AIDS: Exponential vs. Polynomial Growth Models

Harry H. Panjer
University of Waterloo
August 16, 1988

Abstract

Epidemic theory suggests that the early part of an epidemic can be characterized by exponential growth in the number of infections. This model is tested using aggregate data of reported AIDS cases in Canada and the U.S.A. Alternative polynomial type growth models are shown to provide a better fit to the data when gamma incubation period models are used.

Presented at 23rd Actuarial Research Conference, Storrs, CT. August 1988.

PRELIMINARY DRAFT

Introduction

In building models of AIDS-related mortality, actuaries need to recognize various components of the development of the epidemic. The stochastic process of the number of persons infected with the human immunodeficiency virus (HIV), (previously called HTLV-III, more recently called HIV-1) has been modelled using epidemic theory. Two recent papers (Anderson, 1988) and (Isham, 1988) provide extensive discussion of theoretical justification for various infection growth models.

Following infection, a person become infectious and may infect others. After some period of time, called the incubation period, a diagnosis of clinical AIDS may be made once the disease is sufficiently developed. Various statistical models of the incubation period have been proposed. These are generally based on retrospective studies of persons who have been diagnosed as having developed AIDS. The time of initial infection is identified accurately as possible for each of these persons.

In general, there is little information available regarding the number of persons infected to date. The number of persons who have developed clinical AIDS are reported in the U.S.A. by the Centre for Disease Control (CDC) and in Canada by the Federal Centre for AIDS (FCA). Using this data and information about the incubation period, it is possible to estimate the number of persons who are already infected but who have not yet been diagnosed as having developed clinical AIDS.

This information can be used by the actuary to estimate the number of potential AIDS cases that may be in a subpopulation of insureds in order to develop estimates of additional reserve requirements resulting from AIDS infection.

In this paper, we limit ourselves to an analysis of the effect of the choice of infection models on the estimate of the number of persons who are already infected with HIV.

The Mathematical Model

In developing a model, we shall be guided by the principle of parsimony which, simply stated, requires that a model used to explain variation in a set of data should be as simple as possible and include no more parameters than can be justified by the data. In this way the results of the analysis do not become overly model-dependent; that is small changes in data will not result in large changes in the results.

For a given (sub)population, let $M(t)$ be a stochastic process denoting the number of HIV+ infecteds at time t and let $N(t)$ denote the number of persons who have developed AIDS at time t. Then the set of persons represented by $N(t)$ is a subset of the persons represented by $M(t)$; the remainder $M(t)-N(t)$ representing those persons who are somewhere in the incubation period.

Let $G(t)$ denote the distribution function of the time from infection to development of clinical AIDS. This will be called the incubation period (or time) model. Throughout this paper we shall consider a gamma distribution of incubation times. This is one of two incubation period models considered by Medley *et al.* (1987), the other being the Weibull. They considered the Weibull and the gamma models to be indistinguishable on the basis of the data which they were analyzing. Other possible distributions for the incubation period include the normal distribution (Rees, 1987) and the generalized Erlang distribution (Panjer, 1987). The probability distribution function of the gamma distribution is given by

$$g(t) = \frac{d}{dt}G(t) = \frac{q^p}{\Gamma(p)}t^{p-1}e^{-qt}, \quad t>0 \tag{1}$$

in the notation of Medley *et al.* (1987).

In the interest of simplicity, we use this model for all age groups and for all subgroups (male homosexuals, IV drug abusers, recipients of blood products) combined.

Using epidemic theory, various authors justify the use of logistic growth models for homogeneous subgroups (cf. Anderson, 1988 and Isham, 1988). These logistic growth models are approximately exponential in the early stages of the epidemic. When a population consists of a number of different subgroups with different characteristics (i.e. different rates of growth), the resulting mixture of subgroups is no longer exponential. Panjer (1988) used exponential models and showed that the observed rate of growth of AIDS cases in Canada, Australia and the U.S.A. was not constant and appeared to be slowing. This is supported by observations that the "doubling time" is increasing.

We shall assume that the number of infections $M(t)$ follows a Poisson process with mean $\Gamma(t) = \int_{-\infty}^{t} \gamma(s)ds$ as is done by Medley *et al.* (1987). Then the number of AIDS cases $N(t)$ is also a Poisson process with mean

$$\Lambda(t) = \int_{-\infty}^{t} \gamma(s)G(t-s)ds = \int_{0}^{\infty} g(r)\Gamma(t-r)dr \tag{2}$$

(cf. Ross, 1985).

In the next section, this gamma incubation period model is used in conjunction with exponential and polynomial infection growth models.

Estimating the Incubation Model Parameters

We consider three growth models for the mean of infection process;

I. Exponential

$$\Gamma(t) = ke^{\beta t}, \tag{3}$$

II. "Truncated" Exponential

$$\Gamma(t) = k\{e^{\beta(t-t_0)}-1\}, \quad t>t_0 \tag{4}$$

III. Polynomial

$$\Gamma(t) = k(t-t_0)^{\beta}, \quad t>t_0 \tag{5}$$

The exponential growth model was used by Medley *et al.* (1987), Kalbfleisch and Lawless (1987) and others studying models of the incubation period. Using (2), the mean of the stochastic process of reported AIDS cases can be computed. The results are

I. Exponential

$$\Lambda(t) = k(\frac{q}{q+\beta})^p e^{\beta t}, \quad t>-\infty \tag{6}$$

II. "Truncated" Exponential

$$\Lambda(t) = k(\frac{q}{q+\beta})^p e^{\beta(t-t_0)}\, I(p;(q+\beta)(t-t_0)) - k\, I(p;q(t-t_0)), \quad t>t_0 \tag{7}$$

where $I(p;x)$ is the incomplete gamma function (Abramowitz and Stegun, 1970, p. 261)

$$I(p;x) = \int_0^x \frac{1}{(p-1)!} y^{p-1} e^{-y} dy = \frac{x^p e^{-x}}{p!} M(a,1+a,x) \tag{8}$$

where $M(a,b,z)$ is defined below.

III. Polynomial

$$\Lambda(t) = \frac{q^p (t-t_0)^{\beta+p} e^{-q(t-t_0)}}{(p-1)!\binom{\beta+p}{p}} M(\beta+1,\beta+p+1,q(t-t_0)), \quad t>t_0 \tag{9}$$

where $M(a,b,z)$ is the coefficient hypergeometric function (see Abramowitz and Stegun, 1970, p. 505)

$$M(a,b,z) = \int_0^1 \binom{b}{a} e^{zt} t^{a-1} (1-t)^{b-a-1} dt = \sum_{k=0}^{\infty} \frac{(a)_n z^n}{(b)_n n!} \tag{10}$$

where $(a)_n = a(a+1)(a+2) \cdots (a+n-1)$ and $(a)_0 = 1$.

The truncated growth models each have an "origin" of x_0. This origin may be loosely interpreted as the time of the initial entry of the virus into the country under study, although it should more correctly be considered as a mathematical intercept of the growth curve.

Maximum likelihood estimates for CDC and FCA data as at June 6, 1988 are obtained using the data that appear in Tables 1 to 4. Let $-\infty = t_0 < t_1 < t_2 < \cdots < t_3$ denote the end points of time intervals and n_i denote the number of reported AIDS cases in the time interval $(t_{i-1},t_i]$; $i = 1,2,\ldots,m$ with $n = n_1+n_2+\cdots n_m$. The likelihood function is written as

$$L = \prod_{i=1}^{m} \frac{\{\Lambda(t_i)-\Lambda(t_{i-1})\}^{n_i}}{n_i!} e^{-(\Lambda(t_i)-\Lambda(t_{i-1}))} \tag{11}$$

the log likelihood is

$$l \propto \sum_{i=1}^{m} n_i \log\{\Lambda(t_i)-\Lambda(t_{i-1})\}-\Lambda\{t_m\}. \tag{12}$$

This function is maximized with respect to the parameters under consideration.

For the two parameter exponential model, this function is maximized by differentiating (12) with respect to each parameter and setting these derivatives to zero. This procedure is described by Panjer (1988). For the two models with a third parameter (the origin t_0), the log likelihood is maximized numerically over all three parameters. Although, this is rather tedious and time consuming, it can be done quite easily on a computer.

The results of the estimation are given in Tables 1 to 4. Tables 1 and 3 are based on a mean and standard deviation of the incubation period of 8.2307 years and 3.6585 years respectively. These are the parameter values obtained by Medley et al. (1987) when they used a Weibull model for adults (5-59 years). Tables 2 and 4 are based on mean and standard deviation of 6.4059 years and 2.8294 years respectively. These are based on Medley *et al.* (1987) for all ages combined.

Discussion of Results

Tables 1 to 4 give both the number of fitted AIDS cases as well as the predicted number of HIV infections for the time periods inherent in the data. Both models with an origin have an MLE of the origin of early 1977 for the polynomial distribution and late 1978 for the exponential model.

The exponential model has an MLE of .672 for a doubling time of about one year. The estimate of 56 thousand infecteds to date in Table 1 is reasonably consistent with most estimates of about 50 thousand. With the shorter average incubation period, the number is reduced to about 30 thousand. It should be noted that the significance level of the chi-squared test of fit is less than .1% indicating a very poor fit. The fitted cases exceed the observed cases in early and late periods and the reverse holds in the intermediate periods, indicating that the epidemic has grown at a slower rate than that given by an exponential growth model. Furthermore, any short-term forecasting would be seriously biased.

When an "origin" is introduced, the exponential parameter β is reduced to .448 in Table 1 and .497 in Table 2 indicating an approximate doubling time of about 1 1/2 years. This indicates a slower growth of the epidemic. The significance level of the chi-squared statistics indicates that the fit, although better is still marginal. Furthermore, an examination of the residuals reveals a similar but less serious pattern. It should be noted that the fitted number of cases is approximately halved by moving to this model.

The polynomial model fits very well and cannot be rejected at any reasonable significance level. The exponent of 3.1 and 3.4 in Tables 1 and 2 indicate that the epidemic of HIV infection is approximately cubic. The fitted number of cases is very close to the observed number of cases and the residuals appear to be reasonably random. However, the predicted number of cases is reduced even further.

This suggests that if the popular figure of 50,000 infections is to be believed, the incubation period model needs to have a significantly larger mean.

Tables 3 and 4 which deal with the U.S.A. can be examined for comparative purposes. The exponential parameter β has consistently lower values than for Canada. Although it is known that the epidemic started earlier in the U.S.A., the "origin" for the U.S.A. is about one year later than for Canada. Furthermore the chi-squared values, although reduced, remains too large to be acceptable. Further, improvement of the model is required. It should be noted that the number of predicted HIV infections of 1.7 million in Table 3 is consistent with popular estimates of 1.5 million. Changing to the better polynomial model reduces this number by a factor of almost four.

Conclusions

Although the fitted number of AIDS cases depends upon both the infection or incubation models, it appears that serious fitting problems of the exponential can be overcome (for Canada) or partially overcome (for U.S.A.) by using a simple polynomial growth model. Suspicious final results of predicted HIV infection suggests that further refinements of the incubation model may be required. Such improvements will undoubtedly appear in the scientific and medical literature in due course.

References

Abramowitz, M. and Stegun, I.A. (1970). *Handbook of Mathematical Functions*, over, New York.

Anderson, R.M. (1988). The Epidemiology of HIV Infection: Variable Incubation Plus Infectious Periods and Heterogeneity in Sexual Activity. *J.R. Statist. Soc.* A (to appear).

Isham, V. (1988). Mathematical Modelling of the Transmission Dynamics of HIV Infection and AIDS: A Review. *J.R. Statist. Soc.* A, 151 (to appear).

Kalbfleisch, J.D. and Lawless, J.F. (1987). A Comment on Estimation of Incubation Periods and Epidemic Parameters Based on Data on AIDS Patients Infected by Blood Transfusion. *Technical Report STAT-87-24*, Department of Statistics and Actuarial Science, University of Waterloo.

Medley, G.F. *et al.* (1987). Incubation Period of AIDS in Patients Infected Via Blood Transfusion. *Nature*, 328, 719-721.

Panjer, H.H. (1987). AIDS: Survival Analysis of Persons Testing HIV+. *Working Paper ACTSC 87-14*. Department of Statistics and Actuarial Science, University of Waterloo. To appear in Transactions of the Society of Actuaries, 1988.

