CHAPTER III

REVIEW OF LITERATURE

3.1 Introduction

The statistical studies on HIV throughout the world have focused on modeling and methodology of HIV epidemic. From the beginning of the AIDS epidemic in developed countries, statistical and mathematical models have played a considerable role in a variety of areas of research on HIV. They include the analysis of epidemiological patterns and study of the natural history and clinical course of HIV related disease, the prediction of future trends and the design of trials and programmes for intervention and surveillance. There are basically two different types of modeling approaches: the deterministic and stochastic. Deterministic models assume as the number of susceptible, infected, and AIDS cases in the HIV epidemic are deterministic functions of time, ignoring completely randomness of these variables and randomness of all risk factors. These models are usually described by a system of difference or differential equations.

Stochastic models assume that the variables are a family of random variables indexed by time. Since the nature of HIV epidemic is basically stochastic and many variables are subject to stochastic variations, stochastic
models are more realistic than deterministic models. Therefore the purpose of the study is to provide systematic development of stochastic models for HIV transmission. Large number of studies have been made available on HIV pathogenesis and HIV epidemic, the therapies, strategies and HIV incidence, survival probability, progression of the disease, HIV incidence from prevalence, incubation distribution, transmission probabilities, and long-term and short-term projections across the world. The literature is rather very enormous; and it would be a formidable task to review all such results contributed by various authors. Hence it is proposed to give a brief outline of some chosen papers in this chapter.

The research papers which are taken up for review may be broadly classified into the following headings.

i. Models relating to the transmission and dynamics of HIV spread.

ii. Models relating to the distributions of random variables (Latency period, Incubation period, Seroconversion time).

iii. Models relating to infectivity, projection and other aspects.

3.2 Models Relating to the Transmission and Dynamics of HIV Spread

Anderson (1988) discusses on the epidemiology of HIV infection taking into consideration the variable incubation periods plus infectious period and the heterogeneity in sexual activity. The author considers that a homosexual
community of size N can be divided into $X =$ number of susceptible, $Y =$ number of infected and $A =$ AIDS patients ($N = X + Y + A$). The simplest set of equations that are used for the transmission of HIV in a closed community (i.e. no influx of susceptible from outside and no deaths other than those who die due to AIDS) is given by

$$\frac{dX}{dt} = -\lambda X$$

$$\frac{dY}{dt} = -\lambda X - \nu Y$$

$$\frac{dA}{dt} = -\nu Y - \alpha A$$

$$\frac{dN}{dt} = -\alpha X$$

Where the parameter $\nu = 1/D$ defines the rate of leaving the infected class $Y$ to join the AIDS class $A$, where $D$ is the average duration of stay in $Y$. The rate of defines mortality in the AIDS class.

Here $\lambda$ is the per capita force of infection defined as

$$\lambda = \frac{\beta CN}{N}$$

where $C$ is the average number of sexual partners per unit of time, and $\beta$ is the probability of infecting a susceptible partner. The expression for $R_0$ which is
the average number of secondary cases of infection is obtained as \( R_0 = \beta CD \) with the assumption that the infected individuals are infectious throughout the duration of their stay in class Y. The expression for doubling time \( t_d \), is given by

\[
    t_d = t \left[ \ln(2) / (R_0 - 1) \right]
\]

The author has also discussed various extensions of this model taking into consideration variable incubation, infection periods and also the variations in the rate of sexual partners changed.

The use of mathematical models to understand the AIDS epidemic is by Hymann and Stantley (1988). The authors have indicated how the various factors, qualitative and quantitative, should be incorporated into a mathematical model that is used for depicting the spread of AIDS epidemic. The concepts like population risk structure, the sexual activities and its impact on risk, the drug use etc., have been discussed in detail in this paper. In addition to this the authors have given some simplified mathematical models for spread of the epidemic.

Isham (1988) has given a review of the mathematical modeling of the transmission dynamics of HIV infection and AIDS. In this paper a simple epidemic model of deterministic type is discussed initially. A fixed population of size \( n \) is separated into two groups namely a group of \( X(t) \) the susceptible
and \( Y(t) \) infectives where \( X(t) + Y(t) = n \) and \( X(t), Y(t) \) are sufficiently large so that they can be regarded as continuous variables. Assuming the population mixes homogeneously so that in any small time interval \((t, t+\delta)\) the number of contacts between susceptible and an infective is proportional to both \( X(t) \) and \( Y(t) \) (and \( \delta \)) and that a fixed proportion of these contacts results in the susceptible becoming infected. It has been shown that the number of new cases of infection in the time intervals is \( \alpha X(t) \), \( Y(t) \), \( \delta \) for some \( \alpha \) which is a constant of proportionality, \( X(t) \) satisfies the differential equations

\[
dX(t)/dt = -\alpha X(t)/Y(t) = -\alpha X(t) [n - X(t)]
\]

and the solution of this equation is given by

\[
X(t) = nX(0) \{X(0) + [n - X(0)] \exp (n\alpha t)}^{-1}
\]

dy(t)/dt is the rate at which the new infections occur; and it can be obtained from the solution of the differential equations given above. Assuming \( X(0) \approx n \), for \( t \)

\[
Y(t) \approx Y(0) e^{n\alpha t}
\]

The expression for doubling time is also obtained as

\[
t_d \approx (n\alpha)^{-1} \ln 2
\]
The author has also discussed a simple model for the spread of AIDS in terms of stochastic behaviour and obtains a deterministic approximation to the stochastic process. In doing so, a number of assumptions are made. They are as follows.

i) The spread of HIV infection is within a closed male homosexual community.

ii) The population of a fixed size \( n \) at time \( t \) is divided into \( X(t) \) susceptible, \( Y(t) \) infectives.

iii) The latent period of infection is considered to be negligible.

iv) Each susceptible acquires new sexual partners at a rate \( k \).

v) The population mixes homogeneously so that at time ‘t’ the probability that such a partner is an infective is \( Y(t)/n \).

vi) \( \beta \) is the probability of getting infected from a particular infective partner.

A deterministic approximation of the stochastic process is given by differential equations

\[
\frac{dX(t)}{dt} = \left[ -\beta K X(t) Y(t) \right] / n
\]

satisfied by \( X(t) \). Assuming that at time 0, the number of infectives as small, the expression for \( Y(t) \) is
\[ Y(t) \approx Y(0)e^{-\beta t} \]

The expression for doubling time is

\[ t_d \approx (\beta k)^{-1} \ln 2. \]

Further variations of this model are also discussed. The heterogeneity of sexual activity and its influence are also considered in this paper. The incubation distribution is also considered.

