PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Estimation of the Incidence of HIV Infection

Valerie Isham

Phil. Trans. R. Soc. Lond. B 1989 **325**, 113-121
doi: 10.1098/rstb.1989.0076

Phil. Trans. R. Soc. Lond. B **325**, 113–121 (1989) [113]
Printed in Great Britain

ESTIMATION OF THE INCIDENCE OF HIV INFECTION

By VALERIE ISHAM

Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, U.K.

The aim of the method of 'back projection' is to provide estimates of the number of new infections with the human immunodeficiency virus (HIV) as a function of time, by using the numbers of diagnoses of the acquired immune deficiency syndrome (AIDS) together with information on the distribution of the incubation period between infection and diagnosis. Here, the method is investigated with particular reference to cases of HIV infection and AIDS in the United Kingdom.

1. Introduction

Knowledge of the prevalence of HIV infection in a particular population, and the rate at which new infections are occurring is clearly of great importance. However, for most populations of interest this information is not available. The serological data that have been collected usually relate to extremely specific groups of mostly high-risk individuals and these data cannot easily be combined to predict seroprevalence in a wider, more heterogeneous, population.

It is reasonable to suppose that new cases of HIV infection occur in a point process. For the moment we consider only those cases for which AIDS will eventually be diagnosed and denote the intensity of the corresponding point process by $h(t)$. Let the lengths of the incubation periods (between infection and diagnosis) for these individuals be independent, identically distributed variables with probability density function f. Then new diagnoses of AIDS will occur in a point process with intensity $a(t)$, where $a(t)$ is given by

$$a(t) = \int_0^\infty h(t-u) f(u)\, \mathrm{d}u. \tag{1}$$

Thus if the density of the incubation periods were known, together with the HIV infection rate $h(u)$ for $u \leqslant t$, we could calculate the distribution of the number of new diagnoses of AIDS in any time period up to time t. In many ways this would be the most natural method of predicting AIDS incidence.

Conversely, we can use equation (1) to deduce h if the functions a and f are known. This forms the basis of the method known as *back projection* in which knowledge about AIDS incidence and the distribution of incubation periods is used to make inferences about the incidence of HIV infection. It must be stressed, however, that because the proportion of those infected who will ultimately have AIDS diagnosed is not known, this method only provides information about the process of HIV infections that do subsequently lead to diagnoses of AIDS.

A related non-parametric method in discrete time of estimating the number of those already infected with HIV, which is then used to project the short-term future number of AIDS cases in

the United States, is described by Brookmeyer & Gail (1986, 1988). Their work yields a lower bound on the size of the epidemic in the sense that no allowance is made for further infections.

Earlier work using back projection methods includes that of Iverson & Engen (1986) who estimate the number of those in the United States infected with HIV by blood transfusion; Rees (1987) who discusses the total numbers of those infected in the United States and in the United Kingdom, and Boldsen *et al.* (1988) who consider the total numbers of those infected in the United States and in Denmark. The assumed forms of the functions a and f can have a large influence on the estimates obtained and various possibilities have been investigated by these authors.

This paper uses the method of back projection specifically in the context of the AIDS epidemic within the population of the United Kingdom, partly to see what can be inferred about the number of individuals who have been infected with HIV over the past few years and of those who may become infected in the near future, but also to investigate the implications of specific assumptions about the functions a and f. Theoretical aspects of the method will be considered in §2 and numerical results discussed in §3.

Rates of progression from HIV infection to AIDS have been published by the Centers for Disease Control, CDC (1987). These progression rates combine the probability that an infected individual will develop AIDS with the conditional distribution of the incubation period given that AIDS is eventually diagnosed. These rates can therefore be used in a back projection calculation to produce estimates of all those infected with HIV, rather than only those ultimately developing AIDS. Results of some discrete time calculations for the United States are given in CDC (1987). The use of the CDC progression rates in the context of the epidemic in the United Kingdom is discussed in §4, whereas §5 contains some general remarks and concluding comments.

2. Convolution inversion in specific cases

Suppose that $\tilde{a}(s)$ denotes the Fourier transform of $a(t)$,

$$\tilde{a}(s) = \int_{-\infty}^{\infty} \mathrm{e}^{ist} a(t)\,\mathrm{d}t,$$

and similarly for $\tilde{h}(s)$ and $\tilde{f}(s)$. Then it follows immediately from (1) that

$$\tilde{a}(s) = \tilde{h}(s)\tilde{f}(s), \tag{2}$$

and therefore that $h(t)$ can be obtained by taking the inverse transform of $\tilde{a}(s)/\tilde{f}(s)$.

