition for the stability of the no epidemic point for system (2). How valid is this condition for the Cuban AIDS data? Assume $\alpha = \frac{A}{N}$, where $N$ is the sexually active population (15 years old or more) estimated to be 8.8 millions, take $\delta$ as the average recruitment into the class of the sexually active population for the last ten years, 150,000 and 8.66 per thousand for $\mu$ the death rate of the sexually active Cuban population [1]. Then we have $R_0 = 2.06$. Therefore according to model (2) there is an HIV/AIDS epidemic in Cuba.

If the value of $k$ is increased to the value 1.0643, then the value of $R_0$ would be less than one, this would mean that the average time for a seropositive from infection to detection would be reduced to less than 11 months.

The recruitment into the class of the sexually active population, $\delta$, shows a decreasing tendency due to a decrease in birth rate. If this trend continues and the value for $\delta$ reaches a level of 100,000 per year then if the mean time from infection to detection is kept at an average of less than 18 months, then the value of $R_0$ would drop below 1 and the epidemic would be controlled.

Even in the epidemic case, variations in the values of $k$ will influence the number of infected persons in the future. Table III presents predictions based on models (10) and (4) for the HIV epidemic in Cuba in the year 2000 according to the variations of the value of $k$.

If we look at Table I there is a drop in the number of new HIV infections detected in the years 1993–95, this was due in our opinion to the difficult economic

<table>
<thead>
<tr>
<th>Change in $k$</th>
<th>Total HIV+ in the year 2000</th>
<th>Change in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 %</td>
<td>1683</td>
<td>-3.5</td>
</tr>
<tr>
<td>+10 %</td>
<td>1641</td>
<td>-4.7</td>
</tr>
<tr>
<td>+20 %</td>
<td>1603</td>
<td>-6.8</td>
</tr>
<tr>
<td>+30 %</td>
<td>1568</td>
<td>-8.7</td>
</tr>
<tr>
<td>+50 %</td>
<td>1507</td>
<td>-10.4</td>
</tr>
<tr>
<td>+70 %</td>
<td>1447</td>
<td>-12.4</td>
</tr>
<tr>
<td>+90 %</td>
<td>1394</td>
<td>-14.5</td>
</tr>
<tr>
<td>-30 %</td>
<td>1834</td>
<td>+9</td>
</tr>
<tr>
<td>-50 %</td>
<td>1962</td>
<td>+16.5</td>
</tr>
</tbody>
</table>
conditions in Cuba at that time. If we follow our model and the effect that a change in the coefficient $k$ produces in the future of the epidemic, as seen in Table 111, then the consequences of this reduced detection in the years 1993–95 should produce an increase in the number of new infections in the future. This is already evident in the years 1996–97 with 597 new HIV positives detected and for the year 1998 the number of new cases detected is 362.

An increase of 50% on the coefficient $k$ means that the mean time to detect a seropositive is reduced by 8 months from 26 to 18 months. Nonparametric estimation for the time of the detection of a seropositive made in 1993 gave a value between 18 and 20 months, this value has increased since, but no reliable estimation can be made yet because, for the seropositives recently detected, the time of infection is very difficult to determine. By giving more resources to the program (this is being done since 1997) a 50% increase of $k$ could achieve a reduction of the epidemic of HIV by 10.4% in the year 2000.

Acknowledgment

The authors acknowledge partial support from the Fondation pour la Recherche Médicale (Sidaction), France.

The authors would like to thank the referees for their suggestions and comments.

References:

Submit your manuscripts at http://www.hindawi.com