

A STUDY OF STATISTICAL METHODS
IN AIDS MODELLING

by

ONG HONG CHOON

Thesis submitted in fulfilment of the
requirements for the degree
of Master of Science

January 1997

MIKROFIS

4594

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Low Heng Chin and co-supervisor, Professor Quah Soon Hoe, for their invaluable help, guidance, understanding, advice, concern, support and encouragement. I am grateful to Assistant Professor Dr. Peter Bacchetti, of the Department of Epidemiology and Biostatistics, University of California, San Francisco, for allowing me to use his backcalculation program and the Deputy Director, Disease Control Division, Ministry of Health Malaysia, for the permission to use the Malaysian data.

I am also grateful to Professor Hassan Said, our Deputy Vice-Chancellor for Academic Affairs (the former Dean of the School of Mathematical Sciences), Universiti Sains Malaysia, for allowing me to convert from a coursework to a research programme and to the current Dean of the School of Mathematical Sciences, Associate Professor Dr. Rosihan M. Ali, for allowing me to use the facilities in the school. I would like to thank my friends Wong Ya Ping (for helping me in the computer programming), Lawrence H.T. Chang, Tan Ewe Hoe, Catherine Lee and many others from Universiti Sains Malaysia, Penang, for the joy of their fellowship and mutual help and support.

I would like to dedicate this thesis to my mother and other relatives for their love and support. Above all, I thank God, for His mercy, grace and the opportunity to further my studies.

TABLE OF CONTENTS

	<u>Page</u>
Acknowledgements	ii
Table of Contents	iii
Abstrak	v
Abstract	vii
 CHAPTER 1. INTRODUCTION	 1
1.1 A Brief History and the Epidemiology of AIDS	1
1.2 AIDS in Malaysia and the World	5
1.3 Main Problem Areas	14
1.4 Format of Presentation	17
 CHAPTER 2. NATURAL HISTORY OF AIDS	 20
2.1 An Overview of the Natural History of AIDS	20
2.2 Survey of HIV Seroprevalence and Seroincidence	21
2.2.1 <i>Non-Response Bias</i>	21
2.2.2 <i>Biases in Seroprevalent Data</i>	21
2.3 Mathematical Models for the Incubation Period	24
2.3.1 <i>Introduction</i>	24
2.3.2 <i>Non-Parametric Models</i>	25
2.3.3 <i>Parametric Models</i>	26
2.4 Analysis of Interval Censored and Doubly Censored Data	29
2.4.1 <i>Introduction</i>	29
2.4.2 <i>Estimation of Seroconversion Time and Its Effects</i>	29
2.4.3 <i>Analysis of Missing Data</i>	32
2.4.4 <i>Non-Parametric Estimation</i>	35
 CHAPTER 3. AIDS INCIDENCE DATA	 39
3.1 Introduction	39
3.2 A Parametric Approach for Estimating Reporting Delays	40
3.3 An Approach Based on Poisson Regression Methods	43
3.4 Non-Parametric Method of Estimation	47
3.5 Adjusting for Changes in the Surveillance AIDS Definition	50
3.6 Other Issues	52

CHAPTER 4. MECHANISTIC MODELS	54
4.1 Introduction	54
4.2 Deterministic Models	55
4.3 Stochastic Models	65
CHAPTER 5. THE METHOD OF BACKCALCULATION	69
5.1 Introduction	69
5.2 Background of the Method of Backcalculation	70
5.3 Deconvolution Techniques	73
5.4 Statistical Models of the HIV Infection Curve	76
5.4.1 <i>Introduction</i>	76
5.4.2 <i>Strong Parametric Models</i>	76
5.4.3 <i>Weakly Parametric Models</i>	78
5.5 Extension to Account for a Non-stationary Incubation Distribution	80
5.6 Limitations and Usefulness of Backcalculation	82
5.7 An Application of the Backcalculation Method to the Malaysian Data	84
CHAPTER 6. MARKER MODELS	88
6.1 Introduction	88
6.2 Marker Trajectory	90
6.3 Statistical Models of Markers	94
6.4 Other Uses of Markers	99
CHAPTER 7. GENERAL DISCUSSION	101
7.1 Summary	101
7.2 Type of Method for Assessment	101
REFERENCES	104
GLOSSARY	112
APPENDICES	
Appendix A Definitive Diagnostic Methods for Diseases Indicative of AIDS (1988)	115
Appendix B Main Backcalculation Program (Without Monthly Effects)	117

SUATU KAJIAN KAEDAH STATISTIK DALAM PEMODELAN AIDS

ABSTRAK

Tujuan penyelidikan ini adalah untuk melakukan suatu kajian kritikal tentang kaedah-kaedah statistik yang dicipta berhubung dengan pemodelan wabak HIV/AIDS dan untuk mengkaji perkembangan pelbagai kaedah yang digunakan dari tahun 1985 hingga tahun 1995 supaya dapat memahami wabak tersebut dengan lebih baik. Suatu kajian pelbagai isu statistik berkaitan dengan pengumpulan dan sifat-sifat data HIV/AIDS telah dibuat. Data-data tersebut diubahsuai sebelum digunakan di dalam sesuatu model. Isu-isu statistik tersebut merangkumi berat sebelah dalam prevalen kohort, penapisan data, kelambatan melapor, perubahan dalam definisi pengawasan AIDS dan banyak lagi.

Tiga kaedah dalam pemodelan AIDS dibincangkan. Pada satu ekstrem, kaedah pertama cuba menyuaikan suatu fungsi seperti fungsi polinomial atau lengkungan mudah lain kepada fungsi lengkung, bilangan kes AIDS terhadap masa kalendar. Pada ekstrem yang lain, kaedah kedua cuba memodelkan mekanisma penjangkitan wabak tersebut di dalam populasi secara menyeluruh sambil memberi pandangan mendalam kepada evolusi wabak dan mengenalpasti pembolehubah yang menentukan bilangan kes pada masa hadapan.

Kaedah ketiga, yang merupakan perantaraan kaedah pertama dan kaedah kedua ialah kaedah pengiraan balik. Kaedah pengiraan balik ialah suatu kaedah bagi menganggarkan kadar bilangan yang menghidapi HIV yang lalu dari data bilangan insidens AIDS dan suatu anggaran taburan tempoh pengeraman. Kaedah pengiraan balik tersebut diterapkan pada data insidens AIDS di Malaysia (yang agak tepat) dengan menggunakan program pengiraan balik daripada

Bacchetti (1993) dan taburan tempoh pengeraman daripada Brookmeyer (1991) supaya menghasilkan kadaran jangkitan HIV terkini di Malaysia (sehingga Ogos 1996).

Dalam pengukuran perkembangan jangkitan HIV, penanda seperti bilangan sel CD4 T, telah digunakan untuk mengukur perkembangan jangkitan HIV. Kami membincangkan penggunaan penanda tersebut, terutamanya bilangan sel CD4 T, dalam pemodelan perkembangan jangkitan HIV bagi menganggarkan masa baki untuk menghadapi AIDS bagi kes yang menjangkiti HIV.

Sebagai mengakhiri tesis ini, kami memberi suatu ringkasan bagaimana pelbagai kaedah digunakan dalam keadaan yang berlainan.

ABSTRACT

The objective of this research is to do a critical survey on some of the statistical methods that have been developed on the modelling of the HIV/AIDS epidemic and to study the development of various methods which have been used from the year 1985 until the year 1995 to better understand the AIDS epidemic. A study of how the various statistical issues related to the collection and the nature of the HIV/AIDS data is made. These data are adjusted before they are incorporated into the model. The statistical issues include biases in cohort prevalence, censoring of the data, delay in reporting, changes in AIDS surveillance definition and many others.