We shall assume some basic parametric forms for a and f. In particular, for $a(t)$ we suppose either that

$$a(t) = a_0 \exp(a_1 t - a_2 t^2), \tag{3}$$

or that

$$a(t) = (b_0 + b_1 t)\,[1 + \exp(b_2 - b_3 t)]^{-1} \tag{4}$$

(see Cox & Medley 1989). There are good theoretical grounds for expecting that the curve of $a(t)$ should be close to an exponential curve in the early part of the epidemic and that this rapid growth should gradually slow down as the infection spreads (see Isham (1988) for a review of mathematical models for the AIDS epidemic). Those who are at highest risk will tend to be infected early so that the epidemic will then progress more slowly amongst those who are left. The pool of those susceptible to the infection is likely gradually to diminish and behavioural changes too, will have an effect. The quadratic exponential function defined by equation (3)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

is a mathematically convenient way of representing a curve that, for low values of t, increases exponentially but has a slower growth rate as t increases. The linear logistic function given by equation (4) also increases exponentially for low values of t, but becomes more nearly linear for higher values of t. This curve corresponds approximately, for moderate values of t, to the solution of a fairly simple epidemic model (see Isham 1988, section 4). We do not assume that either of the functions represented by equations (3) and (4) is appropriate for arbitrarily high values of t but only over the range of values for which the form of h is to be deduced.

Two flexible parametric families of distributions have often been used to model the incubation period; the gamma distribution, denoted by $\Gamma(\alpha, \lambda)$, has density

$$f(t) = \lambda(\lambda t)^{\alpha-1} \exp(-\lambda t)/\Gamma(\alpha) \quad (t \geqslant 0), \tag{5}$$

whereas the Weibull distribution, denoted Wei (β, ρ), has density

$$f(t) = \beta\rho(\rho t)^{\beta-1} \exp[-(\rho t)^{\beta}] \quad (t \geqslant 0). \tag{6}$$

See, for example, Lui *et al.* (1986); Blythe & Anderson (1988); Medley *et al.* (1987); Anderson & Medley (1988); Giesecke *et al.* (1988) and Lui *et al.* (1988) for discussion on modelling the incubation period distribution.

If $a(t)$ is given by equation (3) then

$$\tilde{a}(s) = a_0(\pi/a_2)^{\frac{1}{2}} \exp[(a_1 + is)^2/(4a_2)]. \tag{7}$$

The gamma distribution (equation (5)) also has a simple Fourier transform and it is straightforward to check that in this case $h(t)$ satisfies

$$h(t) = a_0 \exp[a_1^2/(4a_2)][1 - (2a_2/\lambda)\, \mathrm{D}_{a_1}]^{\alpha} \exp[-(a_1 - 2a_2 t)^2/(4a_2)], \tag{8}$$

where D_{a_1} is the partial differential operator, with respect to a_1. In particular we have

$$h(t)/a(t) = \begin{cases} x(t) & \alpha = 1 \\ x^2(t) - 2a_2/\lambda^2 & \alpha = 2 \\ x^3(t) - 6a_2 x(t)/\lambda^2 & \alpha = 3 \\ x^4(t) - 12a_2 x^2(t)/\lambda^2 + 12a_2^2/\lambda^4 & \alpha = 4 \end{cases} \tag{9}$$

where $x(t) = 1 + (a_1 - 2a_2 t)/\lambda$, $x^n(t) \equiv [x(t)]^n$ and $a(t)$ is given in equation (3).

If we choose to use the Weibull distribution (6) for the incubation periods, then analytic determination of $h(t)$ by using equation (2) is not feasible and we proceed numerically. A simple method that can be used if the index β of the Weibull distribution is a (small) integer is the following, which we describe for the case $\beta = 2$. First, write equation (1) in the alternative form

$$a(t) = \int_{-\infty}^{t} h(u) f(t-u)\, \mathrm{d}u. \tag{10}$$

On differentiating twice with respect to t, we obtain the integral equation

$$a''(t) = h(t) f'(0) + \int_{-\infty}^{t} h(u) f''(t-u)\, \mathrm{d}u, \tag{11}$$

where $a'(t) = \mathrm{d}a(t)/\mathrm{d}t$, $a''(t) = \mathrm{d}^2 a(t)/\mathrm{d}t^2$ and so on. Since for $\beta = 2$, $f'(0) = 2\rho^2 \neq 0$, equation (11) is a Volterra equation of the second kind that can be solved by a process of successive

approximation, yielding a numerical evaluation of $h(t)$ for known functions a and f (Tricomi 1957, chapter 1). This method is widely applicable and is not restricted to the particular functions given by equations (3) or (4), together with equation (6) with which we are concerned here.

