CHAPTER ONE

INTRODUCTION

1.1 OVERVIEW

Acquired Immunodeficiency Syndrome (AIDS) and its etiologic agent Human Immunodeficiency Virus (HIV) are two viral infections that have been most studied world wide since their first discovery in 1981 and 1983 respectively with 60 million people infected, 20 million deaths from AIDS and 14,000 daily new infections of which 95% is in developing world countries (WHO HIV statistics 2007). Statistics showed that around 33.2 million people were living with HIV at the end of 2007 of which 2.5 million were children, and a greater proportion of the population infected coming from Africa and Asia continents. Compared to approximately 47,000 cases of AIDS in the United States with 58% of the patients already dead since the first cases were reported in summer 1981 until 1 December, 1987 (Barker et al. 1998, Fauci 1988, Feinberg 1996). This shows that HIV appears to be progressive and irreversible with a high mortality rate that may approach 100 percent over several years if not put in check.

Presently only two types of HIV are known to infect humans, namely the HIV-1 and HIV-2. These two viruses evolved independently and may have crossed the monkey-human species barrier at several independent occasions (Groot, 2006). HIV-1 is believed to have originated from wild chimpanzees (Pan troglodytes) virus (SIVcpz) of the southern Cameroon of West Africa, while HIV-2 originated from sooty mangabey monkey (cercocebus atys) virus (SIV variant) of Guinea-Bissau, Gabon and Cameroon of Africa (Groot, 2006, HIV Wikipedia 2008).

Studies have reported cases of infection with HIV-2 such as the Portuguese man infected in Guinea-Bissau who had a clinical latency duration of 19 years (Ancelle et al. 1987), a Portuguese woman infected through blood transfusion and she had a clinical latency of 27

years (Mota-Miranda et al., 1995) and a Japanese man (first report of HIV-2 infected Japanese individual) with clinical latency of 35 years (Utsumi et al. 2007). One can effectively induce that HIV-2 has a longer clinical latency period than that of HIV-1. HIV-2 also has a lower transmission rate (this was pointed out by Utsumi et al. (2007), when they noted that the man had sexual acts with his 77 years old wife and both she and their 37 years old son were HIV negative). Also HIV-2 has less immune activation (see Levy 2009). Hence these have posed as very good advantages for the research into vaccine and cure of HIV-2 unlike its counterpart which has a shorter clinical latency period (10 years on average) and it is the more common viral infection in the world (http://en.wikipedia.org/wiki/HIV). Thus concentration has been placed on the HIV-1 strains among human and not the HIV-2 strains in this thesis solely because HIV-1 is more virulent, easily transmitted, has a lesser clinical latency duration (of say 10 years) and according to WHO reports, it is the cause of the majority of HIV infections globally. Unlike HIV-2 which is quite mild in nature, not easily transmitted, has a clinical latency duration of 20 – 40 years and it is less common among humans.

Our aim in this thesis is to use stochastic modelling to determine number of uninfected T4 cells, infected T4 cells and free HIV in an infected individual by examining the pathogenesis, progression and combined treatment of HIV. This is important because it helps in determining the efficacy of methods used in the research of pathogenesis, progression and combined treatment of HIV. We also looked at different ways that research has tried to go about eliminating the virus (section 2.3) in an infected person.

1.2 ACRONYMS AND TERMINOLOGIES

AIDS **A**cquired **I**mmune **D**eficiency **S**yndrome

APOBEC3G	**Apo**liprotein **B** mRNA **e**diting enzyme **c**atalytic polypeptide-like **3G**
$CD4^+$T cells	CD4 positive T lymphocytes
$CD8^+$T cells	CD8 positive T lymphocytes
DNA	**D**eoxyribo**n**ucleic **a**cid
Env	**Env**elope; precursor to envelope glycoproteins
Gag	**G**roup - **a**nti**g**en ; precursor to internal structural proteins
HIV	**H**uman **I**mmunodeficiency **V**irus
IT	**In**te**t**egrase
LTR	**L**ong **t**erminal **r**epeat
mRNA	**M**emory **R**ibo**n**ucleic **a**cid
PR	**Pr**otease
Pro	PR enzyme
Pol	**Pol**ymerase; precursor to RT and IT enzymes
Rev	regulates splicing/RNA transport
RNA	**R**ibo**n**ucleic **a**cid
RRE	**R**ev **r**esponse **e**lements
RT	**R**everse **t**ranscriptase
TAR	**T**rans**a**ctivation-**r**esponse element
Tat	activates transcription
Vif	affects infectivity of viral particles
vpr and/or vpx nef	is present in viron; has nuclear localization signal; facilitates infectivity in quiescent cells; triggers CD4 endocytosis, alters signal transduction in T cells; enhances viron infectivity
vpu	integral membrane protein; triggers CD4 degradation; enhances viron release

1.3 ROLE OF MATHEMATICAL AND STATISTICAL MODELING IN HIV/AIDS EPIDEMIC

Mathematical and statistical models of HIV/AIDS infection have become extremely important not only because medical scientists cannot combat the problems of these viruses alone (since not all problems can be replicated or solved experimentally as human lives are involved), but also to give better understanding of the HIV/AIDS epidemic and for reasons such as:

i. the models based on underlying transmission mechanism of the HIV/AIDS infection can help the medical and/or scientific world to understand and evaluate the epidemiology of these viral infections hence giving insight into different strategies of prevention and control that can be applied according to the severity of the epidemic in the different areas (Tan 2000)

ii. the mathematical and statistical models can provide qualitative insights even when data are lacking or not readily available, and this can help prioritize data collection (Hyman and Stanley 1988, Tan 2000)

iii. the models can be used to provide in-depth understanding of some basic features and principles of the epidemic and its pathogenesis, thus aiding in the study of suitable treatments and/or vaccine and maybe a cure in the near future (Tan 2000)

iv. the models can help reveal important parameters and co-factors of the infection and also shed light on their consequences

v. the impacts of risk factors may be assessed, thereby screening for important risk variables for the purpose of prevention and control of the infections (Hyman and Stanley 1988, Tan 2000)

vi. mathematical and statistical models based on the transmission of the infections can show how early or late infection, behavioural changes and medical advances such as

treatments and vaccines will affect the future course of the HIV/AIDS epidemic (Hyman and Stanley, 1988)

vii. the knowledge of parameters and co-factors can help develop both mathematical and statistical models which can give computer simulations to compare different treatment outcomes etc.; these computer simulations can save time, lives and resources as compared with using other means such as animals and running trials on humans

viii. models can be used to estimate unknown data on the basis of the known facts. For example, the past distribution of HIV infection can be estimated from the current AIDS caseload and the distribution of times from infection to AIDS (see Back-calculation method in section 3.4). To determine the consistency of the generated data a formal mathematical model similar to the one that was designed is required. The available data can also be assessed indirectly to determine their internal consistency by leaving some data out, generating estimates of the missing data based on one or more models, and then comparing the two data sets (lifted directly from Hyman and Stanley, 1988).

1.4 HIV MODELING

In the bid to combat the two deadliest viral infections in the 20th and 21st century, the onus have not been only on the medical scientist to find a cure but also partnership with mathematical, statistical, computational and engineering scientist have become inevitable. Hence the mathematical and computational modeling of HIV/AIDS have become a novel approach with great impact in the different areas of study of the epidemic. Among those who pioneered mathematical modelling (quantitation) of HIV is David Ho. His research into HIV/AIDS in the last 27 years has helped formed the basis for combined antiretroviral

treatments and also in understanding the dynamic nature of HIV replication *in vivo*. Research into the dynamics of HIV *in vivo* has helped in further understanding of the pathogenesis and growth of HIV, the replication and progression of the virus, means of determining HIV progression by T4 cell count or HIV-1 RNA viral load count, results of single and combined antiretroviral treatments, when to commence such treatment and also mathematical and computational modeling of these processes.

To ascertain the progression of HIV in an infected individual, the T4 cell count, the HIV-1 RNA viral and viral decay approach have not only become common but also reliable means used to predict the outcome of a patient in terms of duration to regressing to AIDS and also to determine when to commence ARV or HAART. Also methods that have permitted missing data analysis have become extremely important in HIV modelling because most patients don't know when they are infected and data on some patients are incomplete due to inconsistency in attending ART clinics. De Gruttola et al. (1991) modelled the progression of HIV infection using the T4 cells as its measure, more likely because of the availability of data on the T4 cell counts. They used the parametric linear growth curve model because it permits analysis of incomplete data assuming the data are missing at random. Also autoregressive models were fitted to short series of the T4 cell counts because this method allowed the estimation of annual decline averaged over all individuals. The setback of these methods was that the variability in the rates of decline of the T4 cell cannot be estimated and the modeling of the entire process from infection to AIDS cannot be done.