Panjer, H.H. (1988). AIDS: Some Aspects of Modelling the Insurance Risk. *Research Report 88-10*, Institute of Insurance and Pension Research, University of Waterloo. Presented at Conference on AIDS and Insurance, May 1988.

Rees, M. (1987). The sombre view of AIDS. *Nature*, 326, 343-345.

Ross, S.M. (1985). *Introduction to Probability Models*, Academic Press, Orlando.

TABLE 1

PREDICTED HIV INFECTIONS - CANADA

Source: "Surveillance Update: AIDS in Canada", Federal Centre for AIDS, June 6, 1988

Incubation Period Model: Gamma Distribution

Mean = 8.2307

Standard Deviation = 3.6585

			Exponential		Exponential		Polynomial	
Year	Quarter	Observed AIDS Cases	Fitted AIDS Cases	Predicted HIV Infections	Fitted AIDS Cases	Predicted HIV Infections	Fitted AIDS Cases	Predicted HIV Infections
to 1979	4	1	9	365	0	408	0	329
1980	1-4	4	8	350	1	559	2	553
1981	1-4	6	16	685	9	874	8	976
1982	1-4	22	32	1,341	30	1,369	28	1,523
1983	1	14	12	501	13	449	12	481
	2	15	14	593	16	502	15	525
	3	11	17	701	20	562	19	571
	4	14	20	829	24	629	23	619
1984	1	30	23	981	29	703	28	669
	2	34	28	1,161	34	786	34	721
	3	40	33	1,373	40	880	40	775
	4	43	39	1,624	47	984	47	831
1985	1	59	46	1,921	55	1,100	55	890
	2	77	54	2,272	63	1,231	64	950
	3	93	64	2,688	73	1,377	74	1,012
	4	96	76	3,179	83	1,540	85	1,076
1986	1	91	90	3,761	95	1,722	97	1,142
	2	122	106	4,449	108	1,926	110	1,211
	3	127	125	5,262	123	2,155	125	1,281
	4	143	148	6,225	140	2,410	141	1,353
1987	1	138	175	7,363	158	2,696	158	1,428
	2	162	207	8,710	179	3,015	176	1,504
TOTAL		1,342	1,342	56,335	1,342	27,876	1,342	20,420
Origin			—		1978.81		1977.37	
$\hat{k}$			186.5		580.0		17.017	
$\hat{\beta}$			.672		.448		3.062	
x^2			80.7		28.5		23.2	
df			19		16		16	
Sig. Level			< .1%		2.5%		> 10%	

TABLE 2

PREDICTED HIV INFECTIONS - CANADA

Source: "Surveillance Update: AIDS in Canada", Federal Centre for AIDS, June 6, 1988

Incubation Period Model: Gamma Distribution

Mean = 6.4059

Standard Deviation = 2.8294

			Exponential		Exponential		Polynomial	
Year	Quarter	Observed AIDS Cases	Fitted AIDS Cases	Predicted HIV Infections	Fitted AIDS Cases	Predicted HIV Infections	Fitted AIDS Cases	Predicted HIV Infections
to 1979	4	1	9	198	0	186	0	142
1980	1-4	4	8	190	1	291	2	275
1981	1-4	6	16	371	10	478	8	529
1982	1-4	22	32	726	32	786	28	886
1983	1	14	12	272	14	266	12	290
	2	15	14	321	17	301	15	321
	3	11	17	380	20	341	19	355
	4	14	20	449	24	386	23	390
1984	1	30	23	532	29	437	28	427
	2	34	28	629	34	494	34	465
	3	40	33	744	40	560	40	506
	4	43	39	880	46	634	47	550
1985	1	59	46	1,041	54	718	55	595
	2	77	54	1,231	62	813	64	642
	3	93	64	1,457	72	920	74	691
	4	96	76	1,723	82	1,042	85	743
1986	1	91	90	2,038	94	1,180	97	797
	2	122	106	2,411	108	1,336	110	853
	3	127	125	2,852	123	1,512	125	911
	4	143	148	3,373	140	1,713	141	971
1987	1	138	175	3,990	159	1,939	158	1,034
	2	162	207	4,720	181	2,196	176	1,099
TOTAL		1,342	1,342	30,527	1,342	18,526	1,342	13,471
Origin			—		1978.93		1977.29	
$\hat{k}$			101.1		266.0		4.589	
$\hat{\beta}$			.672		.497		3.44	
x^2			80.8		31.7		23.3	
df			19		16		16	
Sig.Level			< .1%		1%		> 10%	

TABLE 3

PREDICTED HIV INFECTIONS - U.S.A.