3. Numerical results

The functions $f(t)$ and $a(t)$ are the rates at which infections with HIV and diagnoses of AIDS occur. Thus the expected numbers of infections and diagnoses in any particular time period can be obtained by integrating these rates appropriately. We shall take time measured in units of one year, with $t = 0$ corresponding to the start of 1979, and denote by $H_i(A_i)$ the expected incidences of infections (diagnoses) in year i. Thus

$$H_i = \int_{i-1979}^{i-1978} h(u)\,\mathrm{d}u, \tag{12}$$

and similarly for A_i.

The functions $a(t)$ given by equations (3) and (4) have been fitted by Cox & Medley (1989) to AIDS diagnosis data for the United Kingdom reported up to June 1988. The parameter estimates obtained are as follows:

equation (3) quadratic exponential:

$$a_0 = 0.08073 \quad a_1 = 1.6447 \quad a_2 = 0.06558,$$

equation (4) linear logistic:

$$b_0 = 77.645 \quad b_1 = 176.117 \quad b_2 = 6.871 \quad b_3 = 0.8233.$$

It must be emphasized that the standard errors attached to these parameter estimates are relatively large and that the fitted values A_i are sensitive to small changes in the estimates, especially in a_2 (Cox & Medley 1989).

For the years 1980–1993, the expected numbers of diagnoses, A_i, corresponding to these curves are tabulated in table 1. Over the period 1980–1987 the values of A_i for the linear logistic curve are very close to those for the quadratic exponential curve. After 1987, the approach to linearity of the former values starts to appear in contrast to the latter values that peak and then begin to decrease.

Table 1. Expected annual incidence of AIDS diagnoses, A_i

year i	quadratic exponential	linear logistic
1980	1	1
1981	4	5
1982	12	13
1983	36	37
1984	97	94
1985	227	225
1986	465	470
1987	838	840
1988	1327	1261
1989	1845	1644
1990	2252	1954
1991	2416	2206
1992	2276	2420
1993	1883	2614

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

In each case the rate of diagnoses $a(t)$ is given by equations (3) or (4); the parameter values are as given earlier; A_i is obtained by integrating $a(t)$ over the appropriate year.

Anderson & Medley (1988) fitted both the gamma and Weibull distributions, given by equations (5) and (6) respectively, for the incubation period using data for transfusion recipients available up to April 1988, together with a parametric model for the rate at which infected transfusions took place, and obtained the parameter estimates; $\Gamma(\alpha, \lambda)$, $\alpha = 2.70$, $\lambda = 0.19$, (mean 14.3); Wei (β, ρ), $\beta = 2.33$, $\rho = 0.12$, (mean 7.4).

For numerical convenience we shall compare expected annual incidence of HIV infections, H_i, assuming the incubation periods have either the $\Gamma(2, 0.14)$, $\Gamma(3, 0.21)$ or Wei $(2, 0.12)$ distributions that have means 14.3, 14.3 or 7.4 respectively. The Weibull and gamma distributions with index 2 will be used to calculate values of H_i with both forms for $a(t)$. Results using the gamma distribution with index 3 will only be given when $a(t)$ has the quadratic exponential form. Values of H_i for all these cases are given in table 2.

TABLE 2. EXPECTED ANNUAL INCIDENCE OF HIV INFECTIONS, H_i

$f(t)\ldots$	$\Gamma(3, 0.21)$	$\Gamma(2, 0.14)$		Wei $(2, 0.12)$	
$a(t)\ldots$	quadratic exponential	quadratic exponential	linear logistic	quadratic exponential	linear logistic
year, i					
1980	365	106	123	62	71
1981	1083	348	347	201	196
1982	2741	980	912	556	514
1983	5830	2364	2242	1320	1255
1984	10273	4851	4935	2652	2725
1985	14544	8366	8918	4455	4784
1986	15488	11877	11592	6110	5845
1987	9945	13312	9492	6516	4188
1988	−2076	10546	4940	4726	1822
1989	−15499	3198	2685	895	1462
1990	−22669	−6226	2830	−3320	2470
1991	−19244	−13394	3649	−5674	3464
1992	−8016	−15036	4351	−4997	4029
1993	3489	−11208	4836	−2101	4265
total 1980–1987 inclusive	60269	42204	38561	21872	19578