Tan and Wu (1998) developed a stochastic model for the interaction between $CD4^{+}$T cells and the human immunodeficiency virus. Stochastic differential equations were obtained for the numbers of uninfected T cells, latently infected T cells, actively infected T cells and free

HIV through binomial and multinomial distributions. They modelled the generation of new uninfected T cells by a pure birth process (Poisson process) and the growth of uninfected $CD4^{+}$T cells by simulating the antigens using a stochastic logistic pure birth process. The results of the Monte Carlo simulations showed that the probability distributions of the CD4+T cells and free HIV were skewed in the earlier stage of infection and eventually converged to normal distributions in later years.

Sridharah and Jayashree (1993) also used the stochastic point process to model the population of infected T4 cells. In the model, they made use of phases with special types of time-dependencies whose durations were independent and exponentially distributed. The first and second moments of the infected T4 cells were generated from explicit differential equations obtained.

Wu and Ding (1999) gave a model with a sum of exponentials which gave a good fit to the observed clinical data of HIV-1 dynamics i.e. HIV-1 RNA copies after starting antiretroviral treatments. The other advantage about this model was that it can also be used as a biological compartment model for the interaction between HIV and its host cells. Thus enjoying both worlds of biological interpretability and mathematical simplicity after re-parameterization and simplification. Finally the use of hierarchical nonlinear mixed-effect model approach for parameter estimation and other statistical inferences was illustrated using real life data.

Wu et al. (1999) revised four model-fitting procedures for biphasic viral decay data in clinical studies. This was because the estimates obtained when these methods were applied differed significantly. The methods were Single method, Perelson steady state (PSS) method, Wu and Ding (WD) method and Perelson and Neumann steady state (PNSS) method. Pros and cons

of these methods were discussed. For example, the simple method which fitted a bi-exponential model to the biphasic viral load was good because it included all data from baseline onward. The disadvantage of this method was the biased estimates of viral load obtained due to effect of the initial 'shoulder' that was ignored. Ding and Wu (1999) suggested the fitting of the model only after the effect of the 'shoulder' is considered.

Ding and Wu (1999) also worked in detail on the four model-fitting procedure given in Wu et al. (1999), evaluating the performance of these procedures through extensive use of Monte Carlo simulations. Guidelines on how to select appropriate method for data analysis was given and real life data was used to backup the guidelines.

Joshi (2002) derived an optimal control of an HIV immunology model by using a system of ordinary differential equation model taken from Kirschner and Webb (1998). This system of ODE described the interaction of the T4 cells and HIV in the immune system. He used the boundedness of solutions of the ODE system for finite time interval to prove the existence of an optimal control pair. Thus the optimal control pair obtained gave an optimal treatment strategy for the HIV infected patient under two types of drug treatments, namely, treatment that aimed at reducing viral population and treatment that aimed at improving the immune response. Joshi (2002) solved the optimality system by using an iterative method with a Runge-Kutta fourth order scheme. Joshi (2002) noted that the format of the optimal controls he obtained agreed with those of Butler et al. (1995), Kirschner et al. (1997) and Fister et al. (1998) where only one control instead of two was used.

Bortz and Nelson (2006) considered six deterministic models and made comparisons with respect to their ability to represent HIV infected patients undergoing antiretroviral treatment,

(to be precise reverse transcriptase mono-therapy). Bortz and Nelson (2005) created a statistical model using the hierarchical mixed-effects approach to characterize factors such as inter-individual and intra-individual variability in the patient population. Their aim was to derive mathematical model(s) of *in vivo* HIV infection dynamics. Bortz and Nelson (2005) were able to obtain higher viral clearance rate c as was done in earlier work by Louie et al. (2003) by using linear parameter fits as opposed to non-linear parameter fits.

Other method that have been used is the Bayesian modeling. Frost (2001) used this method to model the viral dynamics and evolution of HIV, Putter et al. (2002) estimated parameters in HIV dynamic models and Han et al. (2002) developed the Bayesian analysis method for the population dynamic HIV. Also Huang and Wu (2006) examined the Bayesian approach for estimation of antiretroviral efficacy.

1.5 THESIS OUTLINE

In this thesis, we have combated some of the issues of the two most deadly viruses namely human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS), that have invaded the human race in the last thirty years by concentrating on the stochastic modelling of the dynamics of the viruses. Although availability of efficient vaccines or cure for these infections is still like groping in the dark, medical scientists, pharmacologist, epidemiologists and even the mathematical and social scientists are eagerly working hand in hand to see a dream come true. The collaboration of medical scientists with scientist and theorists have in recent times made a big positive influence in better containing the viruses. This thesis has six chapters and they are outlined below.

In the first chapter an overview of the HIV is given and HIV modelling done by some scientists are reviewed. The second chapter deals with the pathogenesis of the viruses by delving into the genetic variation of the HIV. This is because the pathogenesis of the HIV infection can be understood only when the genetic variation in HIV and the receptor-specific HIV infection are given their due importance.

In chapter three, incubation period and seroconversion time are determined by using data on homosexuals given in Lui et al. (1988). Two stochastic models are used to determine the distribution function of the gay-life and the incubation period. Also the back-calculation method was used to project AIDS incidence.

Chapter four deals with the formulation of stochastic model of the dynamics of HIV in an infected individual. In this chapter, two stochastic models are proposed and analysed for the dynamics of the viral load in a HIV infected person and the multiplication process of the virons inside an infected T4 cell. Also numerical illustration of these stochastic models is given.

In chapter five, the T4 cell count which is considered one of the markers of disease progression in HIV infected individual is examined. WHO has recently advocated that countries encourage HIV infected individuals to commence antiretroviral treatments once their T4 cell count is 350 cells per ml of blood (was formerly 200 cells per ml of blood). This is because when the T4 cell count is low, the T4 cells are unable to mount an effective immune response against antigens and any such foreign matters in the body (Kirschner 1996) and consequently, the individual becomes susceptible to opportunistic infections and lymphomas. Thus, the T4 cell count can be considered a marker of disease progression in an

infected individual and the loss of T4 cells accounts for a major part of the immunosuppressive effect of HIV. As such, a stochastic catastrophe model is developed to obtain the mean, variance and covariance of the uninfected, infected and lysed T4 cells. Also obtained are the amount of toxin produced in a HIV infected person from the time of infection to the present time. Numerical illustration of the correlation structure between uninfected and infected T4 cells, and infected and lysed T4 cells is portrayed.

To combat the persistent death of humans before any cure can be obtained, antiretrioviral drugs were introduced to suppress the havoc done by these viruses in the human body. Treatment with single drug failed due to the fact that HIV evolved rapidly because of its high replication rate of an average of 10^{10} viral particles per day. Thus drug resistance to single therapeutic treatment in HIV infected individuals has promoted the research into combined treatments. Hence, in the sixth chapter a stochastic model under combined therapeutic treatment by extending the model of HIV pathogenesis under treatment by anti-viral drugs given by Perelson et al. (1996) is derived. Mean numbers of free HIV, infectious free HIV and non-infectious free HIV are obtained. Variance and co-variance structures of our parameters were obtained unlike in previous work of Perelson et al. (1996) and Tan and Xiang (1999). Comparison of simulated data for before and after treatment indicates the efficacy of our model in combined treatment.

CHAPTER TWO

PATHOGENESIS OF HIV

2.1 INTRODUCTION

The human immunodeficiency virus (HIV) is the early stage of the acquired immunodeficiency syndrome (AIDS) in which within 24 hours of contact, the virus replicates its RNA into the victim's DNA and as such the protein (gp120) on the virus binds to the protein on the $CD4^{+}$T cell thus affecting the immune response of the victim (Kirschner 1996). New virus particles then bud from the host cell after the duplication process. Thus the HIV virus replicates, mutate, recombine and bud off the host cell and it is the budding and maturity that determines both the duration (of transition from HIV to AIDS) and stage of infection either as the human immunodeficiency virus (HIV) or the acquired immunodeficiency syndrome (AIDS). This is because the HIV infection could be asymptomatic for years and only develops to AIDS when the $CD4^{+}$T cells fall so low due to increase in the viral load of the host cell. Holmes (1998) stated that mutation, recombination and natural selection produced a multitude of different genomes which allow the virus to continually evade immune response and to infect a variety of cell types, and the potential of HIV to evolve at a rate of about 1 million times faster than human nuclear DNA have undermined attempts to produce effective vaccines and allowed the development of resistance to some antiviral treatments within a matter of months. Thus, medicine, science and engineering have continually researched the pathogenesis of the human immunodeficiency virus (HIV) infection and mechanisms of genetic variations of HIV.

