CHAPTER I

INTRODUCTION

1.1 INTRODUCTION

Acquired Immune Deficiency Syndrome (AIDS) is a devastating disease caused by Human Immune deficiency Virus (HIV). The disease is transmitted through sexual contact, any form of contact where exchange of bodily fluids has occurred or using infected blood products. Following a few recognized cases among homosexual men in the United States in the year 1981, new cases of AIDS were subsequently reported in a many countries throughout the world. This epidemic ranks as one of the most destructive scourges in human history and has posed a formidable challenge to the biomedical research and public health communities of the world. It has now reached a pandemic proportion, as no country in the world is now free from HIV/AIDS.

The global pandemic of HIV infection comprises many different epidemics, each with its own dynamics e.g., time of introduction, population density, cultural and social issues. The spread of the epidemic has varied considerably between developed and developing countries, depending on the culture as well as other social and behavioral patterns. Incidence rates have been the highest in developing countries where heterosexual transmission is most common.

The first AIDS case in India was reported in the year 1986. Now India's entry
into the third phase of the HIV epidemic, as envisaged from the increasing number of HIV infections detected even among housewives and children, signals a major AIDS crisis in the offing. The first phase of HIV was recognized when rising trend of HIV prevalence was established among the Commercial Sex Workers (CSW) in 1988 and the professional blood donors in 1989. The second phase started after 1989 when several of the clients of CSW and blood recipients were found to be infected.

1.1.1 Biological Aspects of HIV/AIDS and Its Transmission

AIDS is a condition in which the inbuilt immune mechanism of the human body breaks down completely. The process is gradual but ultimately suppresses the immune ability of the individuals. It is a medically accepted fact that the HIV is the causal agent of AIDS. Those who are affected by AIDS are susceptible to other opportunistic infections like Candida, Mucormycosis and Aspergillus etc.

Within a few weeks of its entry into the blood stream of an individual, the HIV sets an insidious and progressive attack on the immune system of the individual by binding itself to the T4 helper cells. The viral attachment to the cell is initiated by means of antibodies to the virus. On infecting a T4 cell, the viral RNA is injected into the target cell along with viral transcriptase and viral integrase. The viral reverse transcriptase transcribes the viral RNA into viral DNA and the viral integrase bind the viral DNA with that of the host. The integrated viral DNA may remain latent or in an activated form. In the activated form, genomic RNA and messenger RNA are transcribed from the integrated viral DNA. The regulatory proteins such as tat and rev are translated from messenger
RNA. These proteins together with the genomic RNA cause the production of new HIV viruses, which bud on the cell wall. These buds accumulate on the cell wall and after a random length of time from the time of infection, the infected cell disintegrates (i.e. the cell undergoes a lyses) releasing a random number of HIV viruses and this process continues indefinitely. This mechanism by which the killing of the T4 cells takes place, the number of free virus in the blood destroys progressively the immune competence of the host. The viral load (the amount of free HIV in the blood) increases following infection and peaks at the time of sero-conversion (i.e. the time at which detectable levels of antibodies develop in the blood) and falls to a level (called set-point) by about 2 years thereafter and then remains at that level throughout the asymptomatic period. A low level of set point corresponds to a low risk of AIDS and a high level of set point corresponds to a high risk of AIDS. The viral load measured at arbitrary base line is a good prognostic marker at all stages of the development of AIDS. On the other hand, the T4 cell counts in the blood decrease rapidly in the initial months following infection and thereafter at a slower rate. After a period of several years, when the level of T4 cells declines to roughly 400 cells/mm$^3$, the person exhibits symptoms and signs associated with HIV disease. These symptoms and signs do not meet the surveillance definition of AIDS and we say that the person is suffering from AIDS related complex (ARC). The level of T4 cells further continues to decrease to roughly below 200 cells/mm$^3$ and AIDS defining diseases such as *Kaposi's sarcoma and Pneumocystis carinii pneumonia* occur. At this stage, the patient is said to have developed AIDS.
The current situation of HIV in India is different from when the first HIV/AIDS case was identified in Chennai, the capital of Tamil Nadu state, in 1986. Adult HIV prevalence in India declined from 0.41% in 2000 to 0.31% in 2009. The 2008-09 India HIV estimates developed by NACO with support from National Institute of Medical Sciences, National Institute of Health and Family Welfare, UNAIDS and WHO utilised improved methodology and updated epidemiological data from the latest rounds of HIV Sentinel Surveillance and other information on High Risk Groups for more accurate understanding of the Indian epidemic. It is estimated that India had approximately 1.2 lakh new HIV infections in 2009, as against 2.7 lakh in 2000 (NACO 2011). According to NACO, prevalence was 0.41 percent through 2000, 0.36 in 2006 and 0.31% in 2009. All the states with high prevalence rates show a clear declining trend in adult HIV prevalence. However, Chandigarh, Orissa, Kerala, Jharkhand, Uttarakhand, Jammu & Kashmir, Arunachal Pradesh and Meghalaya with low prevalence rates show rising trends in adult HIV prevalence in the last four years (NACO 2011).