There are a number of points to note. Firstly, negative values of H_i are obtained in later years when the quadratic exponential function is used for $a(t)$, which means that this function is not compatible with the various assumed incubation period distributions over the whole 1980–1993 time period. Essentially, in such cases, if the early values of $h(t)$ are determined to give the quadratic exponential function $a(t)$ for low values of t, then too many AIDS diagnoses will occur later. To compensate for these, negative numbers of infections are then needed if the function $a(t)$ is to have the chosen increasing doubling time. When $a(t)$ has the linear logistic form, no negative values of H_i are obtained over the 1980–1993 time period, although there is still an implausible oscillation in the later values. However, the function $a(t)$ has been fitted using data only up to June 1988 and only the lower tail of the incubation period distribution curve can be fitted. This lack of compatibility is not therefore surprising. On the other hand, the figures for expected HIV incidence up to 1986 might be hoped to be reasonably reliable. Note also that the similarity between the fitted values for the two forms of $a(t)$ up to 1987 results in a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

corresponding similarity in the values of H_i up to 1986 (using a particular distribution for the incubation periods).

Secondly, with the assumption of the gamma distribution, $\Gamma(2, 0.14)$ or $\Gamma(3, 0.21)$, only 41 % or 35 % respectively of incubation periods will be of length 10 years or less, as compared with 76 % for the Weibull distribution. Thus the annual incidence of HIV infection with either of the gamma distributions must be much higher than that using the Weibull distribution during the early stages of an epidemic, to produce the same function $a(t)$. If we compare the total numbers of HIV infections occurring during 1980–1987 given in table 2, we see that the totals for the gamma distributions are almost double, or more, those for the Weibull distribution.

Thirdly, it is of interest to note that the use of an exponential curve for $a(t)$ with no quadratic term, together with the Weibull distribution for the incubation period, results in a total of some 22000 HIV infections over the years 1980–1987. This number is very similar to those obtained using the quadratic exponential or linear logistic curves although the numbers in individual years follow a different pattern.

4. The CDC progression rates

In a paper by the Centers for Disease Control (1987), estimates of progression rates to AIDS are given based on data for a subgroup of HIV-positive men taking part in the San Francisco City Clinic Cohort Study. (There is apparently an error in the rates quoted that has been corrected in rates given by Curran *et al.* (1988).) Thus if p_i is the probability that an infected individual develops AIDS between i and $i+1$ years after infection, estimates of p_i, $i = 0, \ldots, 9$ are as given in table 3.

Table 3. Probabilities of progression from HIV infection to AIDS

(p_i, CDC estimate of the probability of developing AIDS between i and $i+1$ years after infection; F_i, probability of incubation period of length between i and $i+1$ years, for Wei (2, 0.12).)

years after infection, i	0	1	2	3	4	5	6	7	8	9	total $i = 0, \ldots, 9$
$100p_i$	0	2	3	5	5	9	6	6	6	5	47
$100F_i$	1	4	7	8	10	10	10	10	9	7	76
$100 \times (47/76)F_i$	1	2	4	5	6	6	6	6	6	4	

It follows that an estimated 47 % of all those infected will have had AIDS diagosed within 10 years of infection. Note that these probabilities are not conditional upon the event that an infected individual does eventually have AIDS diagnosed. If this latter event has probability p, and if F_i is the probability that an incubation period has a length between i and $i+1$ years, then $p_i = pF_i$. For comparison, the values of F_i for the Wei (2, 0.12) distribution are also given in table 3, rounded to the nearest hundredth.

If p were to be $47/76 \approx 0.6$, then the values of pF_i for this Weibull distribution, again given to the nearest 1 % in table 3, are broadly similar to the CDC values p_i.

A discrete time approximation to equation (1) by using a year as the unit of time is

$$A_i = \sum_{j=0}^{\infty} H_{i-j} F_j = (1/p) \sum_{j=0}^{\infty} H_{i-j} p_j. \tag{13}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

For a given set of values $\{A_i\}$ with the year i going back to the start of the epidemic, and a set of p_i values as in table 3, equation (13) can be used to deduce the corresponding values of H_i/p; these are estimates of the annual incidence of all infections with HIV regardless of whether or not these lead to subsequent diagnoses of AIDS. For example, using the annual AIDS incidence A_i for the quadratic exponential model given in table 1 (the corresponding A_{1979} is 0) with the CDC values p_i from table 3, we obtain the values of H_i/p listed in table 4.

