



MODELLING FOR HIV SPREAD IN HOMOSEXUAL POPULATION

T. Srivenkataramana* & C. Nagaraja Rao**

ABSTRACT

Statistical modelling of HIV infection is useful to understand HIV spread mechanism, for predicting HIV and AIDS counts and for forecasting healthcare needs. This article:

- i. stresses the need for modelling of HIV and AIDS*
- ii. discusses the appropriateness of a stochastic model for HIV transmission and*
- iii. examines the applicability of diffusion of news and rumors model of Taga & Isii (1959), reported in Bartholomew (1967) and presents the adapted version of this model to suit HIV spread in homosexual populations. Three different cases are considered under this model.*

Keywords: AIDS, Incubation time, HIV infection, Homosexual population, Mixing, Modelling, Spreaders, Susceptibles.

* Professor of Statistics, Bangalore University, Bangalore - 560 056.

** Professor of Statistics, Vijaya College, Bangalore - 560 004.

1. Introduction

It is more than two decades since the early cases of AIDS were recognized. In many countries, HIV/AIDS represents the deadliest emergency and the greatest social, economic and health crisis of modern times. The AIDS endemic is increasingly female, young and poor. While HIV affects people of all ages, half of all newly infected people are 15 to 24 years old. The impact of this endemic goes beyond the lives of infected people. The disease changes community dynamics, undermines the structure of the family and threatens the future of children. Homosexuals form the second most serious HIV high risk group followed by heterosexuals. In many parts of the USA, homosexual transmission is the single largest mode of HIV spread. The spread virus in this group is of considerable interest for planning interventions.

Modelling of HIV Spread

Epidemic modelling is an exciting, active and rapidly expanding field. Models based on transmission mechanisms of AIDS virus can help medical and scientific community to understand and anticipate its spread. Since AIDS is expected to place enormous demand on healthcare facilities, there is a pressing need to develop flexible prediction models. These are an integral part of long range scientific research and are equivalent to the hypotheses to be tested (Bailey, 1990). Modelling has been in use for a long time (Farr, 1840; Hamer, 1906; Ross, 1911; McKendrick, 1926; Kendall, 1956 and Bailey, 1950, 53). Mollison et al., (1994) have extensively reviewed epidemic models. They discuss the choice among alternative intervention strategies which play a crucial role in quantifying uncertainties of AIDS projection.

In developing a model one creates a logical structure to organize existing information into a framework and suggests what new data that must be collected. Thus, models provide insight into the systems even when data are lacking. Hence statistical models are essential for understanding AIDS and the relation between social and biological mechanisms that influence its spread.

Deterministic theory of HIV epidemic

Incidence of HIV seroprevalence and AIDS counts is a large scale phenomenon. The variability due to large number of susceptibles and infectives on such a phenomenon has to be reduced. We assume that, for a given number of susceptibles and infectives and for given rates of contact, infection and removals, certain definite numbers of new HIV and AIDS cases will result during a specified period. A deterministic approach as developed by Kendall (1956) has been used to evaluate probability of HIV infection in heterosexual relations and for estimating HIV infections in perinatal transmission by Srivenkataramana and Rao (2002).

2. Need for Stochastic Modelling

Many a phenomena in life sciences are now studied not only as random but also as changing with time or space. For example, the growth of HIV infected population over time may be considered to be a stochastic process. The growth of HIV in blood is a determining factor to confer seropositivity. Once this occurs, its progress to AIDS is characterized by several stages associated with depletion of CD_4^+ cells. This forms a discrete space continuous time stochastic process providing a basis for modelling of AIDS. Moreover, as AIDS epidemiological data become more extensive and on occasions deal with smaller risk groups than those relevant for large areas, the elements of random variation become more prominent.