Tan and Hsu (1989) have discussed a stochastic model for AIDS spread in a homosexual population. In doing so, the authors have used Kolmogorov’s forward equation. The authors also state that many biological factors such as incubation time and social factors affecting AIDS spread are subjected to considerable random variation. In developing the stochastic models, the authors assume a population consisting of four types of persons, namely S (Susceptible person), L (Latent person), I (Infective person), A (AIDS case) S(t), L(t), I(t) and A(t) are taken to be the number of persons in the four categories at time t. A set of assumptions are also given by the authors before developing the model. These assumptions include (i) homogeneous mixing of persons among the category S and I, (ii) S(t) is assumed to be large for all \( t \geq t_0 \) so that S(t) is assumed to be a deterministic function of t. Under some assumptions regarding AIDS spread (S\rightarrow L), regarding latency (L\rightarrow I), regarding incubation (I\rightarrow A), the probability generating function of L(t), I(t) and A(T) at time \( t_0 \) is given by
\[ q(t_0, t) = Q(x, y, z, t_0, t) \]

Using this p.g.f, the first order differential equation called Kolomogorov’s forward equation for the p.g.f is obtained. The expected value of \( L(t) \), \( I(t) \) and \( A(T) \) are also obtained. The effects of changing contact rate between susceptible person and infective person is also investigated in this paper.

Billard and Zhao (1991) have discussed a three stage stochastic epidemic model with applications to AIDS. In the three stage model discussed, the assumption is that at time \( t \) the population size is \( N \), of which \( X(t) \) are infected individuals, and \( Y(t) \) the total number of AIDS cases so that \( S(t) \) is the number susceptible individuals in the population is

\[ S(t) = N - X(t) - Y(t) \]

It is also assumed that AIDS epidemic is a time continuous Markov process. The infinitesimal transition probabilities in \((t, t+h)\) are

\[ \Pr\{X(t+h) = x + 1, Y(t+h) = y \mid X(t) = x, Y(t) = y\} = \lambda(x, y, t) h + o(h) \]

\[ \Pr\{X(t+h) = x - 1, Y(t+h) = (y+1) \mid X(t) = x, Y(t) = y\} = \mu(x, y, t) h + o(h) \]

\[ \Pr\{\text{two or more changes in } (t, t+h)\} = o(h) \]

where \( o(h) \) is defined by \( \lim_{h \to 0} o(h)/h = 0 \).
Let us denote

\[ P_{x,y}(t) = \Pr \{ [X(t) = x, Y(t) = y] \mid [X(0) = x_0, Y(0) = y_0] \} \]

then the forward differential difference equations can be written as

\[
\frac{d}{dt} P_{x,y}(t) = -\left\{ \lambda(x,y,t) + \mu(x,y,t) \right\} P_{x,y}(t) + \left\{ \lambda(x-1,y,t) P_{x-1,y}(t) + \mu(x+1,y-1,t) P_{x+1,y-1}(t) \right\} \text{ for } (x,y) \in B,
\]

where \( B \) is the state space of the Markov process. The solution of the set of a
differential difference equations is also obtained in this paper. The expression
for the mean and variance of AIDS cases in a closed population and also in an
open population are obtained.

Mode (1991) has discussed a stochastic model for the development of
AIDS epidemic in a heterosexual population. The author has listed out a
number of classical models which have not taken into account the concept of
bisexual population. The author has introduced a two sex model for the spread
of HIV in a heterosexual population. The progression of HIV disease in any
individual is described using six states \( e_0, e_1, e_2, e_3, e_4 \) and \( e_5 \) where \( e_0 \) is the
susceptible state, \( e_1 \) is the infected but not seropositive state, \( e_2 \) is the
seropositive but asymptomatic state, \( e_3 \) is the AIDS related complex state, \( e_4 \)
state of full blown AIDS and \( e_5 \) is the death due to AIDS. The conditional
probabilities governing the transitions among the states are assumed to be
constant over time. Couples in partnership are taken into account in a
heterosexual population; and the state of the couples is denoted j, k which means that the female in state $e_{jk}$; j, k = 0, 1,2,3,4 which represents that both the partners are alive. Based on these assumptions and also taking into account the risk of infection due to use of drug intravenously, a matrix of probabilities B is defined. In this paper using random functions, the course of the epidemic is obtained. The results are mostly by using computer simulation.

Arca et al. (1992) have authored a paper on the interaction between intravenous drug users and heterosexual population. A general mathematical model related to the transition dynamics of HIV infected is given in this paper. A compartmental structure with three compartments namely Susceptible (X), Infected (Y) and Removed (Z) and with two possible transitions from X to Y and Y to Z, is considered. The change per unit in the number X of susceptible is given by

$$\frac{dX}{dt} = -\lambda (X, Y, t) X(t)$$

Here, $\lambda (X, Y, t)$ is the force of infection at ‘t’ is given by

$$\lambda (X, Y, t) = c \frac{Y(t)}{X(t) + Y(t)} + \beta(t)$$

Where, c is the rate at which people establish contacts, $\beta$ is the probability of getting infected in a single contact. Similarly the equations for the infected and removed as

$$\frac{dY}{dt} = \lambda (X, Y, t) X(t) - \nu Y(t)$$
\[ \frac{dZ}{dt} = \nu Y(t) \]

where, \( \nu = [E(D)]^{-1} \) which gives the constant rate at which the infected developed AIDS; and D is the length of the incubation period.