There are gaps and challenges to be addressed in the fight against HIV and AIDS. HIV/AIDS surveillance methods evolve over time, so data from the same source may not be directly comparable year to year. The type of data available and the lag-time in availability may pose challenges to assessing recent impact. Epidemiological measures of HIV/AIDS are numerous and each has important and distinct definitions. Much of the data are estimates only. This is true globally and in all countries, even the United States, due to the lag-time between HIV infection and the development of AIDS, the fact that many do not know their status, stigma
which leads to under reporting, and surveillance systems that may not be complete.

Attention must be given to ranges given around any estimate, as well as any notes that may accompany data, since these may provide important information that can help in the interpretation. Rates and percents, not just numbers, are important.

In 2001, the UNAIDS/WHO introduced a new version of its epidemiological model for all developing countries. The model known as Estimation and Projection Package (EPP) makes use of the surveillance data from each HIV sentinel site in estimating HIV/AIDS prevalence rates. The Spectrum Program also utilizes the information on the birth rate, death rate and the output of EPP for the country in calculating the estimated number of people living with HIV, number of new infections, number of AIDS cases, number of AIDS deaths, number of orphans, etc. The extent to which this data source is representative of the entire adult population of India is doubtful. This is because not all pregnant women have attended antenatal clinics and not all women of adult age were pregnant at the time of the surveys. Also the selection procedure of the survey sites systematically excludes private clinics where many births occur. Therefore, national estimates based on these surveys rely on a fraction of women who attended the selected antenatal clinics. Also the Spectrum Program of UNAIDS and WHO makes use of vital rates which were obtained from a poor vital registration system.

1.1.2 The Role and the Importance of Mathematical Models for AIDS

Mathematical models are useful in a number of ways. Because of the complexity and the seriousness of the disease, mathematical models of the HIV epidemic and
HIV pathogenesis are especially important for a number of reasons. Mathematical models of the HIV epidemic and HIV pathogenesis can be used to provide in depth understanding of some basic features and principles of the HIV epidemic and HIV pathogenesis; it will help reveal consequences of important parameters of the HIV epidemic and HIV pathogenesis (Anderson 1988; Anderson & May 1992b; Hyman & Stanley 1989; Perelson et al. 1998; Wein et al. 1998). One may use mathematical models of the HIV epidemic to assess impacts of many risk factors and to screen for important risk variables for purposes of prevention and control of AIDS (Anderson et al. 1989; Hethcote et al. 1991a; Hethcote et al. 1991b; Tan & Hsu 1989). Mathematical models of the HIV epidemic can also be used to evaluate and compare different strategies of prevention and control of AIDS (Tan & Xiang 1999). One may use mathematical models of the HIV epidemic to project future AIDS cases and future HIV prevalence (Aalen et al. 1997; Brookmeyer & Damiano 1989; Day et al. 1989; De Angelis et al. 2002; Heisterkamp et al. 1992). A summary of the usefulness of AIDS mathematical models have been given by Anderson (Anderson 1988), which we quote:

"Mathematical models of infection and disease serve as illuminating caricatures, as foundations to build on, as analytic tools for the estimation of epidemiological parameters, and as guides to the information needed for improving epidemiological understanding and in planning programmes of control."
Hethcote and Van Ark (Hethcote & Ark 1992) have listed ten important reasons why mathematical models are needed. All these suggest the importance of mathematical models for studying and for controlling the HIV epidemic.