To understand HIV pathogenesis, the unique nature of the causative microbe was studied and compared with lentivirus infections of animals (because it shared features with other members of the non-transforming and cytopathic lentivirus family of retroviruses) to raise further questions regarding human disease such as (Weiss, 1993): Why do some horses permanently recover from equine infectious anemia when the virus evolves immune escape

variants as readily as HIV? Can the wasting syndrome and brain disease of sheep infected with visna-maedi virus be equated to human AIDS without CD4 depletion? And hypotheses such as AIDS being the end-stage disease of HIV, HIV mutating to produce different types of HIV, have been postulated to explain the relationship between HIV and AIDS. However, research is still done on why AIDS finally develop, if it is the HIV alone that brings about AIDS or maybe there are other viruses, why it takes a variable long time for HIV to develop to AIDS, and also the cofactors that influence the rate at which AIDS develop and so on.

Recent research by scientists such as Smith (2006), Sodora and Silvestri (2008), Levy (2009) showed that new data especially from non-human primate studies have raised doubts about the 1990's hypothesized theory that HIV-1 causes CD4+T cell depletion by direct cytopathic effect. Rather it has been shown that the immune activation of the virus causes the cell depletion. Thus shedding light on the research to see if HIV alone brings about AIDS or maybe there are other viruses. Hence they have strongly advocated a full understanding of HIV/AIDS pathogenesis which may lead to novel therapies (partially quoting Smith 2008). Also according to Hoffmann et al. (2007), an understanding of the immunopathogenesis of HIV-1 infection is a major prerequisite for rationally improving therapeutic strategies, developing immuno-therapeutics and prophylatic vaccines. Hence the delving into pathogenesis of the human immunodeficiency virus in this chapter.

2.2 HIV PATHOGENESIS

The pathogenic mechanisms of HIV disease are extremely complex and multifactorial (Fauci 1993, 2003). And in cases of the acquired immune deficiency syndrome (AIDS), marked depletion of $CD4^+$ T cells was recognized as a hallmark of disease early on (Gottlieb et al. 1981, Maseur et al. 1981, Fauci 2003), even before the classic demonstration in 1984 that the

CD4 molecule was the primary receptor for the virus on a subset of T cells and monocytes (Dalgleish et al. 1984, Klatzmann et al. 1984, Fauci 2003). Also much evidence has suggested that other factors were necessary for HIV fusion and entry, but these factors such as the co-receptors and chemokines remained elusive for several years (D'Souza and Harden 1996). According to Fauci (1996), D'Souza and Harden (1996) in the mid-1990s, a number of diverse areas of investigation elucidated the roles of the chemokine receptors CXCR4 and CCR5 in the efficient binding and entry of two different strains of HIV-1 called X4 and R5, respectively. The discovery that HIV could use different co-receptors also helped to explain the occurrence of syncytial (CXCR4-using) and nonsyncytial (CCR5-using) variants of HIV (Fauci 1996). The importance of the CCR5 co-receptor in the pathogenesis of HIV infection was proven by the finding that cells from individuals homozygous for a deletion of 32 base pairs in the CCR5 gene could not be infected in vivo with R5 viruses and that such individuals (who comprise about 5% of white populations) were thought to be extremely resistant to HIV infection even when repetitively exposed to virus until recent research proved otherwise and hence they can be termed as long-time progressors (O'Brien and Moore 2000, Fauci 2003, Hoffmann et al. 2007, Levy 2009).

The ability to measure plasma viremia precisely led to the classic viral dynamics studies of HIV. HIV research by mathematical scientists have tremendously helped in understanding the relationship between virus production and T cell dynamics (Ho et al. 1995, Wei et al. 1995, Fauci 2003). These studies led to a better insight of the HIV pathogenesis, hence making therapeutic treatments better and less toxic. Studies have shown that even in individuals in whom plasma viremia is driven by antiretroviral therapy to levels of less than 50 copies of RNA per ml ('undetectable') for up to 3 years, the viral reservoir persists and the virus rebounds from this reservoir within weeks of discontinuing therapy (Blackson et al. 2002,

Fauci 2003). Hence one may paradoxically say that studies of the immune response to HIV have been both productive and frustrating. Although individuals in whom HIV infection has been established cannot eliminate the virus from their bodies, continual research into better prophylaxic vaccines through the better understanding of the pathogenesis of these viruses still continues (Chun and Fauci 1999, Blackson et al. 2002, Fauci 2003). (Excerpts from Fauci 1993, 2003)

2.2.1 HIV Structure

HIV has a dense cylindrical core. It is around 120nm in diameter (120 billionths of a meter; around 60 times smaller than a red blood cell) and 10kb in length and roughly spherical. It is composed of two copies of single-stranded RNA enclosed by a conical capsid comprising the viral protein p24 (figure 2.1). This conical capsid can be described in layman's language as being bullet shaped. The RNA component is 9749 nucleotides long and it is surrounded by a plasma membrane of host-cell origin. The RNA is part of a protein-nucleic acid complex which is composed of the nucleo-protein p7 and the reverse transcriptase (RT) p66. The single-strand RNA is tightly bound to the nucleocapsid proteins p7 and enzymes such as reverse transcriptase (RT) i.e. p66, protease (PR) i.e. p11 and integrase (IT) i.e. p32 that are indispensable for the replication, proliferation and development of the viron. The nucleocapsid (p7 and p6) associates with the genomic RNA (one molecule per hexamer) and protects the RNA from digestion by nucleases. The ends of each strand of HIV RNA has an RNA sequence called the long terminal repeat (LRT). The LRT has regions which act as switches to control production of new viruses.

Surrounding this capsid is the matrix layer which is made up of the protein p17 and this ensures the integrity of the viron particle. Also enclosed within the viron particle are genes

such as Vpr, Nef, Vif, p7 and viral protease. These are genes that code the proteins used in controling the ability of the virus to infect a cell and produce new copies of virus and/or cause disease. The outer viral envelope which is formed when the capsid buds from the host cell, taking some of the host-cell membrane with it is a coat of lipoprotein membrane fat. Projecting from this viral envelop/membrane are 72 little spikes formed from the glycoproteins gp120 and gp 41 (HIV Wikipedia 2008, Hoffmann et al. 2007 and Smith 2008).

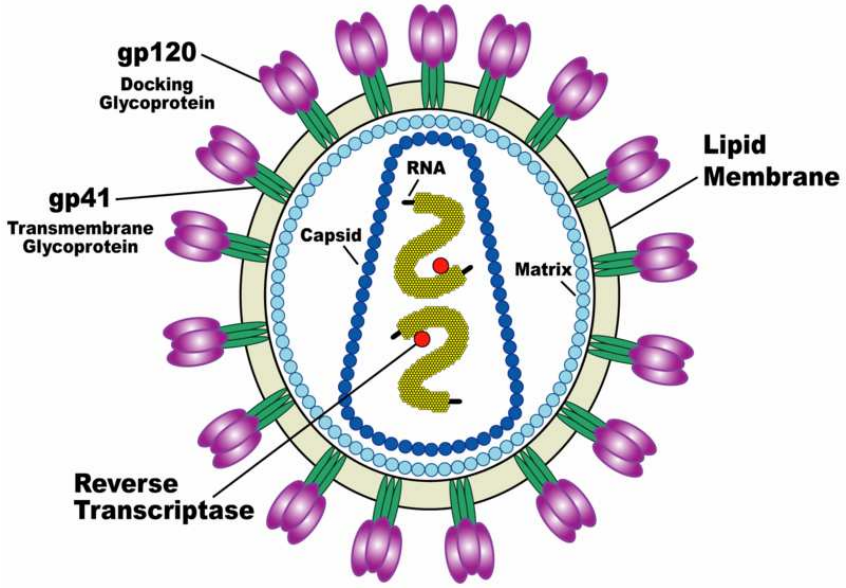

Figure 2.1 HIV genome showing the proteins involved in RNA coding and replication (Excerpt from HIV Wikipedia 2008)

2.2.2 Genes and Enzymes in HIV Entry and Replication

In HIV-1 there are 9 primary genes that encode within the RNA genome, namely: gag, env, pol, tat, rev, nef, vpr, vpu and vif. These genes and certain enzymes play different crucial roles in the entry and replication of the virus in the host cell. According to Fauci (2003), the identification of their relationship to the complex mechanism of HIV replication have been crucial in understanding HIV replication and its relationship to the pathogenic mechanism of the disease. Recent developments in controlling the destroying effects of the virus in the human body via development of effective antiretroviral drugs have also concentrated on some of these genes and enzymes (see section 2.3). These genes and enzymes and their functions are listed below.