Three methods of modelling the AIDS epidemic are discussed. At one extreme, the first method attempts to fit a function of calendar time such as a polynomial or other mathematically convenient curves to the AIDS incidence curve. At the other extreme, the second method attempts to model the full dynamics of the transmission of the epidemic in the population, providing much insight into the qualitative evolution of the epidemic and identifying the key variables that determine the future number of cases.

The third method, which is intermediate between the first two methods, is the method of backcalculation. Backcalculation is a method of estimating past HIV infection rates from the AIDS incidence data and an estimate of the incubation period distribution. The method of backcalculation is applied to the Malaysian AIDS incidence data (which is fairly reliable) using the backcalculation program from Bacchetti (1993) and the incubation period distribution from Brookmeyer (1991) to generate the current HIV infection rate for Malaysia (until August 1996).

In measuring the progression of infection with HIV, markers like the CD4 T cell counts are used to track or measure the progression of the HIV infection. We discuss the uses of markers, especially the CD4 T cell counts, in modelling the progression of the HIV infection to estimate the residual time to AIDS.

To round up the thesis, we give a brief summary of how the various methods are applied to different situations.

CHAPTER 1

INTRODUCTION

1.1 A Brief History and the Epidemiology of AIDS*

The acquired immunodeficiency syndrome (AIDS) emerged as a major and burgeoning global health problem in the 1980s. The first case of AIDS was characterised in June 1981 in five unrelated homosexual men in Los Angeles, California, and early studies suggested that the disease is spread via sexual activity. In 1982, AIDS was observed in transfusion recipients, haemophiliacs and intravenous (IV) drug users and thus provided further evidence of a transmission agent. It was not until 1983 that the etiologic agent, the human immunodeficiency virus (HIV) was isolated and the mode of transmission more firmly established.

The discovery of HIV is a subject of international debate, a legal battle and at one time strained relations between France and the USA. The discovery of the AIDS virus was claimed by Dr. Robert Gallo of the National Cancer Institute of the United States who named it human T-lymphotropic virus III (HTLV III) and by Luc Montagnier of the Pasteur Institute, Paris, France who named it lymphadenopathy-associated virus (LAV). The matter was settled out of court but it is now accepted that Dr. Montagnier discovered in 1983 what has come to be called human immunodeficiency virus (HIV). Varmus *et al.* (1986) in a letter empowered by the International Committee on the Taxonomy of Virus proposed that the AIDS retroviruses be officially designated as the human immunodeficiency virus to be known in abbreviated form as HIV.

* Refer to Appendix A for a brief account of the definitive diagnostic method for diseases indicative of AIDS.

Fauci (1988) noted that infection with the HIV results in a profound immunosuppression due predominantly to a selective depletion of helper/indoor T lymphocytes that express the receptor for the virus (the CD4 molecule). HIV induced immunosuppression results in a host defense defect that renders the body highly susceptible to opportunistic infections and neoplasm. The immune defect appears to be progressive and irreversible with high mortality rate that may well approach 100 percent over several years.

However, the interplay between the HIV and the immune system appears more dynamic than most scientists would have suspected. Nowak and McMichael (1995) gave a plausible hypothesis suggesting that the immune devastation that underlies AIDS stem from continuous and dangerous evolution of the HIV in the body.

AIDS is a highly contagious infection which is transmitted by having unprotected sex with an infected person, by transfer of infected blood and by an infected mother to her unborn child. Theoretically, every HIV infected individual is capable of transmitting the virus which has been isolated from blood, semen, cervical secretions, urine, cerebrospinal fluid, saliva, tears and breast milk. The highest concentration of the virus occur in the male semen which is apparently the most important body fluid for transmitting HIV (Henderson and Saah, 1993). The virus can spread through blood, sperm and vaginal fluid. Saliva and tears has been found to have extremely low concentrations of HIV and no substantiated cases of infection have been traced to either of these body fluids. Although HIV has been found in saliva, the amount present is relatively less. Scientists are not entirely sure whether HIV can be transmitted by wet kissing but deep tongue kissing may possibly carry a risk of infection because the virus contained in the saliva of an infected person may be able to enter the bloodstream of the other person through cuts

or sores in the lining of the mouth or gums. However, there has not yet been any confirmed cases of HIV transmission through kissing.

There is a **window** or **latent** period from the time of infection until the individual becomes infectious. During this period, the antibodies for the virus are not detectable. This is followed by an **infectious period** during which the individual is termed an infective and can pass on the disease to susceptibles. With current technology, the window period, which is normally a few months, is in most cases now only 2 to 4 weeks (Maskill and Gust, 1992). The **incubation period** is the period from infection until appearance of overt symptoms. When the HIV antibodies are detectable in the blood a few weeks after infection, the individual is then seropositive. For the purpose of modelling the epidemic, it is assumed that all those infected with the virus are seropositive, that is the latent period is negligible in comparison to the incubation period (Isham, 1988 and Jewell, 1990).

The incubation period may be shorter for the case of those who received infected blood through transfusion compared to those infected by sexual contact. This may be due to the volume of infected blood contact involved. The incubation period may vary from a few years to over 10 years. No one knows for sure what percentage of people infected with HIV will go on to develop AIDS. Lui, Darrow and Rutherford (1988a) indicated the difficulty in identifying the date of exposure of HIV-1 infection in person other than transfusion recipient (as the date of transfusion can be determined) resulting in a limitation in studying the incubation period for AIDS.

Vaccination provides one of the easiest and simplest ways of fighting a disease by preventing it from occurring as in the case of smallpox. However, HIV presents

virologists with a daunting problem in that it can hide in cells and install its own genes into the genes of its host. It can mutate and attack the very cells which are responsible for fighting infection. Furthermore, clinical trials will probably raise ethical problems and volunteers could be in short supply as thousands of people are often needed to test a new vaccine (Koff and Hoth, 1988). Most vaccines are administered before an infection occurs but are not of any use to people who are already infected. The prospects for a cure at the moment look grim.

People who are HIV+ or have the AIDS condition are generally treated through a combination of health counselling, emotional support and the use of antiviral drugs. Health counselling is to help them adopt more healthy behaviour to help strengthen the immune system and decrease the chances of opportunistic infections. Emotional support is to help them to handle emotional distresses and to maintain a positive mental imagery. Drugs like AZT (zidovudine) reduce the number of opportunistic infections by interfering with the virus's ability to replicate itself in the body. However, AZT has side effects like the suppression of bone marrow function which leads to anaemia and lower white blood counts. DDI (dideoxyinosine) increases the level of T4 cells. This strengthens the immune system and offers hope to those for whom AZT is not effective. Lacking a vaccine or cure for AIDS, the best approach for curbing its spread is prevention.

Today, there are tens of AIDS drugs at various stages of development. Aldhous (1995) gave an upbeat report about drugs that delay the onset of AIDS. There are several dozen new anti-HIV drugs being developed. This expanding range of drugs bring with it the possibility of hitting on the right combination of compounds to counter HIV's tendency to evolve resistance to a single drug given in isolation. In a *Time* magazine article, Gorman (February 12, 1996) reported on a promising new class of drugs called protease

inhibitors which block the production of a key enzyme, protease, that the virus needs to replicate itself. There is still no cure but scientists and survivors are making striking progress. However, even if protease inhibitors live up to their potential, it is estimated that the drugs will cost US\$500 to US\$600 a month probably for the rest of a patient's life in addition to the standard treatment with AZT and its cousins, which cost approximately US\$400 a month. Besides that, hospitalization and other forms of medical care in the final stages of the disease can add a further US\$150,000.