1.1.3. Different Modeling Approaches for Modeling the AIDS Epidemic in HIV-infected Individuals

To develop mathematical models for AIDS and cancer as well as for other systems, there are basically four different types of modeling approaches: The deterministic models, the stochastic models, the statistical models and the state space models (Kalman filter models). For the AIDS epidemic, many deterministic models have been developed and used by applied mathematicians (Anderson 1988; Anderson & May 1992b; Hyman & Stanley 1989; Perelson et al. 1998; Wein et al. 1998; Anderson et al. 1989; Hethcote et al. 1991a; Hethcote et al. 1991b; Hethcote & Ark 1992; Ahlgren et al. 1990; Anderson & May 1992a; Castillo-Chávez 1989; Jager & Ruitenberg 1987; Kaplan & Brandveau 1994; Wilkie 1989).

Since the AIDS epidemic is basically a stochastic process as many risk variables are subjected to stochastic variations, some stochastic models have been developed by some statisticians and probabilists (Tan & Hsu 1989; Tan & Xiang 1999; Blanchard et al. 1989; Billard & Zhao 1991; Balakrishnan 2009; Longini et al. 1989; Longini et al. 1991; Isham & Medley 1996; Mode 1991; Mode et al. 1988; Mode et al. 1989; Statten & Longini 1994; Tan 1991; Tan 1993; Tan & Byers Jr. 1993; Tan & Tang 1993). On the other hand, based on AIDS epidemiological data and survey data, many statistical models have been
developed by statisticians (Brookmeyer & Gail 1994; Jewell et al. 1992; Isham 1989; Longini et al. 1992). To combine advantages from both stochastic models and statistical models, in 1995 Wu and Tan (Hulin & Wai-Yuan 2000) have introduced the state space model into AIDS research. Since then many papers have been published to develop state space models for the HIV epidemic and the HIV pathogenesis (Hulin & Wai-Yuan 2000; Cazelles & Chau 1997; Tan & Xiang 1998; Tan & Xiang 1997; Tan & Xiang 1999).

1.1.3.1. Deterministic Models of the HIV Epidemic

Deterministic models assume that all response variables such as the numbers of susceptible people, infected people 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 or integral equations. These equations are derived by taking into account the biological, the epidemiological as well as the clinical aspects of the HIV epidemics. By analyzing these equations, one may then study the behaviour and progression of the HIV epidemic as time progresses.

1.1.3.2. Stochastic Models of the HIV Epidemic

Stochastic models assume that the response variables are a family of random variables indexed by time so that the HIV epidemic is basically a stochastic process. Since nature is basically stochastic and many variables are subject to stochastic variations, stochastic models are more realistic than deterministic
models. However, the mathematics in stochastic models is usually more complicated and difficult than those involved in deterministic models. Under some very special conditions, results of the deterministic models may provide a close approximation to the results of the mean numbers of the stochastic models. In these cases, the deterministic models may be considered as the models dealing with the mean numbers of the responses. In these cases, one may then consider the deterministic models as special cases of stochastic models. However, it has been shown by Mode et al. (Mode et al. 1988; Mode et al. 1989), Isham (Isham 1991), and Tan and his associates (Tan & Xiang 1999; Tan & Xiang 1998; Tan et al. 1995; Tan & Wu 1998) that for the HIV epidemic and HIV pathogenesis, very often there are significant differences between results of deterministic models and the mean numbers of the corresponding stochastic model. Hence, for most of the cases in AIDS epidemic and HIV pathogenesis, the results of the deterministic model would provide poor approximation to the results of the mean numbers of the corresponding stochastic model.

1.1.3.3. Statistical Models of the HIV Epidemic

Statistical models are models which are constructed from observed data from the system and hence are data dependent. To illustrate, let \( \hat{Y} \) be a vector of observed data set on the response variables. Then, the statistic model expresses \( \hat{Y} \) usually as a sum of the true response vector \( \eta(X,\theta) \) and a vector of measurement error which is \( e \) the vector of random disturbances associated with measuring the responses.

\[
\hat{Y} = \eta(X,\theta) + e
\]
In the above equation, $\eta(X, \theta)$ is usually a vector of deterministic functions whereas, $\epsilon$ a vector of random variables.

The elements of $\eta(X, \theta)$ are usually functions of some risk variables $X$ and some unknown parameters $\theta$. If the function form of the functions in $\eta(X, \theta)$ are known, the above model has been referred to in the statistical literature as linear or nonlinear regression. If the function form of the functions in $\eta(X, \theta)$ are unknown, then the above model has been referred to in the statistical literature as non-parametric regression.