Perelson et al. (1993) have discussed the dynamics of HIV infection of CD4+ T cells. A model for the interaction of HIV with CD4+ T cells is considered taking four different populations namely uninfected T cells, latently infected T cells, actively infected T cells and free of virus. Two steady states such as uninfected state with no virus present and an epidemiologically infected state in which virus infected T cells present are considered in this model. The model mainly uses the fact that the depletion of CD4+T cells as a consequence of HIV infection. A model for T cell growth in an uninfected is first considered. The T cell dynamics is described by the equation

\[ \frac{dT}{dt} = S + rT \left( 1 - \frac{T}{T_{\text{max}}} \right) - \mu_T T \]

where T is the number of CD4+ T cells as measured in the blood. S is the rate of supply of immune complement T cells from precursors in the thymus. \( \mu_T \) represents the average per capita death rate of T cells, r is the average specific T cell growth rate. A model for the influence of HIV on T cell growth is also considered. The model is given by means of a set of differential equations.
\[ \frac{dT}{dt} = S - \mu_T T + rT \left( 1 - \frac{T + T^* + T^{**}}{T_{\text{max}}} \right) - k_v VT \]

\[ \frac{dT^*}{dt} = k_v VT - \mu_T T^* - k_2 T^* \]

\[ \frac{dT^{**}}{dt} = k_2 T^* - \mu_b T^{**} \]

\[ \frac{dV}{dt} = N \mu_b T^{**} - k_v VT - \mu_v \]

where \( T^* \), \( T^{**} \) denote the concentration of latently infected and actively infected CD4+ cells. The concentration of free infectious virus particles is denoted by \( V \). The values of \( T^* \) and \( T^{**} \) are obtained. In addition to this the stability of the infected state, the dynamic of T cells depletion is also discussed.

Billard and Zhao (1994) have introduced a multistage non homogeneous Markov model for AIDS epidemic. To start with, the authors take a five stage stochastic model. The population at risk is of size \( N \) and divided into five categories susceptible, infected, seronegative, seropositive asymptomatic, pre AIDS and AIDS. The number of persons in these categories are denoted as \( X_4(t), X_3(t), X_2(t), X_1(t) \) and \( X_0(t) \) with \( \sum_{i=0}^{4} X_i(t) = N \). \( P_{x_0,x}(t_0,t) \) denotes the conditional probabilities of finding the system at state \( X \) \( (X_4, X_3, X_2, X_1, X_0) \) given that it was at the instant \( t_0 \) in the state \( X_0 \).
\[ P(X, t) = P_{x, \chi}(t_0, t) \] for \( X \in S \) and \( P(X, t) = 0 \) for \( X \notin S \).

The differential difference equations for this model are

\[
\frac{d}{dt} P(X, t) = \left\{ \begin{array}{l}
- \sum_{j=1}^{4} \lambda_j(X, t) + \lambda_1(X + e_4 - e_3, t) P(X + e_4 - e_3, t) \\
+ \lambda_2(X + e_3 - e_2, t) P(X + e_3 - e_2, t) \\
+ \lambda_3(X + e_2 - e_1, t) P(X + e_2 - e_1, t) \\
- \lambda_4(X + e_1 - e_0, t) P(X + e_1 - e_0, t) \end{array} \right\} \text{ for } X \in S
\]

where \( e_i \) is the five component vector with elements.

\[ e_0 = \delta_{i \gamma}, \quad i, \gamma = 4, 3, 2, 1, 0. \]

Thus for example \( e_4 = (1, 0, 0, 0, 0) \) and \( e_0 = (0, 0, 0, 0, 1) \).

The above equations have been solved by using a techniques developed by Severo (1969) and Billard (1973). The results of this model are extended to the case of a general \( m+1 \) multiple stage models.

Modelling the effect of treatment and behavioural change in HIV transmission dynamics has been discussed by Hernandez and Hsieh (1994). In this model the authors considered two types of infected individuals, which are infected but do not know their serological status and therapeutical treatment. The two groups of infectives differ in their incubation time, and probability of
disease transmission. There are several studies related to transmission dynamics of serological status on the sexual activity, behavioural changes and prevalence of the infection (Hethcote et al. (1991), Gail et al. (1990), Anderson et al. (1991) and Brauer et al. (1992)).

In this paper the authors considered only a male homosexual population whose infected individuals belong to either of two compartments. The first is composed of those recently infected individuals that have not changed risky sexual behaviour and the second one is treated individuals. Infected individuals enter this compartment after spending some time in the previous one. The speed at which they are recruited depends on the rate at which newly infected individuals are detected and on the effectiveness of the treatment program. In this mode of transmission the treatment program can reduce the prevalence rate and transmission probability of this compartment. The author also developed the qualitative analysis of the model at the disease-free equilibrium to determine the relative importance of changes in incubation time, probability transmission, and sexual behaviour for an HIV infectious person. The relationship between the extensiveness of the treatment program and prevention of epidemic are discussed.

As mentioned already, one of the models of transmission of HIV is from the mother to the fetus which is called the prenatal transmission. Balasubramanian and Lagakos (2001) have discussed the estimation of the timing of prenatal
transmission of HIV. Knowledge of the timing of prenatal transmission of HIV would be valuable for the determination and evaluation of preventive treatments. According to the authors the estimation of the distribution of the time of prenatal transmission is difficult. The DNA and RNA Polymerase Chain Reaction (PCR) assays and HIV infection. The timing of transmission should be estimated using diagnostic test. Non-parametric and semi-parametric inferences about the distribution of the time of prenatal HIV transmission as well as the cumulative probability of prenatal transmission are developed. The authors give a systematic derivation of the incomplete density of $\tau_1$ denoted as $g(\tau)$ where $\tau_1$ is a random variable denoting the time of transmission of HIV which can range from conception to birth. The random variable I denotes the infant’s true infection status. That is, $I = 1$ if $\tau_1 \leq 0$ and $I = 0$ if $\tau_1 = \infty$. They use the concept of likelihood contribution and derive the conditional probability mass function of $(r, I)$ where $r = (r_1, \ldots, r_n)$, $r_j$ taking the value zero or one depending upon whether the test result at time $r_j$ is negative or positive. The authors have discussed the asymptomatic properties of the ML estimators of the parameters of the cumulative distribution function for the timing of prenatal transmission.

Naresh et al. (2006) have considered a non-linear mathematical model for HIV epidemic that spreads in a variable size population through both horizontal and vertical transmission. Vertical transmission can be accomplished through transplacental transfer of disease agent. In recent years, a few studies of vertical transmission have been conducted to describe the effects of various
epidemiological and demographical factors (Agarwala (2002), Brauer (1995), Busenberg and Cooke (1993, 1998) and Li et al. (1999, 2000)). In this paper the authors devoted a model for transmission of HIV into a population of varying size with vertical transmission and other demographic and epidemiological factors. The purpose of this model is to formulate a model for AIDS epidemic that may be transmitted either directly or vertically in populations, and to study its behaviour qualitatively and numerically.

In the model, the authors assumed that the susceptible become HIV infected via sexual contacts with infectives which may also lead to the birth of infected children. It is assumed that a fraction of new born children are infected at birth; and hence are directly recruited into the infective class with a rate \((1 - \varepsilon) \theta\), and others die effectively at birth \((0 \leq \varepsilon \leq 1)\). They considered only in the vertical transmission. The interaction between susceptible and infectives is assumed to be standard mass action type. It is also assumed that some of the infectives move to join pre-AIDS class, depending on the viral counts, with rate \(\sigma \delta\) and then proceed with a rate \(\mu\) to develop full blown AIDS while others with series infection directly the AIDS class with a rate \((1 - \sigma)\delta\) where \(0 \leq \sigma \leq 1\).

In the view of the above assumptions, the spread of the disease followed by the system of equations:

\[
\frac{ds}{dt} = Q_0 - \frac{\beta_cSI}{N} - \frac{\beta_cSP}{N} - \frac{\beta_cSA}{N} - dS,
\]
\[
\frac{dI}{dt} = \frac{\beta_1 cSI}{N} + \frac{\beta_2 cSP}{N} + \frac{\beta_3 cSA}{N} - (\delta + d)I + (1 - \varepsilon)\theta(I + P + A),
\]

\[
\frac{dP}{dt} = \sigma\delta I - (\mu + d)P,
\]

\[
\frac{dA}{dt} = (1 - \sigma)\delta I + \mu P - (\alpha + d)A,
\]

\[S(0) = S_0, \ I(0) = I_0, \ P(0) = P_0 \text{ and } A(0) = A_0,\]

where \(N(t)\) – population size at time \(t\) with constant inflow of susceptible with rate \(Q_0\). \(S(t)\) – susceptible, \(I(t)\) – infectivity, \(P(t)\) – pre-AIDS patients and \(A(t)\) – AIDS patients with natural mortality rate, \(c\) – average number of sexual partners per unit time, \(\delta\) – the rate of movement from infectious class, so that \(1/\delta\) denotes the average incubation period, \(\beta_i\) \((i = 1, 2, 3)\) are the constant rates of susceptible with infectives, pre-AIDS patients and \(\alpha\) is the disease induced death rate due to AIDS.

Nirav Dalal (2006) observes in his study that many mathematical models have been developed to describe a viral dynamics of HIV-1, mostly using a system of ordinary differential equations. Verotta and Schaedele (2002) used non-linear modes to present the virus dynamics of HIV-1 which can incorporate different factors associated with resurgence. They have a non-linear model of HIV-1 dynamics, that included exposure, compliance to treatment and insurgence of resistance HIV-1 strains. They also showed the application of the
model using real AIDS clinical data involving patients treated with combination of antiretroviral drugs. Very recently Ciupe et al. (2006) discussed the dynamics of HIV-1 infection consisting of three distinct phases starting with primary infection, then latency and finally AIDS or drug therapy. In this paper, the dynamics of primary infection and the beginning of latency were modeled. They showed that with the data available, the results are highly sensitive to the chosen model.

Sani et al. (2007) have discussed the stochastic model to study the spread of HIV in a mobile heterosexual population. A new mathematical model has been developed by them for the spread of HIV that includes the factors such as mobility, heterosexual transmission and varying population size, which are essential for countries such as Indonesia, with its different regions. The model will be stochastic in nature, since it is opposed to the more common deterministic models. However, the authors showed natural stochastic approach, which can be analyzed in more detail, with respect to equilibrium behaviour. They have derived stochastic diffusion approximations, which show that the original process around the equilibrium can be approximated well by Ornstein-Hollenbeck process. Kurtz (1970) and Pollett (1990) have discussed the deterministic and diffusion approximations by using the theory of development process.
Nirav Dalal et al. (2008) analysed the stochastic model representing HIV interval virus dynamics. They showed that the model established in their paper possessed non-negative solutions as it was essential in any population dynamics model. They looked at the stability aspect of the model; and then proved that the number of infected cells and virus particles tended asymptotically to zero exponentially almost surely. Their work showed that stochastic differential equations give another option to model viral dynamics.

Samira Khalili et al. (2008) in their study, discussed the simulation results and investigated the effect of treatment initiation latency on infection dynamics. They observed that immediate initiation of treatment leads to a significant decrease of Infection Probability Percentage (IPP). They concluded that with the stochastic model IPP can be determined; and in order to minimize the IPP with minimum amount of (efficacy) of drugs, an optimization problem was formulated. The results show that scheduling optimal medication is successful in decreasing the IPP with respect to constant medication strategy. The results also show that sooner the optimal medication strategy starts; more successful it is to decrease IPP.

Yongsheng Ding et al. (2008) conclude, in their paper, that the disease transmission rate should be down under its critical value as to decrease the proportion of population infected with HIV against total population. They observe that the AIDS epidemic can die out at last, in this way. The theoretical
results for the stochastic model describing AIDS transmission are improved as far as mathematical biology is concerned. By computer simulation, the authors demonstrate the stability condition for AIDS to die out. Then, the authors conclude that AIDS will perish when they take measures to make the corresponding parameters changing according to their results. In this way, the authors offer a feasible way for AIDS prevention and control.

3.3 Models Relating to the Distribution of Random Variables (Latency period, Incubation period, Seroconversion time)

Mode et al. (1988) have discussed the study of AIDS epidemic using a stochastic model. The authors have stressed in their paper the fact that the distribution of latency period of HIV is a crucial factor in the determination of the course of an AIDS epidemic. Monte Carlo methods have been used by the authors in their study. This paper contains a parametric model for the latency period of HIV. It has also been suggested, that the Weibull and gamma distributions are suitable for the purpose of representing the variable latency periods. Assuming Weibull and gamma distributions, the expressions for survival function and conditional survival probabilities are obtained.

In addition to this, the authors have developed a parametric model of mortality due to AIDS. They have discussed the probability of getting infected by HIV in a random mixing of population of male homosexuals. The expression for the conditional probability of a sexual partner with a susceptible individual getting infected during the time interval (t-1, t] is given by
\[ P(t) = \frac{X(t-1)}{X(t-1) + Y(t-1)} \left[ 1 - G(q) \right] \]

where, \( X(t-1) \) denotes the number of infected persons at time \( (t-1) \), and \( Y(t-1) \) the susceptible individuals. \( G(q) \) is the probability that a susceptible individual escapes infection during \( (t-1, t) \). The authors have discussed the stochastic model for AIDS epidemic and used the Monte-Carlo methods to obtain the realization of the population process. Various curves are drawn to depict the number of infectives with the passage of time in months.