McKendrick (1926) was the first to give a stochastic treatment of an epidemic process. Excellent reviews on the current modelling work on AIDS have been provided by May and Anderson (1987), Anderson et al. (1988), Isham (1988), Bailey (1988) and Mollison et al. (1994). There are several advantages with stochastic modelling as compared to the deterministic approach. For instance, i) the former provides more information, i.e., besides expected values one can compute variances and covariances and assess the effects of various factors, ii) stochastic models along with computers, provide a wider choice in formulating the underlying assumptions and model structure, and iii) Monte Carlo methods may be used to assess whether solutions of some non-linear difference equations are in fact measures of central tendency for random functions of the epidemic process.

3. The Model

The spread of HIV in a heterogeneous population with different mixing pattern is a phenomenon of interest and importance. We examine below the applicability of Taga & Isii (1959) stochastic model for the diffusion of news and rumors which is a special case of the pure birth process, to understand the HIV spread phenomenon among homosexuals. There are many similarities between these two phenomena. In news diffusion model, information is transmitted by the members of the group from a source either at an initial point in time or continuously. The source may be a television, a newspaper, a hoarding or a group of people introduced into the population from outside. Persons who receive information may become spreaders themselves by transmitting information to others whom they meet. Similar is the situation in homosexual population in the context of HIV spread.

The system which we study may be described as follows: There is a closed population of N homosexuals who are at risk of HIV infection. This represents the set of all

adults of the same sex in a given region. The virus is transmitted to the members of this group from a source at some point in time or through homosexual transmission. The sources of infection considered are infected needle, infected blood and other contaminated material. Persons who receive the virus from the source may become spreaders by transmitting the infection to others in the population with whom they mix. The virus spread continues until all in the population are HIV infected. From practical viewpoint, the characteristic of interest is the total number HIV infected at any time t . In this type of HIV transmission, the stochastic element enters at two points. Whether or not a given person gets infected depends on i) the person coming into contact with the source or a spreader and ii) on the infection being transmitted when contact is established. Here neither (i) nor (ii) is a sure event and hence the HIV spread is a stochastic phenomenon. We consider the model with three different cases with respect to the mixing of members in a pair.

3.1. Case (A) : Perfect mixing

In this the assumptions of the model are that (i) all members are equally exposed to the source and (ii) all pairs of members have equal likelihood of acquiring infection. Assumption (i) is not crucial unless the intensity of transmission from source is large compared to the intensity of transmission between the members in a pair. The second assumption is unlikely to hold for human populations in the context of HIV. We later examine the effect of relaxing assumption (ii). First we define two random variables as follows:

$N(t)$: size of the homosexual population at time t which is closed for migration.

$H(t)$: number of persons HIV infected at time t .

t_H : time taken for the number infected to reach level H .

Consider two events as follows:

E_s : {transmission of infection from the source to any member of the population}.

E_p : {transmission of infection by a spreader}.

Assume that the probabilities of these events are proportional to the length of time intervals.

Thus

$$P\{E_s \text{ in } (t, t + \partial t)\} = \alpha \cdot \partial t, (\alpha > 0) \text{ and } P\{E_p \text{ in } (t, t + \partial t)\} = \beta \cdot \partial t, (\beta > 0),$$

where α and β are the intensities of transmission from the source and between the individuals in a pair respectively. It is also assumed that all transmissions are independent of each other. The system is said to be in state H when exactly H persons have received the virus. Here the stochastic process is, $P\{H \rightarrow (H + 1) \text{ in } (t, t + \partial t)\} = \lambda_H \partial t; (\lambda_H \geq 0)$. The infection rate, λ_H , is treated as a parameter of the model. When the total number infected is H , there are $(N - H)$ persons not yet infected. Thus the total contribution to λ_H from the source is $(N - H) \alpha \cdot \partial t$. Also, of all the possible pairs that could be formed in the given population, there are $H(N - H)$ pairs which consist of one infected and one susceptible. Therefore, the total probability associated with these pairs for $H \rightarrow (H + 1)$ is $H(N - H)\beta \cdot \partial t$.

Combining the two contributions, we get

$$\lambda_H = (N - H) (\alpha + \beta H); \quad (H = 0, 1, 2, \dots, N - 1) \quad (1)$$

Thus, in epidemic theory, the source consists of one or more persons who spread the virus. If one person starts the epidemic, we have to put $\alpha = \beta$ and then eqn (1) reduces to

$$\lambda_H = \beta(H + 1) (N - H); \quad (H = 0, 1, 2, \dots, N - 1) \quad (2)$$

Analysis of the model:

Assuming Poisson law for infections to take place, the expressions (Bailey, 1957) for the distribution of $H(T)$ is given by $P(H(t) = 0) = e^{-\lambda_0 t}$ and

$$P(H(t) = H) = \prod_{i=0}^{H-1} \lambda_i \sum_{i=0}^H \frac{e^{-\lambda_i t}}{\prod_{\substack{j=0 \\ j \neq i}}^H (\lambda_j - \lambda_i)} \quad (H = 0, 1, 2, \dots, N - 1) \quad (3)$$

Since λ_i can be found from eqn. (1) in terms of α , β and N , the problem is solved in principle. The feature of the distribution, which is of particular interest, is the mean, $\bar{H}(t)$.

The plotting of $\bar{H}(t)$ for a given population as a function of t gives the expected development in the HIV infection process. If we plot the derivative of $\bar{H}(t)$, we get the rate at which the HIV infection spreads. This may be called the 'HIV epidemic curve', which gives a clear picture of the growth of the epidemic. The expressions for $\bar{H}(t)$ and its derivative are obtained by Haskey (1954) and are reported in Bailey (1957). At time t , if r susceptibles are left uninfected, then changing time scale to $\tau = \beta t$ instead of t for simplicity, the stochastic mean is

$$\bar{H}(\tau) = \sum_{r=1}^{n-1} nCr \{ (n-2r+1)^2 \tau + 2 - (n-2r+1) \sum_{u=r}^{n-r} u^{-1} \} e^{-r(n-r+1)\tau}$$

where $n = N - H$ and r runs to $n/2$ if n is even and up to $(n+1)/2$ if n is odd.

3.2. Case B: Imperfect mixing

The assumption that all pairs of members have equal probability of virus transmission may not be fully justified in human populations. Therefore it is of interest to investigate the effect of relaxing this assumption. As an extreme case, suppose that there is no contact at all among members of the population. The diffusion is thus entirely attributable to the source. This case is covered by the above model and is obtained by setting $\beta = 0$. Then we have the pure birth process and the distribution of $H(t)$ turns out to be binomial with

$$P\{H(t) = H\} = \binom{N}{H} (1 - e^{-\alpha})^H e^{-\alpha t n}; \quad n = 0, 1, \dots, N$$

Here $\bar{H}(t)$ reduces to $N(1 - e^{-\alpha t})$ and the epidemic curve becomes, $\frac{d\bar{H}(t)}{dt} = N\alpha e^{-\alpha t}$. In this, unlike in case A, the rate of spread declines continuously with time instead of first rising to a maximum.

Duration of the epidemic

The time taken for a proportion P of the population to get infected is obtained as

$$E(t_H) = \frac{1}{\alpha} \sum_{i=1}^H \frac{1}{N-i+1} \text{ considering the fact that } t'_H \text{'s are partial sums of the } t \text{'s.}$$

If N is large and $P = H/N$ is not near 1, we get, $E(t_H) = -\left(\frac{1}{\alpha}\right) \log_e(1-P)$.

It may be noted that expected time taken to reach a given proportion does not depend on N unlike in the case A where P will be reached faster for a large N .

3.3. Case C : Spreaders active for a random time

In the pure birth process model in the context of epidemic theory, it is assumed that the spreaders of infection continue to do so indefinitely. As a consequence of this, all members of the population will eventually get infected. But, in HIV transmission,

the infected after knowing their status or after onset of AIDS symptoms, generally stop the risk prone activities and may cease to be spreaders. The other reasons for this include, a) social isolation, as there is a social stigma associated with the disease, b) disinterest in spreading, and c) death due to AIDS.

A comprehensive model considering all possibilities is difficult to develop. Thus the model proposed here considers only reason (iii) and assumes that the spreaders are active only for a random period of time T . As this period is reported to be increasing steadily over time starting with mean T of 4.5 years, the number of new infections will be larger in the course of time. Thus T may be assumed to follow exponential distribution with mean μ^{-1} , where μ is the rate of infected who cease to be spreaders.

The description of the model is as follows: The population of size N consists of infected persons and susceptibles. At any time t , let i = number actively spreading HIV and s = number of susceptibles. Then, $N - (i+s)$ is the number of persons who are infected but have ceased to be spreaders. Since N is fixed, the state of the process at any time t is defined by the values of any two of these variables. For example, the system is in state (i,s) if $i(t) = i$ and $s(t) = s$. Our interest is to know the development of the process in the course of time. In order to do this, we consider the joint probability distribution of i and s , viz.

$$\Pr \{ i(t) = i, s(t) = s \} = P_{i,s}(t).$$

From the state (i,s) , two transitions are possible during the interval $(t, t+\Delta t)$. They are:

i) $(i,s) \rightarrow [(i+1), (s-1)]$ with probability $s(\alpha+\beta i) \Delta t$ for $s = 1, 2, \dots, N$;

$$i = 0, 1, \dots, N-1 \text{ such that } 0 \leq (i+s) \leq N.$$

ii) $(i,s) \rightarrow [(i-1), s]$, with probability $i\mu \Delta t$. Transition (i) takes place when spreader mixes with a susceptible and transition (ii) occurs when a spreader ceases to spread the virus. Using these transition probabilities and relating the joint probability at time $(t+\Delta t)$ to that at t , the following bivariate differential - difference equations for $P_{i,s}^1(t)$ are obtained :

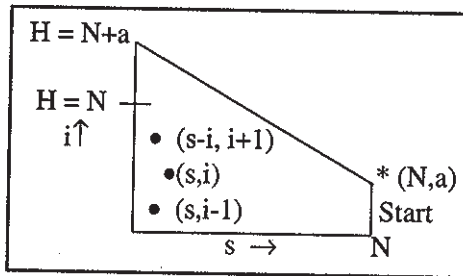
$$P_{i,s}^1(t) = -s \{s(\alpha+\beta i) + \mu i\} P_{i,s}(t) + (s+1) (\alpha+\beta (i-1)) P_{s+1, i-1}(t) + (i+1) \mu P_{s, i+1}(t).$$

$$P_{N,0}^I(t) = -N\alpha P_{N,0}(t) \quad (4)$$

and $P_{N,0}(0) = 1$, where P^I denotes the derivative.

Now, the problem is to solve the set of equations (4). Here if $\mu = 0$ there is no retirement for the spreaders. If $\beta = 0$ or $\mu = \infty$, retirement is irrelevant as no one is ever actively spreading the virus and $i(t) = 0$, always. When $\mu = \infty$, the model reduces to pure birth process with $\beta = 0$. Although the system is difficult to solve, it is easy to find the limiting values of the probabilities $P_{(i,s)}(t)$ as $t \rightarrow \infty$. That is, after a long period, the spread of HIV stops either because every one is infected or all the spreaders have ceased to be active. In either case, $i(\infty)$, the final number of spreaders, is zero with probability 1. Consequently, $P_{(i,s)}(\infty) = 0$ if $i > 0$.

When $i = 0$, $P_{(i,0)}(\infty)$ will be the probability distribution of the terminal number of susceptibles. This can be obtained using embedded random walk over the lattice point (s, i) , which is Markovian since the transition probabilities depend only on the present state of the system.



When the system is in state (s, i) , the two transitions which it can make and their associated probabilities are:

$$i) \quad (i,s) \rightarrow (i+1), s-1 = s / (s + \bar{f})$$

$$ii) \quad (i,s) \rightarrow (s, i-1) = \bar{f} / (s + \bar{f}),$$

where \bar{f} = relative removal rate

$$= \frac{\text{average time taken for a randomly chosen pair to meet}}{\text{average length of time for which a spreader is active}}$$

A large value of \bar{f} indicates that the spreaders cease their activity relatively rapidly and a small value implies the reverse situation.

4. Discussion

The AIDS endemic has raised a number of important health policy issues. Some of these have been raised by scientific advances and others by the societal implications of AIDS. Modelling is the best approach to identify future policy considerations. Modelling should take care of data requirements to know the prevalence of infected people in different population groups by suggesting special studies. We must focus attention on persons with high-risk behavior such as homosexuals. Thus stochastic modelling is handy to study the HIV spread mechanism in a homosexual group. The virus spread phenomenon in homosexuals resembles the diffusion process of news described in Bartholomew (1976) as an extension of pure birth process. Sec.3 modifies this model to suit a homosexual population and considers three cases. Out of these, case C is particularly relevant to the AIDS situation, since HIV spreaders are active for a random period of time, that is, up to diagnosis as an AIDS case. After this the spread activity is very likely to stop.

Case A of perfect mixing assumes equal exposure of all members to the source and also that all pairs have same probability of infection. On the other hand case B examines the extreme situation of infection only from the source and no infection derived from the spreaders. The ground reality is an intermediate situation of infections directly from the source and also from the spreaders. Accordingly the assumptions will have to be modified and the probability expressions derived. In this set-up the probabilities from cases A and B can serve as upper and lower bounds for the actual infection probability.

One way could be on the lines similar to the estimation of probability of HIV infection in the context of heterosexual contact, described by Srivenkataramana and Rao(2002), by evaluating the effective number of pairs giving rise to the new infection.

References

- Anderson, R. M. (1988). The epidemiology of HIV infection plus infectious periods and heterogeneity of sexual activity. *JRSS-A*, 15.
- Bailey, N. T. J. (1988). Simplified modelling of the population dynamics of HIV/AIDS. *JRSS-A*, 151(1), 31-43.
- Bailey, N. T. J. (1950). A simple stochastic epidemic. *Biometrika*, 37, 193-202.
- Bailey, N. T. J. (1953). The total size of general stochastic epidemic. *Biometrika*, 40, 177-185.
- Bailey, N. T. J. (1957). *Math. Theory of epidemics*, London: Griffin.

- Bailey N. T. J. (1990). Application of stochastic epidemic modelling in the public health control of HIV/AIDS. In Gabriel, J-P, Lefevre, C. and Picard P. (eds). *Stochastic Processes in Epidemic Theory. Lecture Notes in Bio-Mathematics*, 86, 14-20.
- Bartholomew, D. J. (1967). *Stochastic models for social processes*. John Wiley & Sons Ltd, London
- Farr, W. (1840), *progress of epidemics*, second report of the Registrar General of England and Wales, 91-98.
- Hamer, W. H. (1906). Epidemic diseases in England. *Lancet*, 1, 733-739.
- Haskey, H. W. (1954). A general expression for the mean in a simple stochastic epidemic. *Biometrika*, 41, 272-275.
- Isham, V (1988). Mathematical modelling of the transmission dynamics of HIV infection and AIDS: a review. *JRSS-A.*, 151, 5-30.
- Kendall, D.G. (1956). Deterministic and stochastic epidemics in closed populations. *Proc. Third Berkeley Symposium on Maths. Statistics, University of California*, 4, 149-165.
- May, R. M. and Anderson R. M. (1987). Transmission dynamics of HIV infection. *Nature*, 326(2), 137-142.
- McKendrick, A. G. (1926). Application of maths to medical problems. *Proc. Edin. Math. Soc.*, 44, 98-130.
- Mollison, D., Isham, V., Grenfell, B. (1994). Epidemic models and data. *JRSS-A*, 157(1), 115-149.
- Ross, R. (1911). *The preventive of malaria (2nd Ed.)*, London: Murray.
- Srivenkataramana, T & Rao, C. N. (2002). HIV spread: some statistical results. *Mapana, Journal of Sciences, Christ College, Bangalore, India, Vol. 1, No. 1, 28-39*
- Taga, Y and Isii (1959). On a stochastic model concerning the pattern of communication-Diffusion of news in a social group, *Ann. Inst. Statist. Math*, 11, 25-43.