i. Gag:

This encodes for the nucleocapsid and the glycoproteins gp 120 and gp 41 of the viral membrane.

ii. Env:

This codes for the glycoprotein gp 160 that is then broken down by a viral enzyme to form gp 120 and gp 41.

iii. Pol:

This codes for the reverse transcriptase (RT) and other enzymes.

iv. Tat:

This is a regulatory protein that accumulates within the nucleus and binds to the TAR found in the LRT of the viral RNA. It is a potent transcriptional activator of the LRT and its importance is in the *in vivo* culture system viral replication.

v. Cyclin TI:

It is a necessary cellular cofactor for tat.

vi. Rev:

This gene regulates splicing. It is also nuclear export factor which is important for switching from the early expression of regulatory protein to the structural proteins that are synthesized later. Both tat and rev stimulate the transcription of proviral HIV-1 DNA into RNA, promote RNA elongation, enhance the transportation of HIV RNA from the nucleus to the cytoplasm.

vii. Nef:

This codes for virus efficient replication. It may induce down-regulation of CD4 and HLA class I molecules from the surface of the infected cells. Thus the virus avoid recognition by CD4+T cells and hence evades any attack mediated by cytotoxic CD8+T cells. It is also essential for the high rate of virus production and progression of disease. It sometimes interfere with T cell activation by binding to various proteins that are involved in intracellular signal transduction pathways, thus helping in the disease progression.

viii. Vpr:

It is used in viral replication in non-dividing cells. it also stimulates the HIV LTR, promotes cellular and viral responses and its important for the transport of the viral pre-integration complex to the nucleus.

ix. Vpu:

This encoded protein influences the formation of new virons by allowing the recycling of gp 160. It also influences the release of new virus particles from infected cells by getting

involved in the degradation of CD4-gp 160 complexes within the endoplasmic reticulum. Thus it is important for the virus budding process.

x. Vif:

It supports viral replication.

xi. APOBEC3G:

Is an enzyme of the intracellular enzymes family. Its function is to deaminate cytosine to uracil in mRNA or DNA. APOBEC3G is expressed in lymphocytes and macrophages which are the primary target cells of HIV infection. In the presence of vif gene, it is complexed, degraded and not incorporated in newly formed virons.

xii. HLA class I:

xiii. HLA class II:

xiv. Others:

These are cellular binding proteins which have been found in the last 10 years (Levy 2009) to be associated with the HIV infection. They include C type lectings – DC-SIGN, Leukocyte function-associated antigens (LFA), Intercellular dhension molecules (ICAMs), $\alpha 4\beta 7$ integrin which acts as an HIV binding site particularly on $CD4^+$ memory T cells.

2.2.3 HIV-1 Strains

HIV-1 strains are classified by the cells they infect. Some of the HIV-1 strains are listed below.

i. Macrophage (M- tropic) strains

They are also known as the Non-syncitia-inducing strains (NSI). They gain entry through the β-chemokine receptor CCR5. Replication of this strain occur in the macrophages and the $CD4^{+}$T cells.

ii. T-tropic isolates strains

They are also known as the syncitia-inducing strains (SI). They gain entry through the α-chemokine receptor and CXCR5. Replication of this strain occur mainly in the $CD4^{+}$T cells and some in the macrophages.

iii Dual-tropic strains

They are also known as the transitional strains of the HIV-1. They use both the CCR5 and CXCR5 for co-receptors. Replication of this strain occur mainly in the $CD4^{+}$T cells and some in the macrophages.

2.2.4 HIV Co-receptors

According to Hoffmann et al. (2007), experiments using non-human cell lines transfected with human CD4 showed that expression of human CD4 on the cell surface of a non-human cell line was not sufficient to allow entry of HIV. Hence the existence of human co-receptors necessary for viral entry was postulated. Co-receptors are chemokines of the cytokine super-family. The chemokines are group of small proteins that mediate leukocyte traffic through specific receptors. They are involved in several human reproductive events such as sperm chemotaxis (i.e. carrying around of sperms), ovulation, implantation of embryo during conception, menstruation e.t.c. Also HIV-1 uses the chemokines as entry into the individual cell(s). There two types of chemokines namely the α-chemokines (these use the α- receptors) and the β-chemokines (these use the β- receptors). In layman's language, co-receptors are

elements that receive the virus or help the virus to gain entry into the targeted cells. Table 2.1 shows the strains of HIV-1 and their chemokines and co-receptors.

Table 2.1 HIV-1 Strains and their Chemokines and Co-receptors.

Strain of HIV-1	Type of chemokine receptor	Type of Co-receptor for entry	Cells tropism
Macrophage	β	CCR5	Macrophages, CD4+T cells
T-tropic	α	CXCR4	CD4+T cells, Macrophages
Dual-tropic		CCR5, CXCR4	
Others		CCR3, CCR2, CCR8, CCR9, STRL33 (Bonzo), Gpr 15 (Bob), Gpr 1	

2.2.5 HIV-1 Subgroup, Recombination and Epidemiological Structure

Although once an individual becomes infected, eradication of the virus still remains impossible despite all the therapeutic advantages achieved during the last decade, knowledge of the epidemiological prevalence can still help to contain the disease to a certain degree (Hoffmann et al. 2007). There are three subtypes of HIV-1 namely: M group or the “major” group, O group or the “outlier” group and N group or the “new” group. Each group is divided into subtypes and their recombined subtype known as the circulating recombinant forms (CRFs). Given below are the HIV-1 subgroups and their epidemiological prevalence.

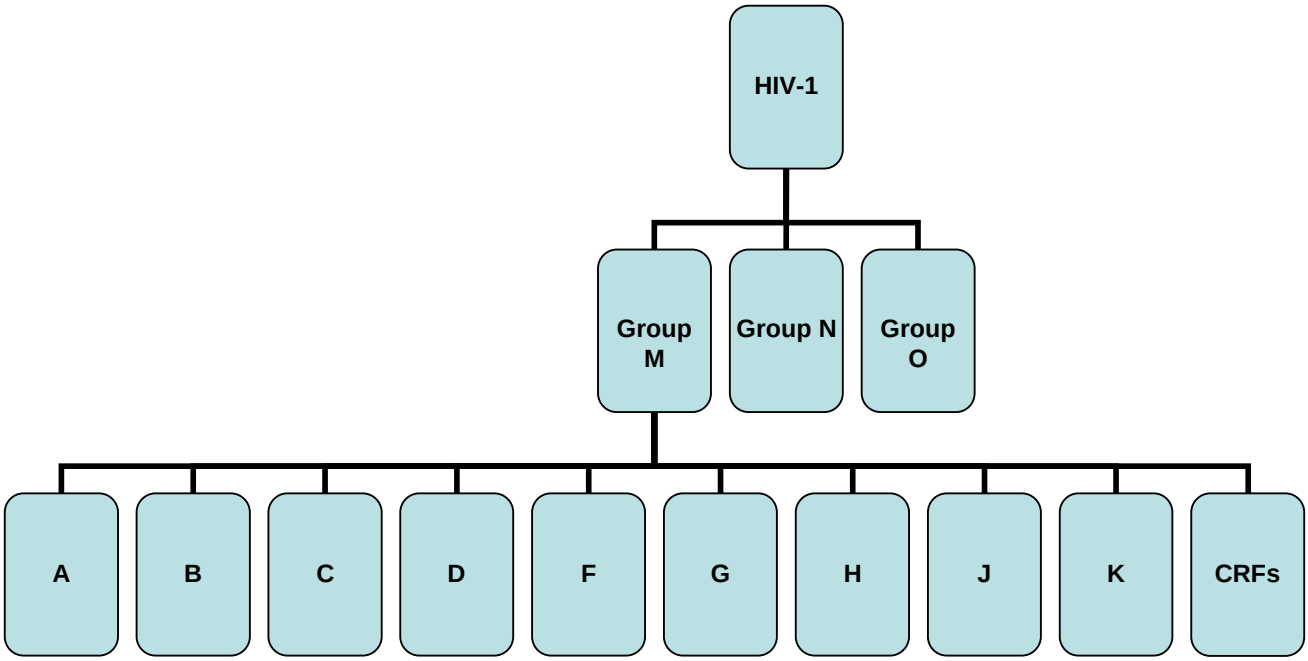

Figure 2.2 HIV-1 Group and Subtypes

Table 2.2 HIV-1 Subtypes and Regional Prevalence

HIV-1 Subtype	Region
A	West Africa, Central Africa, Russia
Crf A/G	West Africa, East Africa, Central Europe
Crf A/E	South-East Asia but originated from Central Africa
B	Europe, America, Japan, Australia
C	Southern Africa, East Africa, India, Nepal
D	East Africa, Central Africa
F	Central Africa, South America, Eastern Europe
G	West Africa, East Africa, Central Europe
H	Central Africa
J	Central Africa
K	Democratic Republic of Congo (DRC), Cameroon

2.3 RECENT DEVELOPMENTS AND PROBLEMS

In recent times scientists have come up with a new way of combating both the human immunodeficiency syndrome (HIV) and acquired immunodeficiency syndrome (AIDS) despite the absence of a cure for them. This recent discovery is still in the pipeline, but it involves the attack of reservoirs of dormant HIV. There are two reservoirs namely macrophages and memory T cells. The macrophages which are antigen scavengers usually engulf antigens in the body and afterwards the macrophages die while the memory T cells retain the whole process of attacking and building forces against antigens in the body so that a reoccurrence of such attack does not come into play (Kirschner 1996).