Taylor et al. (1990) have discussed the estimation of the distribution of time from HIV seroconversion to AIDS using multiple imputation. Multiple imputation is a model based technique for handling missing data problems. The data set used in the analysis is from the Multicentre AIDS Cohort Studies (MACS); and the study consists of 4954 homosexual or bisexual men. The authors define \( F(v|x, \theta) \) as the distribution of time to AIDS measured from the enrolment time, for the seropositive groups given covariates \( x \) and the parameter \( \theta \). Here \( \hat{\theta} \) and \( \text{cov}(\hat{\theta}) \) denote the maximum likelihood estimates and covariance obtained from the observed information matrix. Also they use the relation \( T + U + V \), where \( T \) denotes the time from HIV seroconversion to AIDS, \( U \) the follow up time, and \( V \) the residual AIDS free time. The distribution of \( T \) is estimated using the standard survival techniques. The imputation algorithm is also provided. Using the data from several cities like Baltimore, Chicaco etc., they give the distribution times as numerical illustrations.
Chevret et al. (1992) have discussed a new approach to estimating AIDS incubation time with a particular reference to men with homosexual contacts based on age distributions. Incubation period was expressed as the difference between age at time of diagnosis of AIDS and age at time of contamination. Assuming the independence between age at time of infection and incubation, the age distribution of newly diagnosed AIDS cases is the convolution product between the distributions of the age of freshly infected patients and of the incubation time can therefore be estimated from the age distribution of newly HIV infected subject and newly diagnosed AIDS cases. The p.d.f. of the age at the time of AIDS diagnosis is denoted as ‘g’. The p.d.f. of G is represented as convolution product between the p.d.f. ‘f’ of age at the time of HIV infection at the p.d.f. ‘h’ of incubation time.

Hence \( g(a) = (f * h)(a) = \int_0^a f(a-x)h(x)dx \), where ‘a’ is the age and ‘x’ is the incubation time. Therefore ‘h’ can be obtained by deconvolution between ‘g’ and ‘f’. The data set is that of 2220 AIDS cases diagnosed until 1987 reported to the Ministry of Health, France. Three different families of distributions such as lognormal, Weibull and gamma were used and the parameters are estimated by ML method. The authors have obtained two estimates of incubation period. The two were crude estimates and estimates corrected for seroconversion delay.
The characterization of HIV incubation period distribution and some comparative studies has been discussed by Tan et al. (1996). The authors indicate the importance of the incubation period distribution to estimate HIV infection and to project future HIV prevalence. The progression of HIV inside the body of the infected people is mainly concerned with HIV incubation. It has also been discussed that the stochastic model for the progression from HIV infection to the onset of AIDS ignoring the competing risks. A general stochastic model is considered with 'k' infective stages. Let $\gamma_i(s, t)$, $\beta_i(s, t)$, $\mu_i(s, t)$ denote the transition rates at time $t$ from $I_i$ to $I_{i+1}$, from $I_i$ to $I_{i-1}$ and from $I_i$ to AIDS. The reverse transition is also allowed in this model. The authors have also used a model for HIV incubation under treatment effects. It is assumed that during the $i$th round of treatment the transition rate is,

$$\gamma_i(s, t), \quad \beta_i(s, t), \quad \mu_i(s, t)$$

are given by

$$\gamma_i(s, t) = \theta_i(t) \gamma_i(i, s, t)$$

$$\beta_i(s, t) = \theta_i(t) \beta_i(i, s, t)$$

$$\mu_i(s, t) = \theta_i(t) \mu_i(i, s, t)$$

where $\theta_i(t)$ is a monotonic decreasing function of $t$. The incubation distribution in the absence of drug treatments is also depicted using Markov models and non-Markovian models. The incubation distribution for the infected people who contracted HIV at $t_0$ given by
\[ f(t) = f(t - t_0) \]

\[ = e^{r t} \exp\left[-(t - t_0)A(0)\right] \mu(0) \]

Where \( \mu(0) = A(0) I_{k+1} \) with \( I_r \) being \((r \times 1)\) column vector. The authors have obtained the survival function, and hazard function for the special cases like the general gamma distribution, etc. The incubation distribution under treatment by anti-viral drugs is also discussed. The fitting of the distribution for the data generated by Monte Carlo method is also discussed.

A stochastic model for estimating the expected time to seroconversion has been discussed by Sathiyamoorthi and Kannan (2001). They have used the shock model and cumulative damage process as discussed by Esary et al. (1973). In developing such a model they have taken the antigenic diversity as a random variable following the exponential distribution which has the so called Lack of Memory Property (LMP).

Berzuini and Allemani (2004) have analyzed the data from a seroincident cohort of 457 homosexual men who were infected with the HIV, followed within the multicentre Italian seroconversion study. These data include onset times to Acquired Immuno Deficiency Syndrome (AIDS), longitudinal measurements of CD4+ T-cell count taken on each subject during the AIDS-free period of observation and the period of administration of a Highly Active Antiretroviral Therapy (HAART), for the subset of individuals
who received it. The aim of the study is to assess the effect of HAART on the course of the disease. They analyzed the data by a Bayesian model in which the sequence of longitudinal CD4+ cell count observations and the associated time to AIDS are jointly modeled at an individual subject’s level as depending on the treatment. They discuss the inferences obtained about the efficiency of HAART, as well as modeling and computation difficulties that were encountered in the analysis. These latter motivate a model criticism stage of the analysis, in which the model specification of CD4+ cell count progression and of the effect of treatment are checked. The approach to model criticism is based on the notion of a counterfactual replicate data set $Z^c$. This is a data set with the same shape and size as the observed data, which they might have observed by rerunning the study in exactly the same conditions as the actual study if the treated patients had not been treated at all. They draw samples of $Z^c$ from a null model $M_0$, which assumes absence of treatment effect, conditioning on data collected in each subject before initiation of treatment. Model checking is performed by comparing the observed data with a set of samples of $Z^c$ drawn from $M_0$. Also the model proposed to analyze longitudinal data from 457 HIV-positive homosexual men followed within the multicentre Italian seroconversion study (Rezza et al. (1989), Italian seroconversion study team (1992), Dorrucci et al. (1999) and Pezzotti et al. (2001)). The seroconversion data of each of the subjects in the study was taken to be the middle point between the subject’s last negative HIV test result and his first
positive result. The earliest seroconversion in sample is estimated to have been in 1984. The proposed model reflects ideas from previous work on the joint modeling of longitudinal measurements and event time data (Jewell and Nielsen (1993), Singpurwala (1995), Berzuini and Larizza (1996), Cox (1999), Henderson et al. (2000), Faucett and Thomas (1996) and Taylor et al. (2000)). An additional source of complexity in this model is the representation of the effect of treatment. It would be simple and computationally straightforward to concentrate exclusively on the treatment included changes in CD4+ T-cell counts, ignoring the effect of treatment on the hazard of AIDS and the dependence of this on the CD4+ cell level. Unfortunately, this would incur bias due to informative, AIDS related, drop-out. The following notations are used to this model.

Let the symbol ‘t’ refer to the time since seroconversion with \( t = T_i \) denoting the time at which subject ‘i’ either develops AIDS or is censored, and with \( t = t_i \) denoting the time of initiation of treatment for subject i, whenever appropriate. The model in discrete time by subdividing the time axis into quite short consecutive intervals and by expressing \( t \) as a number of intervals elapsed since seroconversion. The reports of the analysis results are furnished.

Srinivasa Rao and Kakehashi (2005) have discussed the incubation-time distribution in back-calculation applied to HIV/AIDS data in India. Information on accurate population sizes of HIV infected persons and AIDS
cases and the trend of these figures are requisite to the planning of preventive policies and public-health management. Sophisticated statistical models have been developed to facilitate provision of the information. Among the models, a simple extrapolation method is easy to apply and useful for summarizing the trend of the spread of infection, but it is difficult to clarify how long the obtained trend stays unchanged. By comparison, mathematical models of the spread of sexually transmitted diseases use information on sexual behaviour in the population to investigate the effect of behavioural change caused by a preventive program. But mathematical models usually require detailed information on sexual behaviour in the population, which is not always available. In contrast, the back-calculation method connects infection with HIV and the development of AIDS to incubation-time. Because of the long incubation period, this method can provide a very reliable prediction of future AIDS development from present HIV data.

Traditionally, back-calculation method is applied to estimate past HIV trends and to predict future AIDS cases by using reported AIDS cases and the assumed incubation time distribution. Information on HIV incidence is not directly used in the attempt. This is to be expected where detailed information on AIDS cases and incubation time may be more easily obtained about than the figures about HIV incidence. But there is another situation in which information on HIV incidence is more available than information about incubation time. This is likely when HIV surveillance is started but medical
treatment is not generally available or is inadequate. In any case it is useful and helpful to use all these data to obtain more reliable outcomes especially when the quality of each kind of data is insufficient. A recent attempt and effort to take advantage of the information on HIV in back-calculation was made by Cui and Becker (2000), but it required more detailed information on reported HIV.

The mathematical study of a staged-progression HIV model with imperfect vaccine is by Gumel et al. (2006). A staged-progression HIV model is formulated and used to investigate the potential impact of an imperfect vaccine. The vaccine is assumed to have several desirable characteristics such as protecting against infection, causing bypass of the primary infection stage, and offering a disease alerting therapeutic effect (so that the vaccine induces reversal from the full blown AIDS stage to the asymptomatic stage). The model, which incorporates HIV transmission by individuals in the AIDS stage, is rigorously analyzed to gain insight into its quantitative features. Using a comparison theorem, the model with mass action incidence is shown to have a globally-asymptomatically stable disease-free equilibrium whenever a certain threshold, known as the vaccination reproduction number, is less than unity. Furthermore, the model with mass action incidence has a unique endemic equilibrium whenever this threshold exceeds unity. Using the Li-Muldowney techniques for a reduced version of the mass action model, this endemic equilibrium is shown to be globally-asymptomatically stable, under certain
parameter restrictions. The epidemiological implications of these results are that an imperfect vaccine can eliminate HIV in a given community, if it can reduce the reproduction number to a value less than unity, but the disease will persist otherwise. Furthermore, a future HIV vaccine that induces the bypass of primary infection amongst vaccinated individuals (who become infected) would decrease HIV prevalence, whereas a vaccine with therapeutic effect could have a positive or negative effect at the community level.

Krisha Ray et al. (2006) in their study on CD4 / CD8 lymphocyte counts in healthy, HIV-positive individuals and AIDS patients, observed CD4 cell count as a useful predictor of AIDS in Indian conditions and confirmed that a significant percent of AIDS patients had CD4 count below 200/ml.

For the purpose of the study, blood samples collected from 125 HIV exonerative healthy volunteers comprising group I and 425 HIV positive patients divided into group II as asymptomatic and group III of AIDS patients in the age group of 17-60 years were analysed for enumeration of CD4+, CD8 cell/ml by flow cytometry. Differences between means of two groups were compared by Student’s unpaired t-test. The percentages were compared by chi-square test. Data of the earlier studies, wherein sample size, mean and standard deviation are available, were used for comparison.

In the estimation of expected time to seroconversion, there is an important role for the interarrival times between the successive contacts; and it
has a significant influence. Sathiyamoorthi and Kannan (2006) have obtained
the expected time to seroconversion under the assumption that the interarrival
times between the contacts are not independent but constantly correlated.

A stochastic model for HIV transmission under alertness has discussed
by Kannan et al. (2011, 2012). In their paper, the authors have developed a
stochastic model for the estimation of expected time to seroconversion; and its
variance are derived under the assumption that the threshold level of antigenic
diversity is a random variable which follows Mixed Exponential and
Exponentiated Exponential distribution and interarrival times between contacts
follows Exponential distribution. In doing so, the authors have assumed the
following:

i. Sexual contact is the only source of HIV transmission.

ii. During any contact in which a person is in-alert the transmission
    of HIV is a sure event.

iii. An individual is exposed to a damage process acting on the
    immune system and damage is assumed to be linear and
    cumulative.

iv. Damage occurs if a person is in-alert.

v. If the total damage caused when exceeds the threshold level,
    which itself is a random variable, the seroconversion occurs and a
    person is recognized as seropositive.
vi. In any single contact, a person is alert with probability p and in-alert q, so that p + q =1.

Using the above assumptions, the expected time to seroconversion and its variance are discussed. In this study the numerical illustrations are provided.

Claggett et al. (2012) have discussed cross-sectional HIV incidence estimation based on a sensitive and less-sensitive test which offers great advantages over the traditional cohort study for incidence estimation (Brookmeyer and Quinn (1995), Janssen et al. (1998)). Based on the features of the evolving HIV-1 antibody response during the months after primary HIV-1 infection, several Enzyme Linked Immuno Sorbent Assays (ELISAs) have been developed to differentiate early and long-standing HIV infections. Balasubramanian and Lagakos (2010) postulate a longitudinal history statistical model of HIV seroconversion and subsequent reactivity to a less-sensitive assay, and show that the cross-sectional estimator arises as a maximum likelihood estimator under this model. Wang and Lagakos (2010) propose an augmented cross-sectional study design that adds a longitudinal component to the traditional cross-sectional study by following forward those subjects who test positive on the sensitive assay and test negative on the less-sensitive assay until either they become reactive to the less-sensitive assay or for a predetermined length of follow-up time T, whichever occurs first. The impact of varying follow-up time on estimating HIV incidence within the context of an augmented cross-sectional design, evaluates the robustness of incidence
estimators to the assumptions about the existence and size of the subpopulation where infected subjects will remain negative permanently on the less-sensitive test, and propose a new estimator based on abbreviated follow-up time.

3.4 Models Relating to Infectivity, Projection and Other Aspects

Many papers on the projection of AIDS epidemic figures have come up, and one such a paper is by Healy and Tillet (1988). This paper uses log linear model and exponential weighting in forecasting the AIDS epidemic. The authors state that in the study of AIDS epidemic there are six key dates (i) date of infection, (ii) date of seroconversion, (iii) date of onset of AIDS symptoms, (iv) date of diagnosis, (v) date of report and (vi) date of death. It has also been observed that the date of infection and date of seroconversion are very rarely known. The date of onset is a vague indication; date of diagnosis is not always recorded. However the authors have used curve fitting with two types of curves. (a) unweighted fit, and (b) a log linear fit. Using the curves the doubling time is found out.

The statistical analysis of HIV infectivity based on partner studies has been discussed by Jewell and Shiboski (1990). In their paper the authors define the per contact transmission probability as infectivity. The estimation of infectivity is very useful to project the spread of the virus and in evaluating the various intervention strategies. For this purpose, the authors have defined two random variables \((Y_t, K_t)\), based on each partnership.
Where,

\[ Y_i = \begin{cases} 
1 & \text{if partner is infected by time } T_i \\
0 & \text{if partner remains seronegative at time } T_i 
\end{cases} \]

\[ K_i = \text{Number of contacts with index case during period } (0, T_i) \]

Here, the index case refers to the partner who is already known to be infected. Assuming that the infection of the partner occurs only through contacts with the index case and not through any other source of infection, the authors have obtained a method of estimating the infectivity taking \( k \) contacts. If each contact carries a fixed risk of transmission denoted as \( \lambda \), they have used a basic model for calculation of infectivity defined as

\[
\Pr(Y_i = 1/K_i = k) = 1 - (1 - \lambda)^k.
\]

Using the concepts of survival analysis, the authors have defined the hazard function as

\[
\lambda(k) = \Pr \left\{ \text{infection occurs after } k \text{ contacts given that infection has not occurred after } k - 1 \text{ contacts} \right\}
\]

and the associated cumulative infection functions.

\[ S(k) = \Pr \{ \text{infection has not occurred after } k \text{ contacts} \} \]

It will be helpful to refer to \( P(k) = 1 - S(k) \) as the prevalence function. This, of course, refers to the proportion of individuals who are infected after \( k \) contacts and is the distribution function of the random variable that counts
the number of contacts at the occurrence of infection of the partner. If follows that,

\[ 1 - \lambda(k) = \frac{1 - P(k)}{1 - P(k-1)} \]

The authors have discussed the estimation of \( \lambda(.) \) using the conditional likelihood based on \( n \) observations. The non-parametric estimation of \( \lambda(.) \) is also discussed. The authors have indicated some additional techniques for examining the fitness of the basic model to the observed data. In doing so, the complementary log-log link to the basic model is discussed. The results have been substantiated using partnership data collected from partnership clubs.

The antigenic diversity is a consequence of HIV transmission; and this kind of antigenic diversity is observed in the reproductive process of the HIV which infects the CD4 cells. Antigenic variation seems to be a major property of many parasitic infections. The parasites evolve the capacity to escape the immunological surveillance by mutating their immune dominant epitaphs continuously during the time of infection. No sooner has the immune system generated cellular or humoral attack against these targets than has the parasite escaped with mutated antigens. This precisely is called the antigenic diversity. HIV appears to disrupt and confuse the immune regulatory network. Nowak and May (1991) have discussed a mathematical model that explores the quantitative consequences of the antigenic drift of HIV.
The basic sets of equations are

\begin{align*}
v_i &= v_i (r - sz - px_i), \quad i = 1, 2, 3, \ldots n \\
x_i &= kv_i - uvx_i, \quad i = 1, 2, 3, \ldots n \\
\beta &= k^1v - uvz, \\
\end{align*}

where

\begin{align*}
v_i &= \text{population size of virus strain } i \\
r &= \text{replication rate} \\
r &= bQ - d \\
bQ &= \text{birth rate} \\
Q &= \text{probability that the replication is done without error} \\
d &= \text{natural death rate of the virus} \\
\nu &= \sum v_i = \text{total virus population density} \\
\end{align*}

The potential existence of a viral diversity threshold has been established. The immune system can control strain \( i \), if \( v_i < 0 \), which can be written as

\[ r - sz - px_i < 0 \]

The immune system can thus control each individual strain only if this holds for all \( i = 1, 2, \ldots, n \), which implies the restriction that

\[ n < n_c (x, z) = p\tilde{x}/(r - sz) \]
Hence there exists an upper limit, \( n_e \) of different strains that can be suppressed simultaneously by the immune system given by

\[
n_e = \frac{pk}{(ru - sk^1)}
\]

It may be observed that the antigenic diversity is responsible for the depletion of the immune level or immune capacity of an individual progressively. So, mathematical models have been developed under the assumption that the magnitude of damage to the immune system is directly proportional to the antigenic diversity; and using this as the basis, the shock models have been introduced in this model.