1.2 AIDS in Malaysia and the World

Since the beginning of the HIV/AIDS pandemic in the 1980s, 192 countries have reported to the World Health Organisation (WHO) Global Programme on AIDS. As of the middle of 1995, 18.5 million adults and more than 1.5 million children are estimated to have been infected by HIV/AIDS (Yelu, 1995). The ravages of the disease have demanded stronger anti-AIDS campaigns to rise up to the challenge with preventive programmes and activities to control the spread of the HIV/AIDS.

Many people think that AIDS is a disease of homosexuals but WHO has in 1987 identified three distinct worldwide patterns of transmission and infection according to (a) the time at which the HIV epidemic began and (b) predominant modes of HIV transmission. In pattern I countries, the main group affected are homosexuals and intravenous (IV) drug users where the extensive spread of HIV in these groups began in the late 1970s and is slowly spreading into the heterosexual population. This pattern is currently seen in North America, Western Europe, Australia and New Zealand. In pattern II countries, HIV/AIDS are found mainly in sexually active heterosexuals. The countries most affected in this pattern are in Sub-Saharan Africa and some parts of the Caribbean where HIV began to enter the population extensively in the mid 1970s.

Pattern III countries includes Eastern Europe, Asia, most Pacific countries (excluding Australia and New Zealand), North Africa and the Middle East where until recently no predominant mode of HIV transmission was discernible. However, this pattern is now changing rapidly. Central and South America originally belonged to Pattern I but heterosexual transmission has increased so rapidly that the area has been reclassified as a separate pattern - Pattern I/II. The above changes illustrates that HIV/AIDS is so new that each passing month brings new experiences, challenges and facts which increase our questions and understanding concerning these infections. The map (Figure 1.1) below from Wing (1991) illustrates the patterns mentioned.

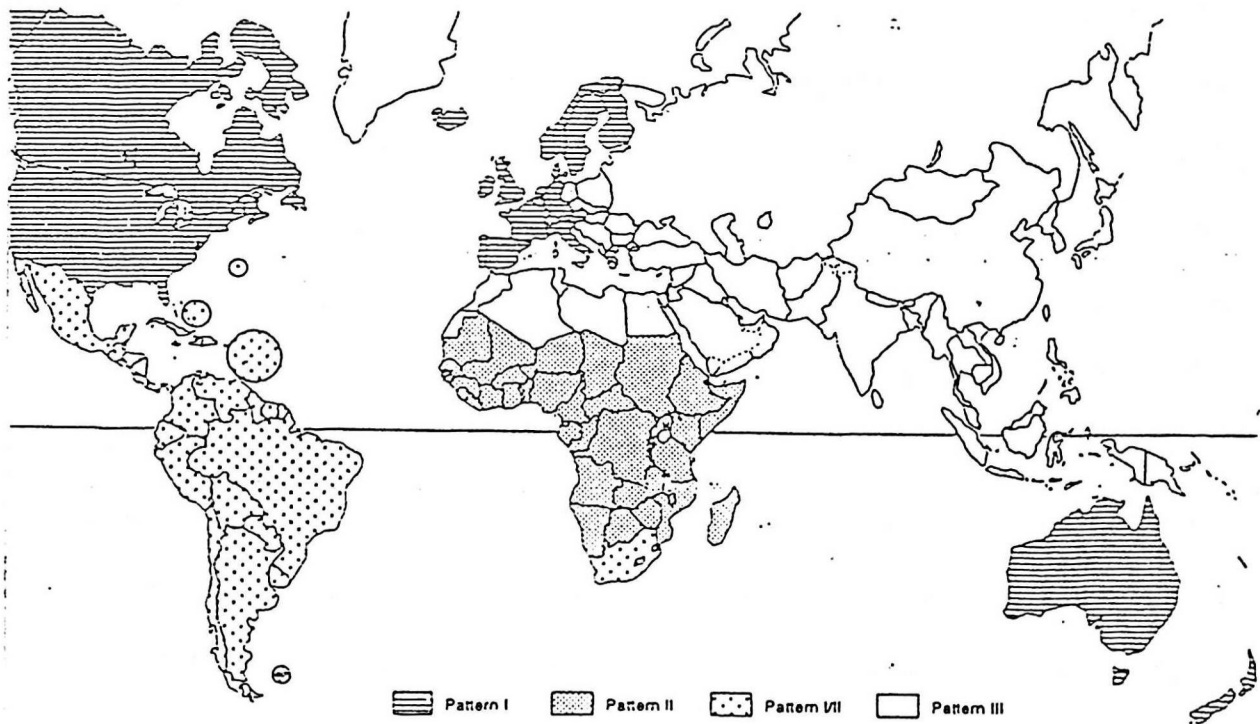


Figure 1.1 Worldwide pattern of HIV/AIDS transmission and infection

Three major global HIV/AIDS trends and patterns have been noted by Chin (1993). Firstly, the decreasing incidence of HIV infections among homosexual men in western countries has continued into the 1990s. Secondly, the vulnerability of IV drug users throughout the world well documented in the 1980s has remained a constant factor. However, the explosive outbreak of HIV among IV drug users in Bangkok, Thailand starting in early 1988 has now spread extensively to contiguous areas in South East Asia. Thirdly, the increase in HIV prevalence among heterosexual population throughout the world, first noted in the mid to late 1980s have continued relentlessly.

The onset of AIDS has resulted in the wasting away of the lives of tens of thousands of young people. AIDS is hitting homes as the disease claims workers in their prime. An estimated 46,000 Thais will die of AIDS in 1995, bringing to nearly 120,000 the number of deaths since the start of the epidemic (Fairclough, 1995). Already, in Chiang Mai, contractors say that they cannot find enough workers as a result of AIDS. A United Nations funded study by a Madras-based organisation revealed that industrial labour is ten times more at risk of contracting AIDS than the rest of the population (Salvi, 1995). The potential impact of AIDS on the Indian economy is just becoming apparent. Other Asian countries with less vigorous health campaigns could pay an even higher price. World AIDS day was observed all over Asia on the first of December, 1994 with a marked desire to educate and create awareness about the disease. These efforts to educate and prevent the epidemic from spreading could not be more timely with the number of HIV cases in Asian countries ripe for explosion once it spreads to the heterosexual population.

In Malaysia, the then Health Minister, Mr. Lee Kim Sai , opened an AIDS exhibition in Kuala Lumpur urging greater support for the victims and more open discussions on the

disease (Malaysian Doctor Annual Special, 1994). A World Health Organisation report recently put Malaysia at the top of its list for new HIV cases reported in the Western Pacific region with 2522 new cases in 1993. Also, in conjunction with the World AIDS day on the first of December 1994, the Malaysian AIDS Council comprising 21 non-governmental organisations (NGOs) under the presidency of Marina Mahathir has put forward aggressive plans for the future. The war against AIDS, however, could have been more successful if it was unhampered by the reluctance to promote the use of condoms in deference to religious sensitivities. It is evident that all over Asia, people are reluctant to have open discussions on sexual issues.

Malaysia, being a conservative Asian country, which was critical of the moral norms of its neighbours and the West, was debating the policies and strategies to use in the fight against AIDS (Huddart, 1992). The first case of AIDS was detected on the nineteenth of December 1986. This was an American Chinese of Malaysian origin who migrated to the United States 30 years ago. He came to Malaysia on a social visit and developed pneumocystis carinii pneumonia (PCP) shortly after arrival. The patient, however, recovered and left Malaysia voluntarily on the fifteenth of January 1987. Subsequently, it was reported by a local news paper that he died following complications after a bone marrow transplant. In the same year, three HIV carriers were also detected.

A proactive surveillance programme was launched in 1989 to screen the injection of drug users in government correctional centres and medical institutions resulting in an apparently sharp rise in HIV+ cases around that time and also among IV drug users. The number of HIV infected persons reported and the classification of HIV infected persons in Malaysia are shown in Table 1.1 and Table 1.2 as reported by Wong (1996) in *The Sun*, a local newspaper.

Table 1.1 Number of HIV infected persons reported in Malaysia (1985-1995)

Year of Reporting	HIV Infections			AIDS Cases			AIDS Deaths		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
1985	0	0	0	0	0	0	0	0	0
1986	4	0	4	1	0	1	1	0	1
1987	5	0	5	1	0	1	1	0	1
1988	22	1	23	3	0	3	2	0	2
1989	175	8	183	3	1	4	3	1	4
1990	648	14	662	11	0	11	11	0	11
1991	1623	63	1686	24	3	27	11	3	14
1992	2324	93	2417	30	0	30	21	0	21
1993	2400	116	2516	35	5	40	24	5	29
1994	3226	150	3376	65	7	72	52	6	58
1995	3357	189	3546	133	9	142	133	9	142
Cumulative Total	13784	634	14418	306	25	331	259	24	283
Percentage	95.6	4.4	100.0	92.4	7.6	100.0	91.5	8.5	100.0

(Source: Malaysian AIDS Foundation)

Table 1.2 Classification of HIV infected persons in Malaysia (1985-1995)

Factor	Classification	HIV Infections		AIDS Cases	
		Number	%	Number	%
Gender	Male	13784	95.60	306	92.45
	Female	634	4.40	25	7.55
	Total	14418	100.0	331	100.0
Age in Years	< 13	33	0.23	7	2.11
	13-19	647	4.49	2	0.60
	20-29	6965	48.31	94	28.40
	30-39	5250	36.41	157	47.43
	40-49	955	6.62	42	12.69
	>50	182	1.26	23	6.95
	Unknown	386	2.68	6	1.81
	Total	14418	100.0	331	100.0
Ethnicity	Malays	10092	70.00	136	41.09
	Chinese	2175	15.09	130	39.27
	Indians	1480	10.26	44	13.29
	Other Malaysians	115	0.80	4	1.21
	Foreigners	486	3.37	17	5.14
	Unknown	70	0.49	0	0.00
	Total	14418	100.0	331	100.0
Transmission Risk Factor	Homo/Bisexual	112	0.78	27	8.16
	IVDU	11088	76.90	161	48.64
	Prostitutes	240	1.66	11	3.32
	Heterosexual	379	2.63	76	22.96
	Transfusion Blood/Blood Product	32	0.22	9	2.72
	Mother to Infant	21	0.15	5	1.51
	Other/Unknown	2546	17.66	42	12.69
	Total	14418	100.0	331	100.0

The summary data in Table 1.1 can be represented graphically in the line graphs in Figures 1.2, 1.3 and 1.4 below. The steep rise in the number of HIV+ cases in the 3rd quarter of 1989 in Figure 1.2 may be due to the mandatory testing of all IVDUs admitted to drug rehabilitation centres, increase in detection through aggressive case finding and increase in awareness of the public which lead people at risk to willingly come forward to be tested (Ismail, Doi and Naganathna, 1995). The vast difference between the number of male and female cases in Figure 1.2 is primarily because the IVDUs are mainly male cases. This has resulted in a subsequently steep rise in the number of AIDS cases and deaths in the early 1990s as can be seen in Figures 1.3 and 1.4.

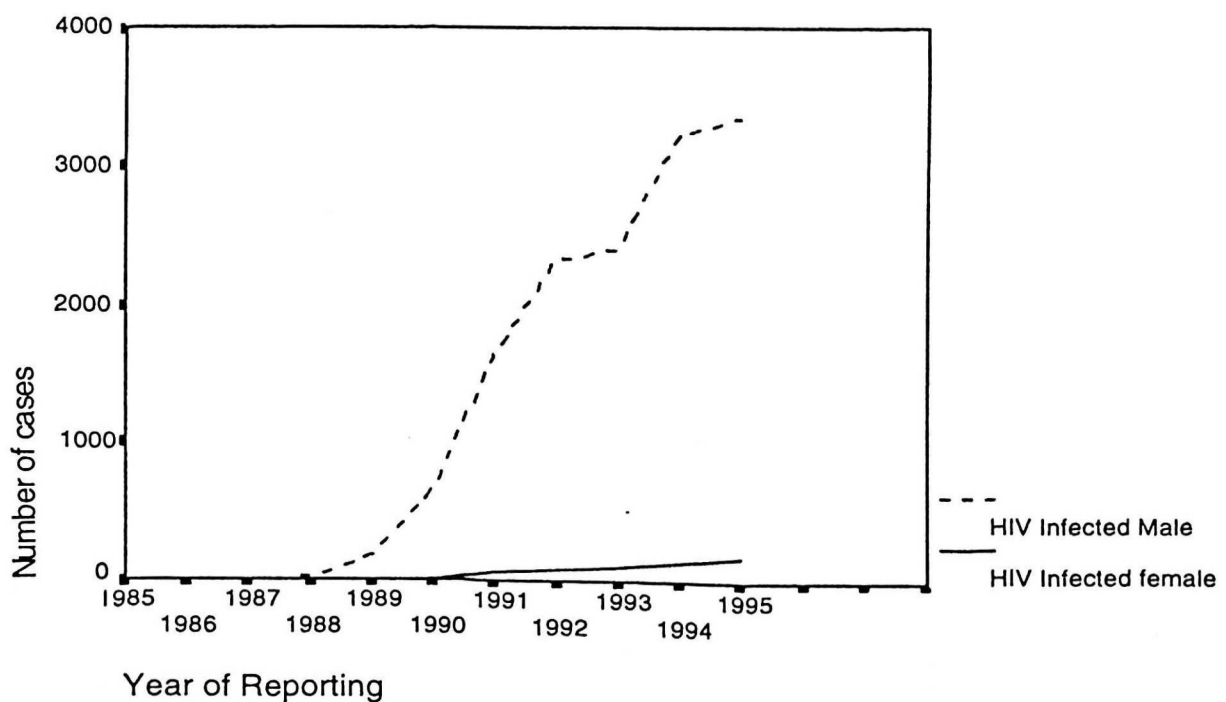


Figure 1.2 Number of HIV+ cases reported in Malaysia

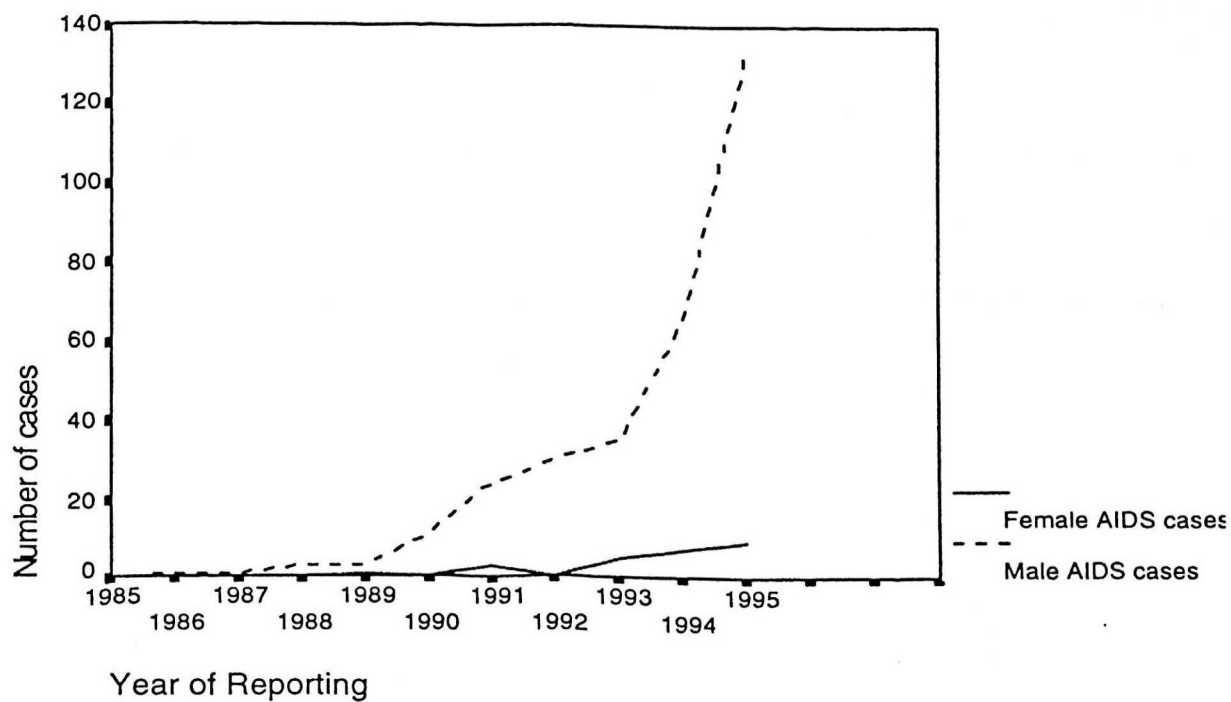


Figure 1.3 Number of AIDS cases reported in Malaysia

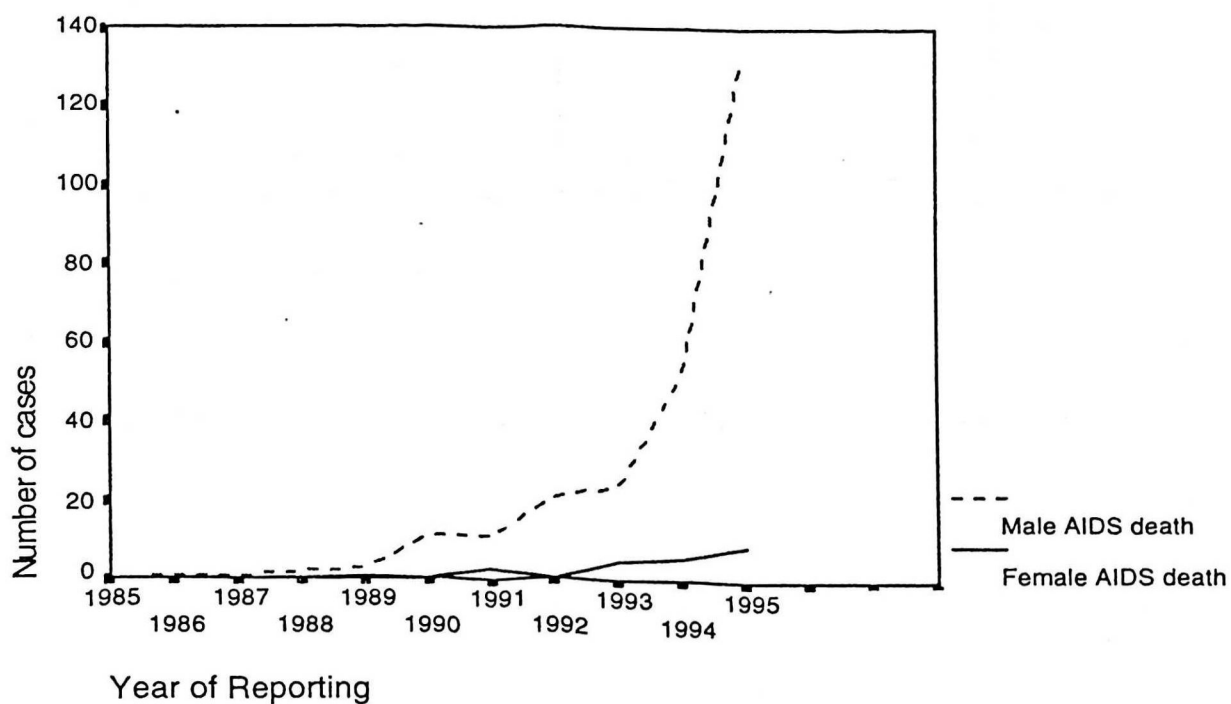


Figure 1.4 Number of deaths due to AIDS reported in Malaysia

The summary data in Table 1.2 can be represented in the bar charts in Figures 1.5, 1.6 and 1.7. It is obvious that the age group most affected are the young people who are in the most productive period of their lives. This is a cause of concern to the government. The large number of IVDUs who are HIV/AIDS cases is primarily due to the mandatory testing of all IVDUs admitted to drug rehabilitation centres and prisons. This has resulted in a bias in the number of cases involving IVDUs.

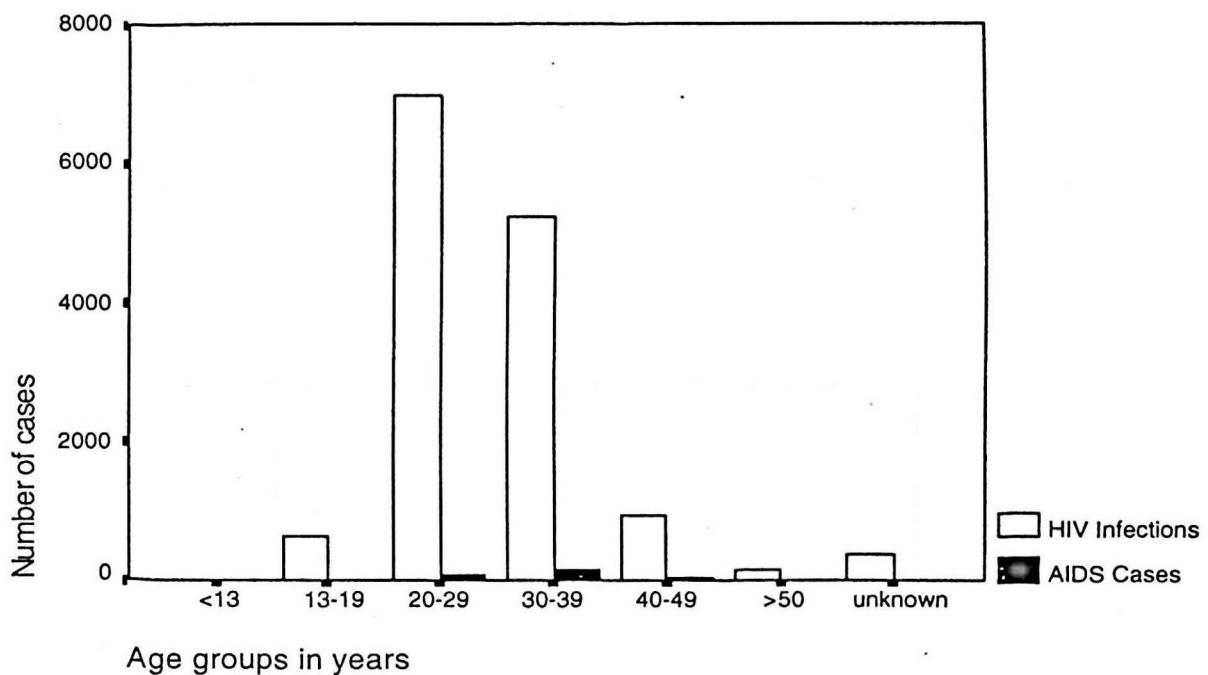


Figure 1.5 Number of HIV/AIDS cases according to age groups in Malaysia

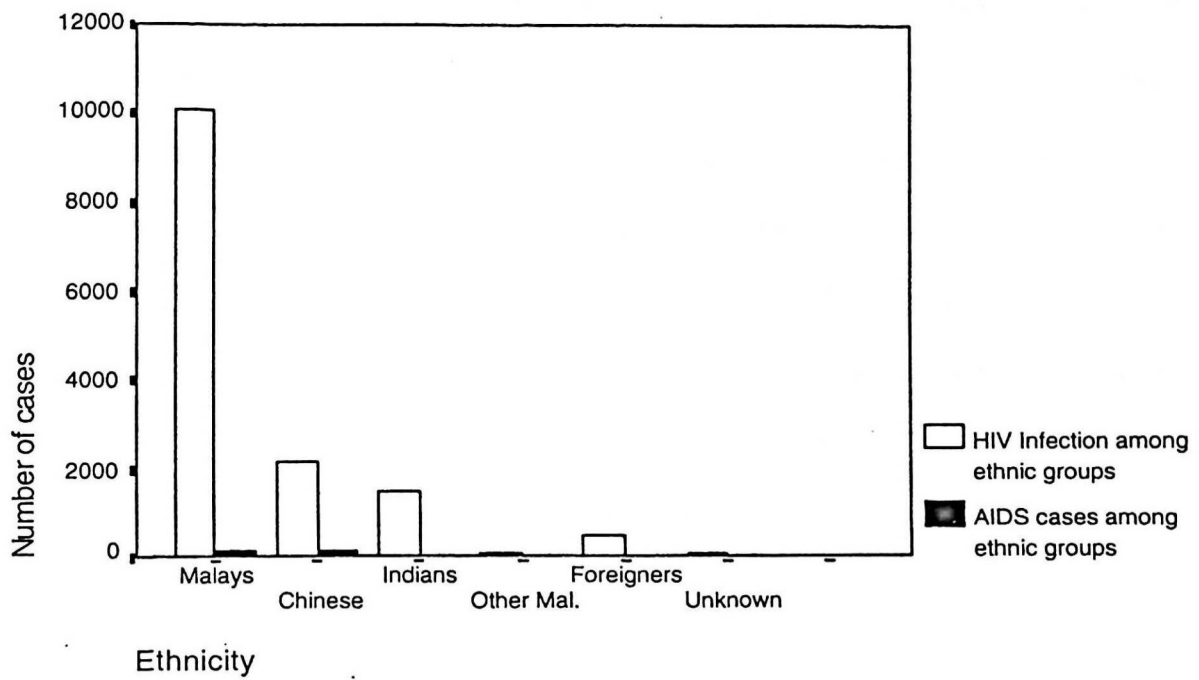


Figure 1.6 Number of HIV/AIDS cases according to ethnicity in Malaysia

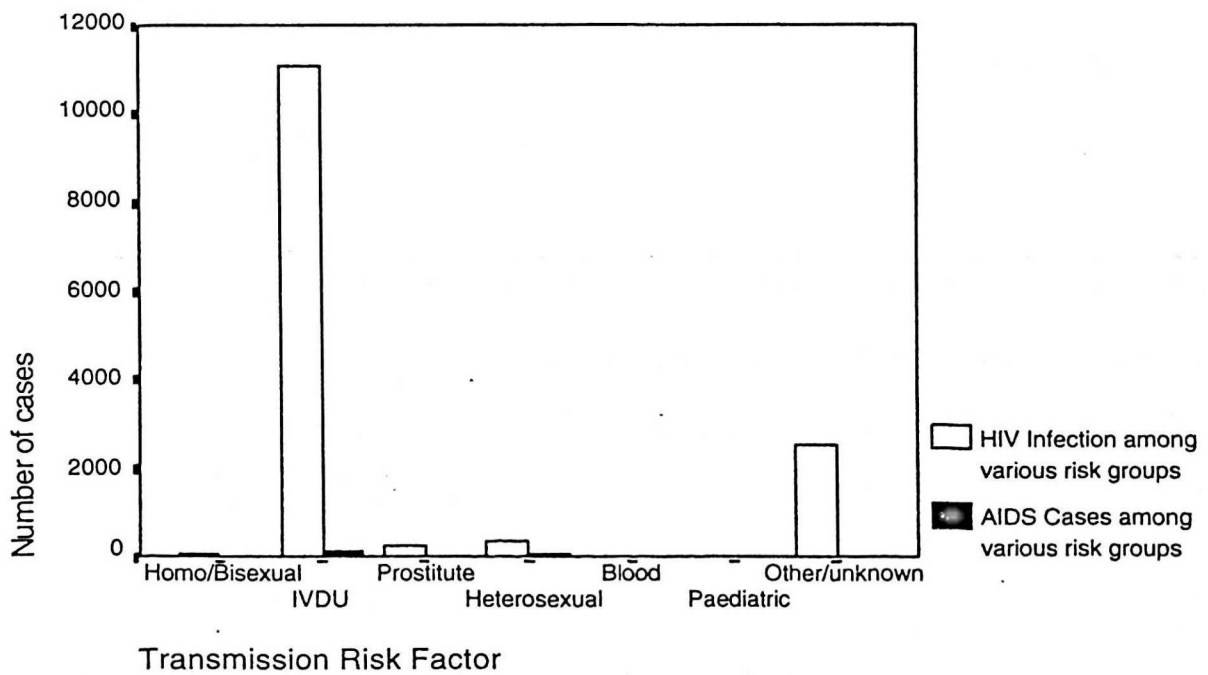


Figure 1.7 Number of HIV/AIDS cases according to transmission risk group in Malaysia

Following the commemoration of the World AIDS day in Malaysia, the Malaysian AIDS Charter was drawn up following a series of workshops. Government departments, the business community, non-governmental organisations, religious leaders and concerned individuals including people living with HIV/AIDS were involved in the drawing up of the Malaysian AIDS charter. Although it is not a legally binding document, it is hoped that the AIDS charter will be morally binding whereby standards are set by which we hope Malaysia will approach the HIV/AIDS epidemic (*Sunday Star*, 3 December 1995). The charter sets out a basic framework of shared rights and shared responsibilities incumbent upon all citizens in Malaysia.

1.3 Main Problem Areas

The discovery of AIDS and the isolation of HIV has generated an area of research that is apparently a rich source of statistical problems of importance and interest. Many papers ranging from the description of the mathematical and statistical models of the growth and extent of the AIDS epidemic to statistical procedures to key components of the disease process have been written as a result of the discovery.

Fusaro *et al.* (1989) in an attempt to analyse this area of research, wrote a paper with an annotated bibliography of over 100 articles containing quantitative methodology related to the AIDS epidemic. The bibliography contains a brief introduction and groups the articles into five areas, namely:-

- I. General Models of Transmission Dynamics
- II. Statistical Modeling of the Magnitude of the Epidemic
- III. Natural History of AIDS
 - (a) Incubation period
 - (b) Infectivity and infection distribution

(c) Survival analysis and disease progression

(d) Association between cases

IV. Screening Test

V. Clinical Trials

Since then, there has been a vast increase in the number of research projects describing the statistical models of the epidemic and key components of the disease process. In trying to summarize these work, Jewell (1990) provided an expanded version of the comments above with particular regard to certain topics. The statistical issues in studies of the epidemiology of AIDS were stated as:-

- (i) the size of the epidemic and the rate of development of new infections over time,
- (ii) the natural history of disease progression including estimation of the distribution of the incubation period,
- (iii) cofactors affecting progression of HIV disease,
- (iv) the length of the period of infectiousness and the level of infectivity during this period,
- (v) cofactors affecting infectivity,
- (vi) staging of the disease process,
- (vii) the study of markers of disease progression and onset of AIDS,
- (viii) treatments and their effects on markers and disease progression.