In the presence of HIV the macrophages still live long past their survival time and hence become a hideout for the virus. The inability of macrophages to die after being infected by the virus is caused by enzyme called Akt which is a protein produced by a cell-survival pathway of the virus. To combat the infected macrophages which is one of the habouring stations of the virus, drugs such as miltefosine and perifosine were used and these two rapidly killed the infected macrophages. Although perifosine is currently being studied as a possible cancer drug, miltefosine on the other hand is known to be safe in leishmaniasis patients, hence further research on the possible effects of using these two drugs to destroy infected macrophages.

Although recombinant viruses forming between HIV clades and groups have occurred due to co-infection and super-infection of cells by two or more virus strains/types usually prior to the establishment of a chronic infection. Recombination between HIV-1 and HIV-2 is impossible because of the differences in the location of the RNA dimmer hairpin sites (Dirac et al. 2002, Levy 2009). Hence recombinant viruses have posed a big problem in controlling the infection by administering antiretroviral drugs for too long because most times resistance to the drugs and poor immune response usually occur (Fultz 2004, Levy 2009).

Eradication of the virus in an infected human body has become impossible because the virus infects not only cells in the body, but also cells in the cellular and immune system. Also the virus is evident not only in the blood, but also in cells and different compartments of the body (Levy 2009).

With the emergence of new antiviral therapies especially the combined treatment, great hope to those at risk of advancing to AIDS have been brought. However long-term therapy

treatments may not be feasible because of toxic drug side effects such as liver damage and drug resistance due to fast mutation of the virus and/or recombination of different types of the virus (Khalili and Armaou 2008, Levy 2009, Hoffmann et al. 2007).

Other avenues explored in eradicating the virus or slowing down cell activation include: Using antibodies that attach to virus-infected cells via gp 120 or gp 41 to directly kill infected cells through antibody-directed cellular cytotocity (ADCC). Targeting intracellular protein needed for HIV replication by using anti-HIV therapy; for example Vpu was shown to reduce the activities of the human cellular membrane protein called tetherin with the help of a calcium-modulating cyclophilin ligand, thus blocking the budding of the virus from the cell surface. Also deficiency of Vif has shown in studies in Australia to help in delaying onset of AIDS whence such individuals have been able to stay as long-term AIDS progressors.

While these researches are in pipeline, work still continue in the detailed understanding of these two deadly viruses that have sacrilegiously and despicably invaded and destroyed the peace of the human race in the last thirty years.

CHAPTER THREE

A STOCHASTIC POINT PROCESS MODEL OF THE INCUBATION PERIOD OF A HIV INFECTED INDIVIDUAL

3.1 INTRODUCTION

Acquired Immune Deficiency Syndrome (AIDS) is a sure fatal but containable disease caused by the retrovirus HIV. It is found that there is a risk of contracting HIV infection from exposure to infected persons. The exposure can be through sharing of intravenous hypodermic needle with infected persons, transfusion of HIV infected blood, mother-to-child transmission at birth or performing a sexual act with HIV infected persons. As sex plays a major important role in human life, the virus has the vulnerability of being quickly transmitted from one infected individual to either an infected or non-infected individual by the pattern of their intimate behaviour. Since the behaviour is highly stochastic, the time for a susceptible to become an infective is unpredictable. Whence, the dynamics of the spread of HIV presents several perplexing difficulties in its comprehension even in the case of a specific community such as a population of transfusion related cases of AIDS (Medley et al. 1988). The foremost difficulty that is baffling the model builders is the incubation period of HIV. The incubation period (**IT**) of HIV in an infected individual is the period from the time of infection to the time of the first diagnosis of an opportunistic disease associated with AIDS. And according to Medley et al. (1988), one of the striking features of acquired immunodeficiency syndrome (AIDS) is that the incubation period appears to be both long and very variable. Usually, the time of infection is not known in several cases. However, the seroconversion time (**ST**) (i.e., the time at which an infected individual becomes HIV positive) may be known in many cases. The latent period, namely, the interval between the time of infection and the time of seroconversion is small (in weeks) compared to the incubation period (in years) of HIV. Hence, the time of infection is taken to be the time of seroconversion.

Studies on HIV incubation period have been carried out. For instance, Medley et al. (1988) in their study observed that the data on the time of infection was incomplete and estimated mean incubation period to be 4.5 years to 15 years.

Chevret et al. 1992 developed a new approach for estimating the incubation period of acquired immunodeficiency syndrome (AIDS) based on age distributions. They expressed the Incubation period as the difference between age at time of diagnosis and age at time of contamination. By assuming independence between age at time of infection and incubation period, the age distribution of newly diagnosed AIDS cases was given as the convolution product between the distributions of the age of freshly infected patients and of the incubation times. Hence, AIDS incubation time could therefore be estimated from the age distribution of newly HIV infected subjects and newly diagnosed AIDS cases.

Lee (1999) estimated the maturity of the HIV infection and the incubation period of AIDS by using data from 363 seroprevalent (i.e. those who were AIDS free at entry) Korean AIDS patients (including 59 seroincident cases). He proposed two methods for imputing the unknown times since seroconversion which were, firstly fitting Weibull regression with the marker of matured CD4+T cell count for seroincident cohorts, and secondly, using a random effects model with CD4+T cell count as a response for repeated measures from which the times since seroconversion can inversely be extracted.

Rao and Kakehashi (2005) estimated HIV incidence density from prevalence data and also the incubation time distribution by using the deconvolution technique and maximum likelihood method to estimate parameters. The difference was that their data was not based on homosexual men/women.

Several mathematical and statistical analyses have been proposed in the recent past to assimilate the data and provide information about the dynamics of the epidemic (Anderson and May 1991). In the statistical analyses of the data, the gamma, Gompertz, Lognormal, Normal and Weibull distributrions were used to model the distribution function F(t) of the incubation period (Brookmeyer and Gail 1994, Anbupalam et al. 2002). The advantages and disadvantages of using each of these models are outlined in Brookmeyer and Gail (1994). In particular, the Weibull model is used in situations where it is hypothesized that the hazard function λ(t) increases indefinitely and is proportional to a power of time from infection (Brookmeyer and Gail 1994). The hazard function quantifies how the risk of AIDS evolves with time from infection and is given by

$$\lambda(t) = \frac{\mathrm{f(t)}}{\mathrm{S(t)}}$$

where $f(t) = \mathrm{F}'(\mathrm{t})$ and S(t) = 1 – F(t) are the probability density function (p.d.f) and the survival function (s.f) of the incubation period respectively. However, as Brookmeyer and Gail (1994) have pointed out, the hazard function λ(t) should be consistent with epidemiological data and with theoretical considerations of the pathogenesis of HIV infection. Not much attention has been paid to the formulation of the distribution functions (hence the hazard functions) of the latent and the incubation periods by considering the stochastic behavioural aspects of the members of the population under study.

In this chapter, two stochastic models are presented namely:

i. Model I which is devoted to the determination of the distribution function of the gay-life (i.e. the time period from the entry of a susceptible in the specified community till he/she tested HIV positive) of a susceptible.

ii. Model II which determines the distribution function for the incubation period (i.e., the period from the time of seroconversion till the onset of overt symptom of AIDS).

Essentially, a two-parameter family distribution function for the gay-life and a one-parameter family of distribution for the incubation period are obtained. It is observed that the distribution function of the incubation period serves as a good fit for the data provided by Lui et al. (1988). Further, the distribution function is used to project AIDS incidence by back-calculation (Brookmeyer and Gail 1994).