Stilianakis et al. (1994) have discussed on the antigenic diversity threshold model for AIDS. The model suggests an antigenic diversity threshold above which the immune system is unable to suppress the virus population. It may be observed at this juncture that this paper has served as the basis for using the threshold models which are discussed in this thesis. The authors of the above cited paper used a model under the assumption that HIV induces two kinds of responses (i) responses against the specific viral strains and (ii) non specific responses against all viral strains. A system of many countable differential equations is used to describe the model for \( i = 1, 2, \ldots, N(t) \).

\[
\frac{dv_i}{dt} = v_i (\alpha - \beta z - \gamma x_i)
\]

\[
\frac{dx_i}{dt} = \lambda v_i - \xi v x_i
\]
\[
\frac{dz}{dt} = v(e - z)
\]

where

\[
v = \sum_{i=1}^{N(t)} v_i
\]

and

\[
v_i(t) = \begin{cases} 
0 & \text{for } \{t; N(t) < i\} \\
 v^* & \text{for } \inf \{t; N(t) = i\}
\end{cases}
\]

The model parameters are

\(\alpha\) - replication rate of HIV of strain i

\(\beta\) - HIV elimination rate per nonspecific immune cell

\(\gamma\) - HIV elimination rate per specific immune cell

\(\lambda\) - number of new specific immune cells per unit of time per virus

\(\varepsilon\) - number of new nonspecific immune cells per unit of time per virus

\(\xi\) - elimination rate per HIV organism of specific and nonspecific immune cells

The variations of mutation rate, the effect of the variation in the size of the initial virus population, and the effects of the initial viral diversity are all discussed in this paper. In addition to this, the authors have discussed the variations of the antigenic diversity threshold also.
Although the application of sophisticated methods to HIV/AIDS data was delayed, recently it has begun. The projection of AIDS (Basu et al. (1998)) is useful, if all India level transmission probabilities are available. However, in a population where the dependable data as mentioned above are not accessible, researchers could assume a reasonable set of scenarios for the behavioral and epidemiological parameters, so that the scope of the epidemic could be determined. Modeling of this kind may not be explicit, but it is important to note that such a model guides one to predict the scope of the epidemic in future, until dependable data become available. The popular back-calculation approach (Brookmeyer and Gail (1986), Brookmeyer and Gail (1988)), that assumes the distribution of incubation time of AIDS to be known and then through convolution project the AIDS cases, is extensively used by researchers. Longitudinal studies on HIV infected individuals with reliable infection dates are necessary to ascertain the incubation period of AIDS. Unavailability of infection dates in India causes problems of left censoring methods to deal with such situations which were developed and discussed by Rao et al. (2005).

In such a situation, application of mathematical models is difficult to carry out. Dynamic transmission models suggested so far have focused on the rate of new infection in a population (Anderson and May (1991)), and these also emphasized mixing patterns of the uninfected and infected individuals in a population (Anderson (1988)); for review see (Isham (1988)). Masayuki Kakehashi used a novel application of such models by incorporating realistic
epidemiological parameters (Kakehashi (1998) and Kakehashi (1999)) incorporating the features of epidemiological models in a basic and lucid way (Hethcote (1989)). Network models studied the spread of sexually transmitted infections and the methods were developed to estimate basic reproductive rates (Bauch and Rand (2000), Bauch (2002)). A recently proposed deterministic model (Kakehashi (2000)) could be a dynamic way of estimating growth of disease in specific situations where information on sexual behavior is available. Moreover, the Indian population is relatively free of AIDS therapy, which reduces the complexity of the analysis.

Adriana Weinberg et al. (2009) discussed and concluded that resistance to antiretroviral was common in antiretroviral-experienced pregnant women, but not in naive women. The 14% prevalence of resistance to zidovudine and lamivudine in antiretroviral experienced women suggests that alternative Nucleoside analogue Reverse Transcriptase Inhibitors (NRTI) are desirable for this group of patients. Shan mei et al. (2010) in their paper introduced the concept of a complex agent network to model the HIV epidemic by combining multi-agent systems, in which the agents represent individuals who could have sexual interactions. They report that the model can be adopted to predict the future trend of HIV prevalence and incidence among (Men who have Sex with Men) MSM in Amsterdam (Netherlands).

Ron Brookmeyer (2010) discusses the current approaches and methods for measuring the HIV/AIDS epidemic and their strength and weakness. The
author summarizes the main sources of errors and problems with these and other approaches and discusses opportunities for improving their reliability. Changing methods and data sources present new challenges, because the incidence and prevalence estimates produced at different points in time are not directly comparable with each other, which complicates assessment of time trends. The methodological changes help explain the changes in global statistics. As methods and data source continue to improve, the development of statistical tools for better assessing the extent to which changes in HIV/AIDS statistics can be attributed to changes in methodology versus real changes in the underlying epidemic which is an important challenge.

Beena Thomas et al. (2011) have discussed the unique socio-cultural issues of Men who have Sex with Men (MSM) in India, and how they relate to HIV risk that could maximize the utility of future prevention effort. This article seeks to elucidate the specific challenges of providing effective HIV prevention program for this diverse and socially marginalize risk group. It highlights the gaps in current HIV prevention efforts by providing insight into patterns of Indian MSM behaviour and sexual partnerships, and the specific cultural and psychological content in which HIV risk is occurring. Understanding the distinct social forces that shape the HIV risk environment could maximize the effectiveness of prevention interventions and heighten the acceptability of these programmers by MSM.
Julia Braun et al. (2012) have discussed the predictive cross-validation for the choice of linear mixed-effects models with application to data from the Swiss HIV cohort study. A predictive cross-validation approach has the advantage that the model has to be fitted just once and not in every cross-validation step. This saves a considerable amount of time, especially in more complicated linear mixed-effects models with random effects and serial correlation. Empirical evidence suggests that this cross-validation approach is good approximation of a true leave-one-out approach, but this of course depends on the nature of the concrete problem and data set. It is a well-known fact that predictions in the context of HIV are particularly challenging, because the viral load and its effect on the immune system can change so fast. Furthermore, one may wish to choose the best model, and for this task, our cross-validation procedure has shown to be useful.