Source:"AIDS Weekly Surveillance Report", Centers for Disease Control, June 6, 1988

Incubation Period Model: Gamma Distribution

Mean = 8.2307

Standard Deviation = 3.6585

			Exponential		Exponential		Polynomial	
Year	Half Year	Observed AIDS Cases	Fitted AIDS Cases	Predicted HIV Infections	Fitted AIDS Cases	Predicted HIV Infections	Fitted AIDS Cases	Predicted HIV Infections
to 1981	1	89	1,013	37,483	83	57,908	98	56,823
	2	182	382	14,139	173	19,062	177	20,635
1982	1	368	526	19,472	372	21,590	368	23,252
	2	655	725	26,817	686	24,452	672	25,761
1983	1	1,241	998	36,933	1,137	27,694	1,114	28,180
	2	1,611	1,375	50,865	1,737	31,366	1,714	30,523
1984	1	2,500	1,893	70,052	2,497	35,524	2,484	32,798
	2	3,261	2,608	96,476	3,422	40,234	3,430	35,015
1985	1	4,573	3,591	132,868	4,516	45,569	4,549	37,179
	2	5,890	4,946	182,987	5,782	51,610	5,837	39,296
1986	1	7,453	6,812	252,012	7,227	58,453	7,281	41,371
	2	8,790	9,381	347,074	8,856	66,203	8,867	43,407
1987	1	10,557	12,920	477,994	10,682	74,980	10,580	45,407
TOTAL		47,170	47,170	1,745,171	47,170	554,645	47,170	459,647
Origin			—		1979.43		1978.96	
$\hat{k}$			27,216		85,871		11,392	
$\hat{\beta}$			.640		.249		1.724	
x^2			2,438		40.5		35.7	
df			10		9		9	
Sig.Level			0		< .1%		< .1%	

TABLE 4

PREDICTED HIV INFECTIONS - U.S.A.

Source:"AIDS Weekly Surveillance Report", Centers for Disease Control, June 6, 1988

Incubation Period Model: Gamma Distribution

Mean = 6.4059

Standard Deviation = 2.8294

			Exponential		Exponential		Polynomial	
Year	Half Year	Observed AIDS Cases	Fitted AIDS Cases	Predicted HIV Infections	Fitted AIDS Cases	Predicted HIV Infections	Fitted AIDS Cases	Predicted HIV Infections
to 1981	1	89	1,013	20,612	73	28,813	96	27,365
	2	182	382	7,775	170	11,139	175	12,118
1982	1	368	526	10,708	376	13,137	366	14,457
	2	655	725	14,747	700	15,494	670	16,830
1983	1	1,241	998	20,309	1,159	18,273	1,113	19,234
	2	1,611	1,375	27,970	1,760	21,551	1,714	21,666
1984	1	2,500	1,893	38,521	2,511	25,417	2,483	24,122
	2	3,261	2,608	53,052	3,416	29,977	3,428	26,602
1985	1	4,573	3,591	73,064	4,486	35,354	4,546	29,102
	2	5,890	4,946	100,624	5,732	41,697	5,832	31,622
1986	1	7,453	6,812	138,581	7,176	49,177	7,278	34,160
	2	8,790	9,381	190,855	8,843	57,999	8,870	36,715
1987	1	10,557	12,920	262,847	10,769	68,403	10,597	39,286
TOTAL		47,170	47,170	959,664	47,170	416,430	47,170	333,279
Origin			—		1979.61		1978.88	
$\hat{k}$			14,966		32,278		3,624	
$\hat{\beta}$			.640		.330		2.10	
x^2			2,438		54.3		36.0	
df			10		9		9	
Sig.Level			0		0		< .1%	