The lay-out of this chapter is as follows: In section 3.2, a stochastic model for the determination of the p.d.f q(t) of the time interval ST between the time of entry of an individual into a population of homosexuals and the time of his/her seroconversion (becoming HIV positive) is proposed. In section 3.2.1 a two-parameter family of the probability distribution function of ST is obtained. The moments of ST are obtained in section 3.2.3 and the problem of estimation of the parameters of q(t) is considered in section 3.2.4. In section 3.3, a stochastic model for the determination of the probability function p_n of the incubation period (IT) is proposed. A one-parameter family of the probability function p_n of IT is obtained in section 3.3.1 while the moments of IT are obtained in section 3.3.2. The problem of estimation of the parameter of p_n is considered in section 3.3.3 and illustrated by a numerical example in section 3.3.4. The method of back-calculation is used in section 3.4 to obtain AIDS projection for a sample data.

3.2 A STOCHASTIC MODEL OF THE PERIOD OF THE GAY-LIFE

Consider a population of homosexuals consisting of susceptibles and infectives. Assume that at time t = 0, a new member who is tested HIV negative enters into the population and makes sexual contacts with members of the population. Assume further that his/her contacts occur at random time points which follow a Poisson process with parameter λ, λ > 0. Let the probability that the individual who has already had n contacts up to time t when he/she tested HIV positive for the first time in the interval (t, t+Δ) be given by

$$n\mu\Delta + o(\Delta),\ \mu > 0.$$

Let the gay life period of the individual be represented by the random variable ST. in the next section, we obtain the probability density function (p.d.f) of ST.

3.2.1 The Probability Distribution Function of the Gay-life

We define the p.d.f of ST by

$$q(t) = \lim_{\Delta \to 0} \frac{\Pr\{t < \mathrm{ST} < \mathrm{t} + \Delta\}}{\Delta}$$

Then q(t)Δ represents the probability that the individual tests HIV positive for the first time in the interval (t, t+Δ). At least one contact is needed to get infected with HIV, and also using probabilistic rules, we obtain

$$q(t) = \mathrm{e}^{-\lambda \mathrm{t}}\lambda \otimes e^{-(\lambda+\mu)t}\mu + \mathrm{e}^{-\lambda \mathrm{t}}\lambda \otimes \times \sum_{\mathrm{n}=2}^{\infty} \mathrm{e}^{-(\lambda+\mu)\mathrm{t}}\lambda \otimes \ldots \otimes e^{-(\lambda+(\mathrm{n}-1)\mu)t}\lambda \otimes e^{-(\lambda+n\mu)t} n\lambda \qquad (3.2.1.1)$$

Taking Laplace transform on both sides of 3.2.1.1 we get

$$q^*(s) = \frac{\lambda}{s+\lambda} \sum_{\mathrm{n}=1}^{\infty} \frac{n\lambda^{n-1}\mu}{(s+\lambda+\mu)\ldots(\mathrm{s}+\lambda+\mathrm{n}\mu)} \qquad (3.2.1.2)$$

Splitting into partial fractions, equation 3.2.1.1 yields

$$q^*(s) = \lambda \sum_{n=1}^{\infty} \frac{1}{(\mathrm{n}-1)!} \left(\frac{\lambda}{\mu}\right)^{(n-1)} \left\{ \sum_{j=0}^{\mathrm{n}} \binom{n}{j} (-1)^j \frac{1}{(s+\lambda+\mathrm{j}\mu)} \right\} \qquad (3.2.1.3)$$

Inverting 3.2.1.3, we obtain explicitly the p.d.f of ST given by

$$q(t) = \lambda e^{-\lambda t}(1 - e^{-\mu t})e^{\lambda \frac{(1 - e^{\mu t})}{\mu}} \tag{3.2.1.4}$$

The frequency curve for ST for various values of λ and μ can be obtained by using

$$t_{mode} = \frac{1}{\mu}\log\left(\frac{2\lambda}{2\lambda + \mu - \sqrt{\mu^2 + 4\lambda\mu}}\right) \tag{3.2.1.5}$$

The distribution function Q(t) is given by

$$Q(t) = \frac{\lambda}{\mu}\int_0^{1-e^{-\mu t}} u(1-u)^{\frac{\lambda}{\mu}-1} e^{\frac{\lambda}{\mu}u}\,du \tag{3.2.1.6}$$

If λ = μ = λ, then

$$Q(t) = 1 - e^{-\lambda t}e^{1-e^{-\lambda t}}$$

In this case, the hazard function λ(t) is given by

$$\lambda(t) = \lambda(1 - e^{-\lambda t})$$

It can be observed that the hazard rate is increasing monotonically, which agrees with Brookmeyer and Gail (1994). In the next section, the moments of ST are obtained using equation 3.2.1.3.

3.2.2 The Moments of ST

The k-th moment of ST is given by

$$E[ST^k] = (-1)^k\left[\frac{d^k}{ds^k}\{q^*(s)\}\right]_{s=0}$$

Consequently, from 3.2.1.3, we obtain

$$E[ST^k] = k!e^{\frac{\lambda}{\mu}}\sum_{j=0}^{\infty}\frac{(-1)^j}{j!(\lambda + j\mu)^k}\left(\frac{\lambda}{\mu}\right)^j \tag{3.2.2.1}$$

For the particular case $\lambda = \mu = \lambda$, the mean and variance of ST obtained from equation 3.2.2.1 are given by

$$E[ST] = \frac{e-1}{\lambda} \tag{3.2.2.2}$$

$$Var[ST] = \frac{2e}{\lambda^2}\sum_{n=0}^{\infty}\frac{(-1)^j}{(j+1)(j+1)!} - \left(\frac{e-1}{\lambda}\right)^2 \tag{3.2.2.3}$$

The parameters of q(t) are obtained in the next section by using the method of maximum likelihood.

3.2.3 Estimation of the Parameters of q(t)

The likelihood function $L(\lambda,\mu)$ for a sample of size n is given by

$$L(\lambda,\mu) = \lambda^n \exp\left(-\lambda\sum_{i=1}^{n} t_i\right)\prod_{j=1}^{n}(1-e^{-\mu t_j})\exp\left\{\frac{\lambda}{\mu}\left(n - \sum_{k=1}^{n} e^{-\mu t_k}\right)\right\}$$

The logarithm of L is given by

$$\log_e L = n\log\lambda - \lambda\sum_{i=1}^{n} t_i + \sum_{j=1}^{n}\log(1-e^{-\mu t_j}) + \frac{\lambda}{\mu}\left(n - \sum_{k=1}^{n} e^{-\mu t_k}\right) \tag{3.2.3.1}$$

When $\log_e L$ reaches its maximum value, the values of λ and μ satisfy the following simultaneous equations:

$$n(\lambda+\mu) - \lambda\mu\sum_{i=1}^{n} t_i - \lambda\sum_{k=1}^{n} e^{-\mu t_k} = 0 \tag{3.2.3.2}$$

$$\mu^2\sum_{j=1}^{n}\frac{t_j e^{-\mu t_j}}{1-e^{-\mu t_j}} + \lambda\mu\sum_{k=1}^{n} t_k e^{-\mu t_k} - n\lambda + \lambda\sum_{k=1}^{n} e^{-\mu t_k} = 0 \tag{3.2.3.3}$$

From equation 3.2.3.2, we obtain

$$\lambda = \frac{n\mu}{\mu\sum_{j=1}^{n} t_j + \sum_{k=1}^{n} e^{-\mu t_k} - n} \tag{3.2.3.4}$$

Substituting 3.2.3.4 into 3.2.3.3, we obtain the following transcendental equations for μ:

$$\mu\left(\sum_{j=1}^{n}\frac{t_j e^{-\mu t_j}}{1-e^{-\mu t_j}}\right)\mu\sum_{j=1}^{n}t_j+\sum_{k=1}^{n}e^{-\mu t_k}-n+n\left(\mu\sum_{k=1}^{n}t_k e^{-\mu t_k}-n+\sum_{k=1}^{n}e^{-\mu t_k}\right)=0 \qquad (3.2.3.5)$$

Equation 3.2.3.5 can be solved using Newton-Raphson algorithm (Sastry 1994). Accordingly, we put

$$\psi(\mu)=\mu\left(\sum_{j=1}^{n}\frac{t_j e^{-\mu t_j}}{1-e^{-\mu t_j}}\right)\left(\mu\sum_{j=1}^{n}t_j+\sum_{k=1}^{n}e^{-\mu t_k}-n\right)+n\left(\mu\sum_{k=1}^{n}t_k e^{-\mu t_k}-n+\sum_{k=1}^{n}e^{-\mu t_k}\right)=0 \quad (3.2.3.6)$$

Then if $\mu^{(0)}$ is an initial approximate value of μ, then the $(l+1)$th iterate of μ is given by the equation

$$\mu^{(l+1)}=\mu^{(l)}-\frac{\phi(\mu^{(l)})}{\phi'(\mu^{(l)})},\ l=0,1,\ldots \qquad (3.2.3.7)$$

The iterative scheme given by equation 3.2.3.7 is the Newton-Raphson algorithm.