TABLE 4. EXPECTED ANNUAL INCIDENCE OF ALL HIV INFECTIONS, H_i/p, USING THE QUADRATIC EXPONENTIAL MODEL AND THE CDC PROGRESSION PROBABILITIES

year, i	1979	1980	1981	1982	1983	1984	1985	1986	1987
H_i/p	50	125	288	931	2197	4295	7168	9339	9640
$(47/76)H_i/p$	31	77	178	576	1359	2656	4433	5775	5962

Note that values of H_i/p for i after 1987 cannot be deduced until estimates of p_i for $i > 9$ are available. Also, as is to be expected given the similarities noted in table 4, scaling the H_i/p values by the factor 47/76 gives a set of values that closely resemble the H_i values obtained in table 2 using the Weibull distribution, again with the quadratic exponential function $a(t)$. Use of the linear logistic curve in place of the quadratic exponential curve in the above calculations gives very similar results over this time period, since the values of A_i up to 1987 are almost identical for the two curves.

5. DISCUSSION

The analytic forms for the rate, $a(t)$, of AIDS diagnoses and the density, $f(t)$, of the incubation period have been chosen and their parameters fitted using data currently available and therefore necessarily lying within restricted ranges of values of t. In particular, information is only available about the shape of the lower part of the distribution curve for the incubation period. Inevitably, then, the predicted values of the rate, $h(t)$, of HIV infection (or the annual incidence, H_i, of HIV infection) obtained by using values of $a(t)$ or $f(t)$ for t lying outside these ranges must be treated with great caution. Even up to 1987 the projected values of H_i vary considerably with the particular incubation period distribution used.

I have already noted the broad similarity in shape between the estimated probabilities p_i given by CDC (1987) and the corresponding discretized Weibull distribution $\{F_i\}$ given in table 3. If one were to believe that these sets of probabilities are approximately correct then one would be led to an estimate of around 0.6 for the probability p that an HIV infected individual will subsequently be diagnosed as having AIDS. It appears from table 2 that, assuming the Weibull distribution for incubation periods, the total number of HIV infections in the United Kingdom between 1980 and 1987 inclusive, which will lead to subsequent diagnoses of AIDS, is perhaps of the order of 20–22 000. If we include all HIV infections, not just those leading to AIDS diagnoses, the corresponding range could be 33–37 000 (if the entries in table 2 are scaled by a factor of 10/6). To be more specific it will be necessary to examine further the relative merits of the parametric forms for $a(t)$ and $f(t)$ that have been used.

It is self-evident that projections of this sort are entirely dependent on the assumptions being made. If the functions a and f were known then the HIV infection rate h could be determined exactly from equation (1). Error in the assumed forms of a or f will be reflected in errors in the estimated annual incidence of HIV infection, H_i. In some cases the assumed forms will be clearly

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

incompatible, at least over part of their ranges, as is apparent from the negative values of H_i given in table 2 which obtain when $a(t)$ has the quadratic exponential form. The linear logistic curve for $a(t)$ gives more satisfactory results although, as remarked earlier, the oscillation in H_i does not seem entirely plausible.

One source of error that has been suggested is that whereas the population being considered is the whole population of the United Kingdom with $a(t)$ being fitted to all AIDS diagnoses in those countries, the incubation period distributions fitted by Anderson & Medley (1988) were obtained by using data on recipients of blood transfusions. The CDC probabilities refer to those (presumably) infected by homosexual transmission. It might be expected that the distribution of the length of the incubation period would vary with the mode of transmission. That $a(t)$ is the total rate of AIDS diagnoses summed over all risk groups is not a problem in the method described in this paper for inferring the rate of HIV infection, as long as the density f of incubation periods is an appropriately weighted mixture of distributions applying to each risk group separately. In particular, the fitted Weibull distribution could be inappropriate for the majority of AIDS cases that have occurred so far. The weights used in combining the distributions might need to change with time as the infection starts to spread into different subgroups of the population (grouped for example by risk or by spatial considerations). Were the necessary data available, back projection of HIV incidence for each subgroup separately would avoid this problem. However, recent work by Anderson & Medley (1988) and Lui *et al.* (1988) suggests that the mode of transmission of infection does not have a strong effect on the distribution of the incubation period and so the assumption of a common incubation period distribution is not too unreasonable as a first approximation. There is also evidence (Darby *et al.* 1989) that the incubation period distribution does vary with the age of the infected individual, but again this is not a problem as long as f is the appropriately weighted mixture of such distributions.