Comparing with the deterministic and stochastic models, the statistical model makes full use of data but very often ignores the mechanism of the system as well as other prior information about the system. On the other hand, the deterministic model and stochastic model take into account the mechanism and other information of the system but usually do not make full use of available data.

1.1.3.4. The State Space Model (Kalman Filter Model)

The state space model (Kalman filter model) consists of two sub-models: The stochastic system model which is the stochastic model of the system and the observation model which is a statistical model based on available observed data from the system. The state space model takes into account the basic mechanisms of the system and the random variation of the system through its stochastic system model and incorporates all these into the observed data from the system. Furthermore, the state space model validates and upgrades the stochastic model through its observation model and the observed data of the system. It follows that
the state space model adds one more dimension to the stochastic model and to the statistical model by combining both of these models into one model. It combines information and advantages from both the stochastic model and the statistical model.

The state space model was originally proposed by Kalman in the 60's for engineering control and communication. Since then it has been successfully used in satellite research and military missile research. It has also been used by economists in econometric research (Harvey 1990) and by mathematician and statisticians in time series research (Aoki 1990) for solving many difficult problems which appear to be extremely difficult from other approaches. In 1995, the state space model was first proposed by Wu and Tan in AIDS research (Hulin & Wai-Yuan 2000). Since then it has been used by Cazelles and Chau (Cazelles & Chau 1997) and by Tan and his associates for modelling AIDS epidemic (Tan & Xiang 1998; Tan & Xiang 1997); it has also been used by Tan and his associates for studying the HIV pathogenesis in HIV-infected individuals (Tan & Xiang 1999; Tan & Xiang 1998).

1.1.3.5 Curve fitting Method

A recent advance in this field is the addition of curve fitting models. Sathian B and Sreedharan J used this method for forecasting several infectious and non infectious diseases(Sathian & Sreedharan 2012; Sathian et al. 2010a; Sathian et al. 2010b; Sathian et al. 2011a; Sathian 2011b; Sathian et al. 2011c). It gives more accurate estimates compared to the other models. Sathian and Sreedharan studied the applicability of the curve fitting method in the projection of reported
HIV cases in Nepal as an alternative to the back calculation method. It showed greater accuracy with the comparison of 2010 actual reported cases and forecasted cases (99% accuracy) (Sathian & Sreedharan 2012; Sathian 2011b; Sathian et al. 2011a).

1.2 RELEVANCE OF CURRENT STUDY

HIV/AIDS is one of the most leading challenges for world public health. It was first identified in United States nearly two decades ago. The majority of HIV and AIDS cases appear in sub Saharan Africa. Statistical methods have expanded spatially in recent years to address large scale worldwide health issues. These methods have a prominent role in the study of the HIV/AIDS epidemic.

It is well known that HIV/AIDS is a fatal epidemic and a successful cure has not been discovered as yet. Hence active precautionary measures have to be taken with thorough care against the spread of the epidemic. In order to determine the current levels and trends in this epidemic, access to the relevant and comprehensive research data is pertinent. Gathering precise information on incidence and prevalence is difficult due to stigma of the disease.

The use of statistical modelling approaches make a valuable contribution for developing better understanding of the levels and trends in the HIV epidemic and the limited information based on the estimates.

The aim of any modelling is to extract as much as information possible from available data. Therefore these statistical methods will be the most reliable source to detect the prevalence and incidence of HIV/AIDS.
Until the last decade conventional research through statistical methods were adopted to understand the trend and prevalence of HIV/AIDS in almost all countries. The recent studies in India showed that the resulting incubation time distribution for back calculation of HIV cases is not significantly different from the weibull distribution.

A detailed study on the application of weibull distribution and other important distributions applicable in this field will be examined in detail and the most convenient distribution will be identified.

1.3 OBJECTIVE OF THE STUDY

The proposed study is planned to find suitable statistical distribution for data regarding HIV/AIDS.

Specific

(1) To demonstrate the application of weibull distribution in the field of HIV incubation time data.

(2) To find suitable distribution for the incubation time data of HIV/AIDS.

(3) To evaluate the characterization properties, asymmetry, skewness and kurtosis of the suitable identified distribution.

(4) To predict the number of cumulative number of AIDS cases with Back Calculation Model using incubation time distribution

(5) To predict the number of HIV cases using curve fitting Method