3.3 A STOCHASTIC MODEL OF THE HIV INCUBATION PERIOD

Assume that an individual has tested HIV positive for the first time at time t = 0. Let the conditional probability that he/she shows the first identifiable symptoms of AIDS during the n-th year given that he/she has not shown any symptoms of AIDS in the previous years be given by

$$1-e^{-n\mu},\ n=1,2,\ldots,\ \mu>0$$

Let IT be the random variable representing the incubation period. In the next section, a one-parameter family of distribution functions of IT is obtained.

3.3.1 The Probability Distribution of the Incubation Period

Let the probability function of IT be defined by

$$p_n = \Pr\{IT = n\}$$

Then p_n represents the probability that the individual shows the first symptom of AIDS in the n-th year. By using probabilistic rules, we obtain

$$p_n = e^{-\mu} e^{-2\mu} \ldots e^{-(n-1)\mu}(1 - e^{-n\mu}), n = 1, 2, \ldots \quad (3.3.1.1)$$

Simplifying equation 3.3.1.1 yields

$$p_n = e^{-\frac{(n-1)n}{2}\mu} - e^{-\frac{n(n+1)}{2}\mu}, n = 1, 2, \ldots \quad (3.3.1.2)$$

The mode l of the distribution is given by

$$e^{-\frac{(l-2)(l-1)}{2}\mu}(1 - e^{-(l-1)\mu}) \le e^{-\frac{l(l-1)}{2}\mu}(1 - e^{-l\mu}) \le e^{-\frac{l(l+1)}{2}\mu}(1 - e^{-(l+1)\mu}) \quad (3.3.1.3)$$

The median θ of the distribution is given by

$$1 - e^{-\frac{\theta(\theta+1)}{2}\mu} = \frac{1}{2} \quad (3.3.1.4)$$

From equation 3.3.1.4, we have

$$\mu\theta(\theta+1) - 2\log 2 = 0 \quad (3.3.1.5)$$

Solving equation 3.3.1.5, the median is given by

$$\theta = \frac{\sqrt{\mu^2 + 8\log 2} - \mu}{2\mu} \quad (3.3.1.6)$$

3.3.2 The Moments of Incubation Period

The mean of IT is given by

$$\begin{aligned} E[IT] &= \sum_{n=1}^{\infty} np_n \\ &= \sum_{n=1}^{\infty} n\{e^{-\frac{(n-1)n}{2}\mu} - e^{-\frac{n(n+1)}{2}\mu}\} \\ &= \sum_{n=0}^{\infty} e^{-\frac{n(n+1)}{2}\mu} \end{aligned} \quad (3.3.1.7)$$

The second moment of IT is given by

$$E[IT^2] = \sum_{n=1}^{\infty} n^2 p_n$$
$$= \sum_{n=1}^{\infty} n^2 \{e^{-\frac{(n-1)n}{2}\mu} - e^{-\frac{n(n+1)}{2}\mu}\} \qquad (3.3.1.8)$$
$$= \sum_{n=0}^{\infty} (2n+1) e^{-\frac{n(n+1)}{2}\mu}$$

3.3.3 Estimation of the Parameter of p_n

Equation 3.3.1.2 represents a one-parameter family of probability distributions and for estimation of the parameter, either the method of moments or the method of maximum likelihood can be used.

3.3.3.1 The Method of Moments

Let $t_1, t_2, \ldots, t_m$ be a random sample of size n drawn from a population of incubation times of HIV infected individuals. Then the sample mean is given by

$$\bar{t} = \frac{1}{n} \sum_{n=1}^{m} t_n$$

Replacing E[T] by $\bar{t}$ in 3.3.1.7, we have

$$\bar{t} = \sum_{n=0}^{\infty} e^{-\frac{n(n+1)}{2}\mu} \qquad (3.3.3.1.1)$$

As the incubation time of an HIV-infected individual can never be greater than 100 years, equation 3.3.3.1.1 can be truncated in the following manner:

$$\bar{t} = \sum_{n=0}^{100} e^{-\frac{n(n+1)}{2}\mu} \qquad (3.3.3.1.2)$$

An approximate value $\bar{\mu}$ of μ can be obtained from equation 3.3.3.1.1 by using the Newton-Raphson algorithm.

3.3.3.2 The Method of Maximum Likelihood

The likelihood function L(μ) for a sample $\{n_1, n_2, \ldots, n_m\}$ of size m is given by

$$L(\mu) = \prod_{j=1}^{m} e^{-\frac{(n_j - 1)n_j}{2}} (1 - e^{-\mu n_j})$$

The logarithm of L(μ) is given by

$$\log_e L(\mu) = -\frac{\mu}{2} \sum_{i=1}^{m} (n_i - 1)n_i + \sum_{j=1}^{m} \log(1 - e^{-\mu n_j})$$

When $\log_e$ L(μ) reaches its maximum value, the value of μ satisfies the following equation:

$$\frac{\partial L}{\partial \mu} = 0 \qquad (3.3.3.2.1)$$

From equation 3.3.3.2.1, we obtain

$$\sum_{i=1}^{m} n_i \frac{e^{-n_i \mu}}{1 - e^{-n_i \mu}} = \frac{1}{2} \sum_{i=1}^{m} m(n_i^2 - n_i) \qquad (3.3.3.2.2)$$

By applying the Newton-Raphson algorithm to equation 3.3.3.2.2, an approximate value $\bar{\mu}$ for μ can be obtained.

3.3.3.3 The Method of Median

The value of μ can be estimated from equation 3.3.1.5. for a sample of incubation times, we obtain the sample median θ* and then replacing θ in equation 3.3.1.5 by θ*, we have the following equation for a crude estimation μ* of μ:

$$\mu^* = \frac{2\log 2}{\theta^*(\theta^* + 1)} \qquad (3.3.3.3.1)$$

A numerical example to compare the three methods is provided in the next section.

3.3.4 A Numerical Example

The data of 84 homosexuals and bisexual men analysed in Lui et al. (1988) is used to obtain the incubation periods of twenty one individuals who developed AIDS prior to the year 1988 (Table 3.1). Estimates for the value of μ by the three methods are obtained and corresponding expected values and standard deviations are determined. The estimates are then used to test the goodness of fit of the distribution obtained.

Table 3.1 HIV Incidence data of 84 homosexuals

Year of HIV Infection	Year of diagnosis									
	1979	1980	1981	1982	1983	1984	1985	1986	Censored	Total
1978	0	0	0	1	0	1	1	0	3	6
1979		0	0	0	0	0	0	1	7	8
1980			0	0	0	1	1	1	9	12
1981				0	2	2	1	5	19	29
1982					1	0	3	0	19	23
1983						0	0	0	2	2
1984							0	0	4	4

From this table, the following incubation times (in years) of 21 persons were obtained as:

4, 6, 7, 7,4, 5, 6, 2, 2, 3, 3, 4, 5, 5, 5, 5, 5, 1, 3, 3, 3.

The sample mean is 4.19 years and the sample median is 4 years. By using Newton-Raphson algorithm in equation 3.3.3.1.2, with table 3.2, we have the optimal value $\hat{\mu} = 0.09$ so that the expected value of IT is 4.19 years with a standard deviation of 2.15 years. On the other hand,

for the same data of 21 persons, by adopting Newton-Raphson algorithm in equation 3.3.3.2.2, we get $\tilde{\mu}=1.01$ so that the expected value of IT is 1.41 years with a standard deviation of 0.59 year. Also, using equation 3.3.3.3.1, we get μ* = 0.07 so that the expected value of IT is 4.80 years with a standard deviation of 2.48 years. The three values of the parameter μ are listed in table 3.2.

Table 3.2 Values of the Parameters of μ

Method	μ	Mean	Standard Deviation
Moments	$\hat{\mu}=0.09$	4.19	2.15
Maximum Likelihood	$\tilde{\mu}=1.01$	1.41	0.59
Median	μ* = 0.07	4.80	2.48

Further, by applying χ^2 test, it was observed that the value of μ obtained by the method of moments fits closely to the observed data. Hence in what follows, we assume μ = 0.09 and proceed to project AIDS incidence by the Back-Calculation Method with a sample data (Bacchetti 1990)

3.4 THE BACK-CALCULATION AND THE INFECTION RATE

One of the methods used in estimating and projecting the infection rate from AIDS incidence data is the back-calculation method (Brookmeyer and Gail 1994). It is an important method of constructing rates of HIV infection and estimating current prevalence of HIV infection and future incidence of AIDS (Bacchetti et al. 1993). This method has been used by many mathematical scientists to obtain and predict the AIDS incidence of different populations. Amongst the work done are those of Verdecchia and Mariotto (1995) who modelled past HIV infections in Italy considering the interaction between age and calendar time. Anbupalam et

al. (2002) also used the Back calculation method to project future AIDS cases in Tamil Nadu by assuming that the incubation distribution was Weibul and Log-logistic. Ong and Soo (2006) estimated the HIV infection rates and projection in Malaysia while Lopman and Gregson (2008) used the Back-calculation method to reconstruct the historical trends in HIV incidence in Harare, Zimbabwe by using mortality data. They also attempted to determine the amount of peakness of HIV incidence and when the peakness occurred in Harare, Zimbabwe.