The eight areas mentioned above aimed to describe the epidemic and understand the natural history of the disease. The paper briefly surveyed some of the statistical methods that have been developed to meet the needs of analyzing the AIDS data. The paper focused on the projection of future AIDS cases and the estimation of the incubation distribution and those of the infectivity associated with transmission. Areas not touched

upon included the design and analysis of clinical trials of therapeutic agents and vaccines, screening programmes and the construction of useful clinical staging systems for the HIV disease.

Many problems are encountered in making future predictions about the HIV/AIDS epidemic because the various key components are complicated by many statistical issues. Anderson *et al.* (1989) gave at least three reasons for the somewhat daunting problems that surround the making of predictions about the future trends in the incidence of AIDS. Firstly, the diagnosis of the disease and its etiologic agent was recent and the key processes that determine the transmission within communities of people remain uncertain. This could be due to the interplay between the HIV virus and the immune system, which turns out to be significantly more dynamic than most scientists would have suspected. However, the HIV gains the upper hand and causes the severe immune impairment that define full blown AIDS. An additional problem is posed for epidemiological study because one of the primary routes of transmission is via sexual contact, given the sensitivities of individuals, societies and government bodies to open discussion and investigation of the behavioural characteristics that influence transmission via this pathway.

Secondly, the quality of the epidemiological data could be affected by poor reporting systems, long incubation periods and many other uncertainties. The poor reporting system could stem from underreporting which is common among developing countries and delay in reporting which is a time span between the actual diagnosis and the diagnosis being recorded in a central body for reporting AIDS cases. In addition, the long incubation period poses a lot of statistical uncertainties as it can range from a few years in the case of blood transfusion recipients to 10 years in the case of sexual transmission. In some cases, it is over 15 years.

Thirdly, the influence of education and publicity on patterns of behaviour and other areas have an impact on the course of the epidemic. In well-educated communities in cities that have been hardest hit by the HIV epidemic, major changes in sexual behaviour have been observed (Henderson and Saah, 1993). This has included reduction in the number of sexual partners and reduced unprotected anogenital intercourse. These behavioural changes were serologically confirmed with results showing a significant decline in the annual incidence of HIV infection. In one study in San Francisco, for example, the seroconversion rate dropped from 8% in 1985 to less than 1% in 1988. However, preventive education does not seem to have such encouraging results in heterosexual groups. These changes are thought to have an impact on the rate of spread of the virus but we are ignorant in quantitative terms of the temporal pattern of change in either incidence of new HIV infections or sexual behaviour.

1.4 Format of Presentation

Having briefly explained the current global situation and the main problem areas in AIDS studies, the purpose of this thesis is to study the various statistical methods and models proposed for the AIDS epidemic and its cofactors through the years 1985 to 1995 to better understand the epidemic and to indicate the problems or limitations as well as the uses for the various methods.

The HIV seroprevalent data is often plagued with various biases and censorings in addition to the fact that the incubation period is subjected to a lot of uncertainties. In Chapter 2, we discuss some of the ways of overcoming the biases and censorings in the data and also study the various mathematical models used to model the incubation period of the HIV infection.

In Chapter 3, we discuss some of the problems encountered in AIDS incidence data like delay reporting and changes in AIDS surveillance definition. We study how these data can be adjusted to give a more accurate figure. The reason is that AIDS incidence data is the most commonly available report and provide the vital source of information for the health department to assess the AIDS epidemic and to plan for future health care needs.

With the AIDS incidence data adjusted and the HIV seroprevalent data corrected, we can then use the information on an HIV transmission dynamics model in Chapter 4. This model gives an insight into the mechanism of the large scale phenomenon of the epidemic and contribute to the understanding of how different factors interact in determining the spread of HIV and the occurrence of AIDS.

In Chapter 5, we discuss the method of backcalculation for forecasting the AIDS epidemic which is intermediate between empirical data fitting and the mechanistic model discussed in Chapter 4. The method estimates past HIV infection rates from available AIDS incidence data and an estimate of the incubation period. The incubation period distribution is then applied to the estimated past HIV infection rate to project future AIDS incidence. The backcalculation method is then applied to the monthly AIDS incidence data in Malaysia from January 1995 until August 1996 released by the Ministry of Health, Malaysia to generate the estimated current cumulative HIV infection for Malaysia using the incubation period distribution from Brookmeyer (1991). The backcalculation program was obtained from Assistant Professor Dr. Peter Bacchetti. Communication with him was done by e-mail. In fact, literature concerning AIDS was obtained by searches on the World Wide Web of the Internet.

There is a need for the development and assessment of markers of disease progression for use as surrogate responses for the onset of AIDS. With this in mind, Chapter 6 discusses the uses of markers, especially CD4 T cell counts in modelling the progression of the HIV infection to estimate the residual time to AIDS.

In Chapter 7, we round up the discussion with a comparison of the various methods used in AIDS epidemic projection and study the type of methods to be used for various situations.

CHAPTER 2

NATURAL HISTORY OF AIDS

2.1 An Overview of the Natural History of AIDS

In the general course of the AIDS disease, when a susceptible becomes infected, he goes through a latent or window period whereby the antibodies for the virus are not detectable for a few weeks. This is also called the **pre-antibody phase**. Once the antibody is detected, the infected individual is said to undergo a seroconversion and he is now termed a **seropositive**. Isham (1988) mentioned of the convenience of assuming all infected as seropositive in modelling. The period from infection until overt symptoms appear is the **incubation period**. AIDS is the advanced stage of the HIV disease and is defined by a number of clinical life-threatening conditions. Many patients experience various symptoms or opportunistic infections which are not AIDS defining before an onset of AIDS. The end of the disease process is death since death is the current prognosis of the majority of AIDS cases. The diagram below gives a simplified schematic of the natural history of AIDS.

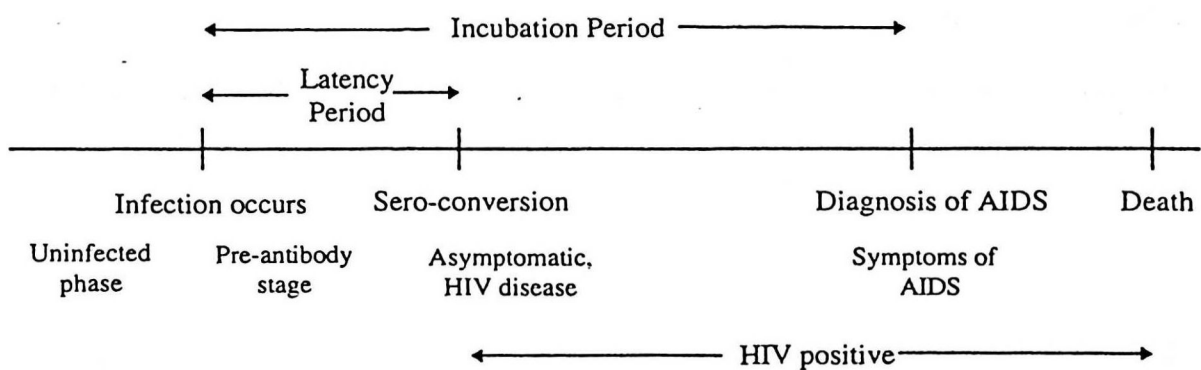


Figure 2.1 A simplified schematic of the natural history of AIDS

2.2 Survey of HIV Seroprevalence and Seroincidence

2.2.1 Non-Response Bias

HIV incidence can be estimated directly by studying cohort data or indirectly by studying the seroprevalence trends and by the method of backcalculation which will be discussed in Chapter 5. The survey data, however, is subjected to various biases and uncertainty. A major problem faced in the survey data is non-response bias whereby a fraction P of the sampled individuals from the population refuse to provide the blood. This will affect the seroprevalence of the entire population. Suppose θ_1 is the seroprevalence rate among the responders and $\theta_0 = \theta_1 f$ is the seroprevalence rate among nonresponders where f is the ratio of the rate of nonresponders compared to responders. Then the seroprevalence of the entire population is $\theta = \theta_1(1 - P) + P\theta_1 f$ and the ratio of true seroprevalence to that of responders is $\theta / \theta_1 = 1 + P(f - 1)$.

The ratio is 1 if everyone responds (i.e., $P = 0$) or if $f = 1$. $f = 1$ means that the responders and nonresponders have the same rate or in other words, the responders are representative of the entire population. Hull *et al.* (1988) studied nonresponse bias among patients in a New Mexico clinic for sexually transmitted diseases. It was noted even among admitted homosexual men, nonresponders had a much higher seroprevalence than responders.

2.2.2 Biases in Seroprevalent Data

The seroprevalent data consists of those who are already seropositive at the beginning of the studies and whose calendar times of infection is not known. The biases which result from follow-up time, from enrolment instead of time infection is thus investigated. Goedert *et al.* (1986) in a three-year AIDS incidence study, mentioned of a significantly

higher AIDS incidence in the New York cohorts compared with other cohorts if the virus entered the New York homosexual population before it entered the populations from which the other cohorts were drawn. Standard analytical tools for censored survival data do not apply for prevalent cohorts because all the individuals are censored.

The three time scales for calendar time (s), time from infection (u) and follow-up time (t) are illustrated in Figure 2.2 below (Brookmeyer and Gail, 1987). The follow-up starts at calendar time $s = Y$. For convenience, the origin of calendar time s , was chosen to coincide with the onset of the epidemic with epidemic density function $I(s) = \alpha r s^{r-1} / (1 + \alpha s^r)$ where $\alpha = 0.028$, $r = 2$ (Brookmeyer and Gail, 1987).

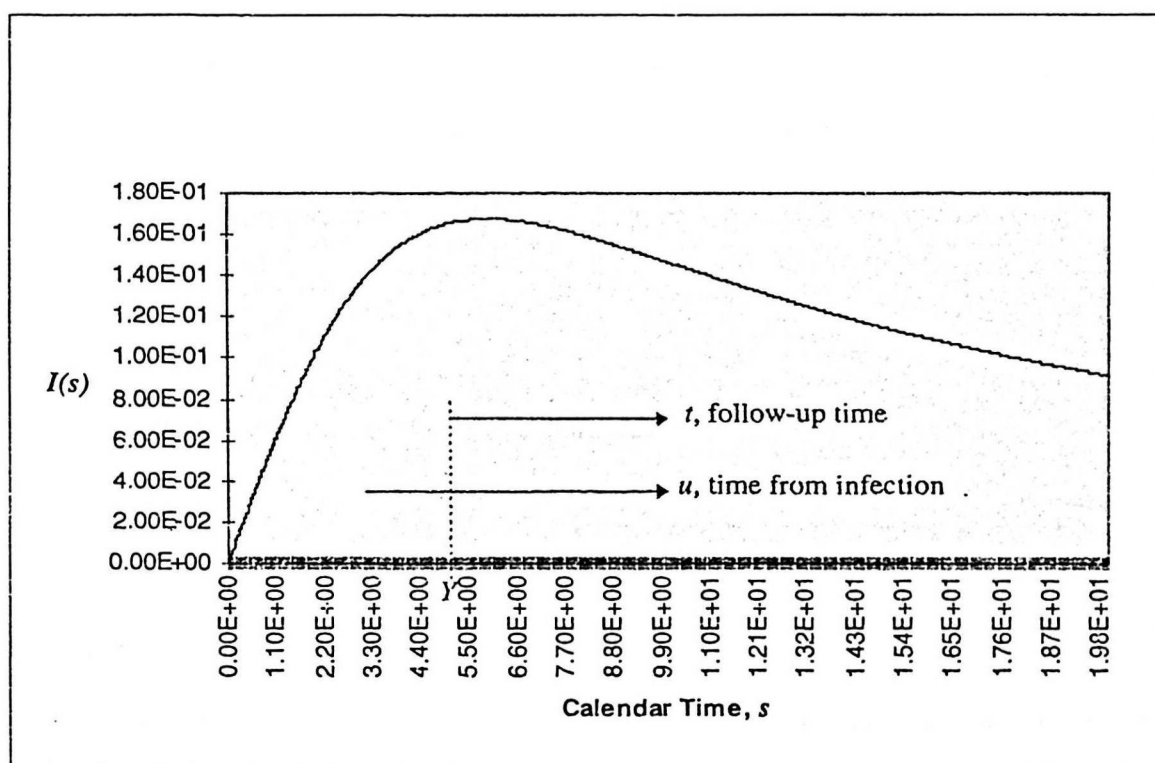


Figure 2.2 Epidemic density function $I(s)$ illustrated schematically in the prevalent cohort study

If we let $S(u) = 1 - F(u)$ be the survival function, $\lambda(u)$ the hazard function, then $S_p(t) = 1 - F_p(t)$, and $\lambda(t)$ are the corresponding survival and hazard functions in the follow-up time. If the hazard function $\lambda(u)$ is constant (an exponential distribution for F), then $S_p(t) = 1 - F_p(t)$ which follows from the lack of memory of the exponential distribution. If the hazard $\lambda(u)$ is increasing, then an individual in the prevalent cohorts group will be at greater risk of developing AIDS compared to the newly infected individual, that is, $S_p(t) < S(t)$. The converse is true for decreasing hazard $\lambda(u)$. If, however, the infected individuals are a mixture of a proportion α who eventually develop the disease and proportion $1 - \alpha$ who never develop the disease then, $S_p(t) \geq 1 - \alpha$.

Assuming that the time to infection and the time from infection to disease onset are independent, the induced survival function in the observed follow-up time scale, $S_p(t) = 1 - F_p(t)$ is the probability of being disease free at calendar time $Y + t$ conditional on being prevalent at calendar time Y and is given by Brookmeyer and Gail (1987) as

$$S_p(t) = \frac{\int_{-\infty}^Y I(s) S(Y - s + t) ds}{\int_{-\infty}^Y I(s) S(Y - s) ds}.$$

The prevalent cohorts provide the crucial information on the course of the disease process because prevalent cohorts are by definition at risk of the disease at the time of recruitment and hence supply more at-risk person-time and disease events for analysis than an incident cohort.

Munoz *et al.* (1992) separated the two subcohorts of prevalence and incidence to assess the effects of imputation of the data of infection on the estimation of the incubation period distribution. They found that the prevalent cohorts provide crucial information on the course of the disease process for they are, by definition, at risk of the disease at the time of recruitment and, hence, supply more at-risk person-time and disease events for analysis than incident cohorts. However, effective survival analysis of the incubation period requires the establishment of the time of onset of the infection. To complete the missing data of seroconversion in the seroprevalent cohort, Munoz *et al.* (1992) used the data provided by an incident cohort.

2.3 Mathematical Models for the Incubation Period

2.3.1 Introduction

The incubation period distribution $F(t)$ is the probability distribution for the time period t , that an individual is infected, which is usually taken as the time of seroconversion to the time AIDS develop. The model gives an understanding of the epidemiology of AIDS mainly by fitting the model to the data collected. The transmission dynamics will be discussed later in Chapter 4. The long and variable incubation may result from differences in the virulence, infectious mode, mode of transmission, host response and frequency of exposure.

If U represents the incubation distribution with cumulative density function $F(t) = P(U \leq t)$, then the survival function is $S(t) = 1 - F(t)$ and the hazard function is $\lambda(t) = f(t) / S(t)$ where $f(t)$ is the probability density function. The hazard function $\lambda(t)$ quantifies the risk of developing AIDS with time t from infection.