MODELLING THE AIDS EPIDEMIC IN MALAYSIA

Ong Hong Choon
School of Mathematical Sciences, University of Sciences of Malaysia, 11800 USM
Penang Malaysia
hcong@cs usm my

Abstract

There are generally three methods of modelling the acquired immuno deficiency syndrome (AIDS) epidemic At one extreme is the attempt to fit a function of calendar time such as a polynomial or other mathematically convenient curves to the AIDS incidence curve while the other extreme 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 method of backcalculation which is intermediate between the first two methods, estimates the past HIV infection rate from the AIDS incidence data and an estimate of the incubation period distribution This method is used on the Malaysian data to model the AIDS epidemic because it makes use of the Malaysian AIDS incidence which is fairly reliable and is more reflective of the trend of the epidemic as compared to the HIV infection rate recorded An application is made in this study on the AIDS incidence data in Malaysia released by the Ministry of Health, Malaysia using a backcalculation program and an approximate incubation period distribution to generate the current HIV infection rate for Malaysia

Keywords Backcalculation, AIDS modeling, HIV infection

1. Introduction

Methods for predicting the future trends in the incidence of AIDS are based on one of three methods These methods are similar in that they all fit some function of calendar time to the incidence of AIDS data but they differ in the degree to which the mechanisms that generatethe data are incorporated into the model

At one extreme, the method fits a function of calendar time such as a polynomial or other mathematically convenient curves to the AIDS incidence curve and then extrapolates into the future (see [10]) Although this method is easily implemented, it has the danger of extrapolating a fitted model outside the range of the observed data The prevalence of HIV infection cannot be estimated with this approach This method ignores what is known about the epidemiology of the disease and it cannot incorporate information that one might have on changingpatterns of transmission

The second method models the full dynamics of 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 short and medium term forecast of the number of cases (see [2]) Unfortunately, the deterministic models proposed for general epidemics are complicated, and the stochastic models are even more complicated

Furthermore, predictions based on such models are particularly sensitive to unknown parameters such as the long incubation distribution from infection to the development of AIDS, the frequency and pattern of sexual activity and behavioural changes which change with time and also the proportion of infected people who eventually develop AIDS with allowance for emigration, immigration and death Such models are complicated and contained many unknown parameters (see [11])

The third approach, to be explained in this study, which is intermediate between these two approaches, is the back projection method (see [3, 4, 9]) 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 requires reliable counts of the number of AIDS cases diagnosed over time and a reliable estimate of the incubation period distribution. The method is popular because it makes use of the AIDS incidence data which represent the most readily available information on the AIDS epidemic as most national AIDS surveillance data systems record only AIDS cases The incubation period distribution is then applied to the estimated past HIV infection rates to project future AIDS incidence

2. The Method of Backcalculation

The basic convolution equation in backcalculation relates the number of new cases of AIDS in time t to $t+dt$ $\left(\text{designated } Z(t)\right)$ and the number of new HIV infections $g(s)$ at each time s since the start of the epidemic $(s=0)$ through the incubation period distribution $f(u)$, where u is the time spent between the initial infection and the eventual diagnosis of AIDS. The basic convolution equation is given as

$$Z(t)=\int_0^t g(s)f(t-s)ds. \qquad (1)$$

From the above equation, for an individual to be diagnosed as an AIDS case by calendar time t, he or she must have been infected at some prior time s, and then have an incubation period less than $t-s$ In other words, the backcalculation method uses the above equation together with knowledge of $Z(t)$ (obtained from the AIDS cases registries) and $f(t)$ (obtained from the epidemiological studies) to give information on past infection rates $g(s)$ If $f(t)$ is known, the above relationship could be inverted to express $g(s)$ for all $0\le s\le t$ as a function of $Z(t)$ In general, a family of values for $g(s)$, $0\le s\le t$ can be constructed which are consistent with a realization of $Z(s)$, $0\le s\le t$.

Let $z_1, z_2, \ldots, z_n$ be the number of AIDS cases diagnosed in the calendar time interval $[t_{i-1}, t_i)$, $i = 1, 2, \ldots, n$ It is assumed that individuals become infected according to a point process. Then the expected number of AIDS cases occurring during the time interval $[t_{i-1}, t_i)$ is given by [7, p.198]

$$E(z_i) = \int_{t_0}^{t_i} g(s)\ \{F(t_i - s) - F(t_{i-1} - s\}ds \qquad (2)$$

where $F(t)$ is defined to be 0 for $t \leq 0$ By convention, we shall define calendar time 0 to be the start of the epidemic (that is $g(s) = 0$ prior to that time) and thus $t_0 = 0$

The method of backcalculation is used because it makes use of the AIDS incidence data which is more reflective of the trend of the epidemic The number of HIV+ cases, on the other hand is dependent on the test made and is unreliable as a trend For example, a steep rise in the number of HIV+ cases may be due to the mandatory testing of all intravenous drug users in drug rehabilitation centres and increase in detection through aggressive case finding

3. Result of the Backcalculation Method on the Malaysian Data

The definitive diagnostic method for diseases indicative of AIDS used in Malaysia is taken from the Ministry of Health's publication *"Plan of Action for Prevention and Control of AIDS"* released in May 1988 This definition is similar to the CDC's 1987 definition and the WHOdefinition and has been in used in Malaysia since 1988 The under reporting rate is assumed to be 10% (around 90% reported) throughout and the reporting delay, which is aound 2 to 4 weeks, from the various districts to the AIDS section, Ministry of Health is asumed to be negligible

The backcalculation method is applied by [13] on the Malaysian data until August 1996 and obtained the estimated number of cumulative HIV infection as shown in Figure 1

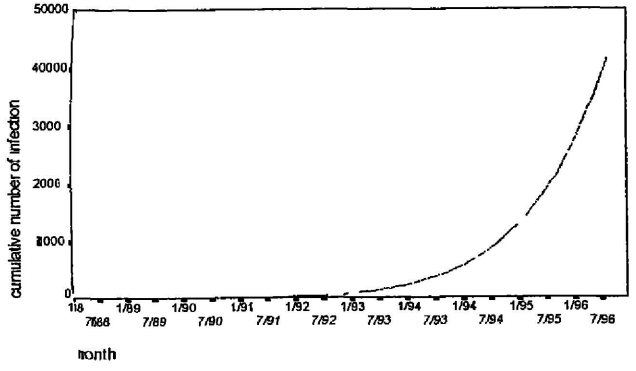

Figure 1 Estimated cumulative number of HIV infection applied to the Malaysian data till August 1996 (see [13])

An obvious result from Figure 1 is that the HIV/AIDS epidemic in Malaysia is in their early stage which is evident in its rapid exponential increase in the number of infected cases until August 1996 A backcalculation program in Fortran from [3] is used again on the Malaysian data in this paper The program is based on the incubation period distribution from [4]. The basis for using the incubation period distribution from [4], which is based mainly on homosexuals, on the Malaysian data which are mainly intravenous drug users is because the incubation period distribution of the two cohorts are similar (see[12])

Table 1 Cumulative HIV+ and AIDS cases recorded from the Ministry of Health, Malaysia (see [1]) and the corresponding estimated cumulative HIV+ cases from the backcalculation method in this study.

Year	Cumulative HIV+ cases recorded	AIDS Cases reported	Estimated cumulative HIV+ Cases
1986	3	1	1.62
1987	5	0	402.26
1988	14	2	1036 68
1989	214	2	1828 84
1990	992	18	2897 21
1991	2786	60	4502 37
1992	5298	73	7208 14
1993	7805	71	12064 01
1994	11198	105	20214 95
1995	15396	233	30897 54
1996	19993	347	40536 07
1997	23917	568	46741 35
1998	28541	875	50055 38
1999	33233	1200	51718 45
2000	38340	1168	52542 59
2001	44278	1302	52950 32
2002	51256	1193	53152 01

There is a slowdown in the increase of the number of estimated infected HIV+ cases in the late 1990s as can be seen from Figure 2. This trend is supported by a slowdown in the increase of the number of AIDS cases (as can be seen in Table 1) which gives a clear picture of the trend of the epidemic Also, there is a narrowing of the gap between the number of recorded cumulative and estimated HIV+ cases This is probably due to an increasing awareness among the population (especially the high risk group) towards the epidemic and also due the efforts by the Malaysian government and non-government organizations to promote this awareness This slowdown in the increase is also similar to the trend in developed countries like the US and UK where the AIDS epidemic had begun earlier (see [3, 8])

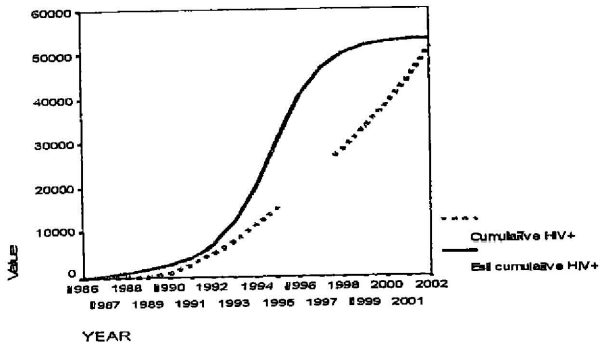

Figure 2 A comparison between the recorded and estimated cumulative number of HIV infections in Malaysia

Beside the uncertainty resulting from inaccurate knowledge of the inputs of AIDS data, the estimates from the method of backcalculation have some inherent stochastic uncertainty. The confidence bounds in Table 2 reflect only the uncertainty that would be present if we actually know the incubation and reporting inputs and the smoothness weights

Table 2 Point wise 95% lower and upper confidence bounds of the estimated HIV+ incidences

Year	Cumulative HIV+ estimated	Lower bound	Upper bound
1986	1.62	0	23 71
1987	402.26	229.72	574.79
1988	1036 68	789.61	1281 75
1989	1828 84	1590 70	2066 98
1990	2897 21	2590 98	3203 44
1991	4502 37	4000 09	5004 66
1992	7208 14	6478 84	7937 44
1993	12064 01	11082 89	13045 12
1994	20214 95	18796 78	21633 12
1995	30897 54	28951 01	32844 07
1996	40536 07	38096 14	42976 00
1997	46741 35	43744 66	49748 03
1998	50055 38	46405.93	53704 83
1999	51718 45	47428.19	56008 72
2000	52542 59	47705.98	57379 20
2001	52950.32	47705 65	58194.98
2002	53152 01	47630 46	58673 56

4. Discussion

There are several potential sources of error and limitations underlying the assumption of the methodology Firstly, the parametric model for the HIV infection rate provide no information about future incidence rate as it only attempts to estimate the historical infection rates It is for this reason that backcalculation is referred to as a method for estimating the minimum size of the epidemic (see [6]) Secondly, there is little information about the recent infection rate because of the long incubation period However, short term projections of AIDS are reliable because such projections depend more strongly on the infection rate in the distant past than the recent ones The third limitation is that the incubation distribution, F, is not known precisely although it is assumed known The incubation distribution may be different for different subgroups of infected individuals with age as a cofactor of disease progression Fourthly, the assumption of dependence between the calendar date of infection and incubation period implicit in the convolution equation would be violated if cofactors of disease progression are identified that are more prevalent among those infected earlier (or later) in calendar time Also, a smaller f needs to be compensated by a larger g in order to fit the cumulative AIDS incidence series The fifth potential source of limitation is in the inaccuracies in the AIDS incidence data over time It could be clouded with issues like reporting delays and changes in AIDS definition, discussed in the previous section

The application of backcalculation to data is useful in several aspects Firstly, the backcalculation method provides a simple conceptual framework for relating the incubation distribution with the AIDS incidence data and the infection rate Secondly, backcalculation leads to short-term projections of AIDS incidence that are robust to changes in the incubation distribution (see [5, 6, 14]) Thirdly, although backcalculation estimates of cumulative infections are known to be highly sensitive to the choice of f, plausible ranges of estimates of g from backcalculation for the number infected in the United States based on data through mid-1987 were in broad agreement with estimates based on surveys in selected populations (see [6, 14])

References

1 http //dph gov my/aids/ Accessed on March 3, 2005

2 M Arca, C A Perucci and T Spadea, "The epidemic dynamics of HIV-1 in Italy Modelling the interaction between intravenous drug users and heterosexual population", *Statistics in Medicine* **11**(1992), 1657-1684

3 P Bacchetti, M R Segal and N P Jewell, "Backcalculation of HIV infection rates", *Statistical Science* **8**(1993), 82-119

4 R Brookmeyer, "Reconstruction and future trends of the AIDS epidemic in the United States", *Science* **253**(1991), 37-42

5 R Brookmeyer and A Damiano, "Statistical methods for short term projections of AIDS incidence", *Statistics in Medicine* **8**(1989), 23-34

6 R Brookmeyer and M H Gail, "A method for obtaining short-term projection and lower bounds on the size of the AIDS epidemic", *Journal of the American Statistical Associations* **83**(1988), 301-308

7 R. Brookmeyer and M H Gail, "AIDS Epidemiology A Quantitative Approach", New York Oxford University Press (1994)

8 F Chiarotti, M Palombi, S Nicola, A Ghirardini, and L Prospero, "Effects of different parametric estimates of seroconversion time on analysis of progression to AIDS among Italian HIV positive haemophiliacs", *Statistics in Medicine* **11**(1992), 591-601

9 N E Day, S M Gore and D De Angelis, "Acquired immune deficiency syndrome predictions for England and Wales (1992-1997) Sensitivity analysis, information, decision", *Journal of Royal Statistical Society A* **158**(1995), 505-524

10 M J R Healy and H E Tillett, "Short-term extrapolation of the AIDS epidemic", *Journal of the Royal Statistical Society A* **151**(1988), 50-61

11 S H Heisterkamp, B J De Haan, J C Jager, J A M Van Druten and J C M Hendricks, "Short and medium term projections of the AIDS/HIV epidemic by dynamic model with an application to the risk group of homo/bisexual men in Amsterdam", *Statistics in Medicine* **11**(1992), 1425-1441

12 A B Mariotto, S Mariotti, P Pezzotti, G Rezza, and A Verdecchia, "Estimation of the acquired immunodeficiency syndrome incubation period in intravenous drug users A comparison with male homosexuals", *American Journal of Epidemiology* **135**(1992) 428-437

13 H C Ong, S H Quah and H C Low, "An application of the backcalculation method to estimate past HIV infection rates in Malaysia", *The Medical Journal of Malaysia* **53**(1998) 385-391

14 J M G Taylor, "Models for the HIV infection and AIDS epidemic in the United States", *Statistics in Medicine* **8**(1989), 45-58