The method in continuous time is based on the convolution equation

$$A(t) = \int_0^t g(s)F(t-s)ds \qquad (3.4.1)$$

where A(t) represents the expected cumulative number of AIDS cases diagnosed by calendar time t, g(s) is the infection-rate at calendar time s and F(t) is the distribution of the incubation period. Equation 3.4.1 is a Volterra integral equation for g(s) and has been obtained by noting that an individual can be diagnosed to have AIDS before calendar time t, provided he/she has been infected at some time s < t and has an incubation period less than t-s. For a given AIDS incidence data, A(t) can be fitted and a model used for F(t) in 3.4.1 so that the rate g(s) can be computed by de-convolving equation 3.4.1. Taking Laplace transform on both sides of 3.4.1, we have

$$A^*(u) = \frac{g^*(u)f^*(u)}{u}$$

so that

$$g^*(u) = \frac{uA^*(u)}{f^*(u)} \qquad (3.4.2)$$

By inverting 3.4.2, we obtain the infection rate g(s).

On the other hand, the back-calculation in discrete time is based on the equation

$$E(Y_j) = \sum_{i=1}^{j} g_i \, p_{j-i+1} \qquad (3.4.3)$$

where Yj is the number of AIDS cases diagnosed in the j-th year [j-1,j], g_j is the number infected in the beginning of the j-th year and p_j is the probability that a person who is infected at the beginning of the 1st year is diagnosed with AIDS in the j-th year. If A_n denotes the expected cumulative number of AIDS cases diagnosed up to the end of the n-th year, then using equation 3.4.3, we have

$$A_n = \sum_{j=1}^{n} E(Y_j) = \sum_{j=1}^{n} \sum_{i=1}^{j} g_i \, p_{j-i+1} \qquad (3.4.4)$$

Equation 3.4.4 is analogous to equation 3.4.1.

We proceed to illustrate the back-calculation in discrete time with the data used in Bacchetti (1990) where the monthly infection rate and monthly AIDS incidence among gay men in San Fransisco in the cohort born from October 1929 through September 1959 were estimated. Taking t = 0 to correspond to January 1978 and the time unit as year, the data is given in table 3.3 below.

Table 3.3 Data on AIDS incidence among gay men in San Fransisco

j	1	2	3	4	5	6	7	8	9	10	11
Y_j	0	0	1	26	93	278	560	840	1264	1464	1455

Table 3.4 Probability distribution of the Incubation Time

n	1	2	3	4	5	6	7	8	9	10
p_n	0.09	0.15	0.18	0.18	0.15	0.11	0.07	0.04	0.02	0.01

For μ = 0.09, the probability distribution of the incubation time is given in table 3.4. Following Brookmeyer and Gail (1994), we proceed to obtain the discrete time infection curve. We assume for simplicity that infections occurring in a calendar year are accounted at a single time point, for example, January 1 of the year and

$$g(2n-1) = g(2n) = \beta_n, \; n = 1, 2, \ldots \tag{3.4.5}$$

Equation 3.4.5 provides a simple smoothness assumption on the annual infection rate. Consequently, equation 3.4.3 leads to the following matrix equation:

$$\begin{pmatrix} E(Y_1) \\ E(Y_2) \\ E(Y_3) \\ E(Y_4) \\ E(Y_5) \\ E(Y_6) \\ E(Y_7) \\ E(Y_8) \\ E(Y_9) \\ E(Y_{10}) \\ E(Y_{11}) \end{pmatrix} = \begin{pmatrix} 0.09 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.24 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.34 & 0.09 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.36 & 0.24 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.32 & 0.34 & 0.09 & 0.00 & 0.00 & 0.00 \\ 0.25 & 0.36 & 0.24 & 0.00 & 0.00 & 0.00 \\ 0.18 & 0.32 & 0.34 & 0.09 & 0.00 & 0.00 \\ 0.11 & 0.25 & 0.36 & 0.24 & 0.00 & 0.00 \\ 0.06 & 0.18 & 0.32 & 0.34 & 0.09 & 0.00 \\ 0.03 & 0.11 & 0.25 & 0.36 & 0.24 & 0.00 \\ 0.01 & 0.06 & 0.18 & 0.32 & 0.34 & 0.09 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{pmatrix} \tag{3.4.6}$$

Using the Poisson Regression Analysis (PRA) (Koch et al. 1986, McCillagh and Nelder 1989), the values of β_j for j = 1, 2, …, 6 are estimated. The method is based on the assumption that the random variable Y_j has a Poisson distribution. Setting $\mu_j = E(Y_j)$, the likelihood function corresponding to the sample $\{n_1, n_2, \ldots, n_{11}\}$ of $\{ Y_1, Y_2, \ldots, Y_{11}\}$ is given by

$$\varphi(\mu_1, \mu_2, \ldots \mu_{11}) = \prod_{i=1}^{11} e^{\mu_i} \frac{\mu_i^{N_i}}{n_i!} \tag{3.4.7}$$

But from equation 3.4.6, we have

$$\mu_1 = 0.09\beta_1$$
$$\mu_2 = 0.24\beta_1$$
$$\mu_3 = 0.34\beta_1 + 0.09\beta_2$$
$$\mu_4 = 0.36\beta_1 + 0.24\beta_2$$
$$\mu_5 = 0.32\beta_1 + 0.34\beta_2 + 0.09\beta_3$$
$$\mu_6 = 0.25\beta_1 + 0.36\beta_2 + 0.24\beta_3$$
$$\mu_7 = 0.18\beta_1 + 0.32\beta_2 + 0.34\beta_3 + 0.09\beta_4$$
$$\mu_8 = 0.11\beta_1 + 0.25\beta_2 + 0.36\beta_3 + 0.24\beta_4$$
$$\mu_9 = 0.06\beta_1 + 0.18\beta_2 + 0.32\beta_3 + 0.34\beta_4 + 0.09\beta_5$$
$$\mu_{10} = 0.03\beta_1 + 0.11\beta_2 + 0.25\beta_3 + 0.36\beta_4 + 0.24\beta_5$$
$$\mu_{11} = 0.01\beta_1 + 0.06\beta_2 + 0.18\beta_3 + 0.32\beta_4 + 0.34\beta_5 + 0.09\beta_6$$

and hence on substitution of these equations in 3.4.7, φ becomes a function of $\beta_1, \beta_2, \ldots, \beta_6$. Differentiating $\log_e \varphi$ with respect to β_j and equating the results to 0, the following system of equations is obtained:

$$\sum_{i=1}^{11}\left(\frac{\mu_i - n_i}{\mu_i}\right)\frac{\partial \mu_i}{\partial \beta_j} = 0,\ j = 1, 2, 3, 4, 5, 6. \qquad (3.4.8)$$

Equations 3.4.8 do not yield an explicit solution and so an iterative method is used to obtain an approximate solution for $(\beta_1, \beta_2, \ldots, \beta_6)$ as given below:

$$\hat{\beta}_1 = 6, \hat{\beta}_2 = 33, \hat{\beta}_3 = 1041, \hat{\beta}_4 = 2583, \hat{\beta}_5 = 3416, \hat{\beta}_6 = 5172.$$

The above values can be used to forecast AIDS incidence on short term. For example, the predicted AIDS incidence in the 12^{th} year is obtained as 6523 by using the following extended equation

$$\hat{Y}_{12} = \hat{\beta}_1(p_{11} + p_{10}) + \hat{\beta}_2(p_9 + p_8) + \ldots + \hat{\beta}_6(p_3 + p_2)$$

3.5 CONCLUSION

In this chapter a two-parameter family distribution function for the gay-life and a one-parameter family of distribution for the incubation period have been modelled. For the model, it was observed that the distribution function of the incubation period using the method of moments serves as a good fit for the data provided by Lui et al. (1988). The only setback of the Back-calculation method in projecting AIDS incidence is the inability to project for a very long time.