I am grateful to Roy Anderson, F.R.S., David Cox, F.R.S., and Graham Medley for providing me with prepublication copies of their papers and for helpful discussion.

References

Anderson, R. M. & Medley, G. F. 1988 Epidemiology, HIV infection and AIDS: the incubation and infectious periods, survival and vertical transmission. *AIDS* **2**, S57–S63.

Blythe, S. P. & Anderson, R. M. 1988 Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (HIV). *IMA J. Math. appl. med. Biol.* **5**, 1–19.

Boldsen, J. L., Jensen, J. L., Søgaard, J. & Sorensen, M. 1988 On the incubation time distribution and the Danish AIDS data. *J. R. statist. Soc.* A **151**, 42–43.

Brookmeyer, R. & Gail, M. H. 1986 The minimum size of the AIDS epidemic in the United States. Lancet(ii), 1320–1322.

Brookmeyer, R. & Gail, M. H. 1988 A method for obtaining short term projections and lower bounds on the size of the AIDS epidemic. *J. Am. statist. Ass.* **83**, 301–308.

Centers for Disease Control 1987 Human immunodeficiency virus infection in the United States: a review of current knowledge. *MMWR* **36** (suppl. no. S-6), 1–48.

Cox, D. R. & Medley, G. F. 1989 A process of events with notification delay and the forecasting of AIDS. *Phil. Trans. R. Soc. Lond.* B **325**, 135–145. (This issue.)

Curran, J. W., Jaffe, H. W., Hardy, A. M., Meade Morgan, W., Selik, R. M. & Dondero, T. J. 1988 Epidemiology of HIV infection and AIDS in the United States. *Science, Wash.* **239**, 610–616.

Darby, S. C., Rizza, C. R., Doll, R., Spooner, R. J., Stratton, I. M. & Thakrar, B. 1989 Incidence of AIDS and excess mortality associated with HIV in haemophiliacs in the U.K. *Br. med. J.* **298**, 1064–1068.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Giesecke, J., Scalia-Tomba, G., Berglund, O., Berntorp, E., Schulman, S. & Stigendal, L. 1988 Incidence of symptoms and AIDS in 146 Swedish haemophiliacs and blood transfusion recipients infected with human immunodeficiency virus. *Br. med. J.* **297**, 99–102.

Isham, V. 1988 Mathematical modelling of the transmission dynamics of HIV infection and AIDS: a review. *J. R. statist. Soc.* A **151**, 5–49.

Iversen, O.-J. & Engen, S. 1986 Epidemiology of AIDS – statistical analyses. *J. Epidemiol. Commnty Hlth*, **41**, 55–58.

Lui, K-J., Lawrence, D. N.,, Morgan, W. M., Peterman, T. A., Haverkos, H. W. & Bregman, D. J. 1986 A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome. *Proc. natn. Acad. Sci. U.S.A.*, **83**, 3051–3055.

Lui, K-J., Darrow, W. W. & Rutherford, G. W. III 1988 A model based estimate of the mean incubation period for AIDS in homosexual men. *Science, Wash.* **240**, 1333–1335.

Medley, G. F., Anderson, R. M., Cox, D. R. & Billard, L. 1987 Incubation period of AIDS in patients infected via blood transfusion. *Nature, Lond.* **328**, 718–721.

Rees, M. 1987 The sombre view of AIDS. *Nature, Lond.* **326**, 343–345.

Tricomi, F. G. 1957 *Integral equations*. New York: Wiley Interscience.

Note added in proof – 22 *May* 1989

The recent issue of *Statistics in Medicine* (1989, **8**, part 1) devoted entirely to AIDS modelling, contains several papers describing the use of the back-projection method in connection with the AIDS epidemic in the U.S.A. In particular, Brookmeyer & Damiano (1989) extend earlier work by Brookmeyer & Gail (1988), and Taylor (1989) uses a similar approach. De Gruttola & Lagakos discuss problems arising from uncertainty about the appropriate form for $h(t)$.

References

Brookmeyer, R. & Damiano, A. 1989 Statistical methods for short-term predictions of the AIDS incidence. *Statist. Med.* **8**, 23–34.

De Gruttola, V. & Lagakos, S. W. 1989 The value of AIDS incidence data in assessing the spread of HIV infection. *Statist. Med.* **8**, 35–44.

Taylor, J. H. G. 1989 Models for the HIV infection and AIDS epidemic in the United States. *Statist. Med.* **8**, 45–58.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES