PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Predictions of the AIDS Epidemic in the U.K.: The Use of the Back Projection Method

N. E. Day, S. M. Gore, M. A. McGee and M. South

Phil. Trans. R. Soc. Lond. B 1989 **325**, 123-134
doi: 10.1098/rstb.1989.0077

Phil. Trans. R. Soc. Lond. B **325**, 123–134 (1989) [123]
Printed in Great Britain

PREDICTIONS OF THE AIDS EPIDEMIC IN THE U.K.: THE USE OF THE BACK PROJECTION METHOD

By N. E. DAY, S. M. GORE, M. A. McGÈE and M. SOUTH
MRC Biostatistics Unit, 5 Shaftesbury Road, Cambridge CB2 2BW, U.K.

Back projection methods are used to predict the yearly number of new AIDS diagnoses and the number of new HIV infections, to the end of 1992. The AIDS, but not the HIV, predictions are insensitive to the choice of incubation period distribution. A wide range of predictions is consistent with the AIDS diagnoses in years up to 1987, but limited ancillary information on the relative rates of new HIV infection in 1984 to 1987 can be used to narrow this range. The range of prediction based on AIDS reports to the end of 1988 is lower and narrower than the range based on reports to the end of 1987. The number of new AIDS cases in 1992 appears likely to fall in the range 1000–3000.

1. Introduction

Accuracy in predicting the future of the AIDS epidemic depends on the extent of the information on which the predictions are based. Methods range from those that simply use the number of reported AIDS cases, to those that model the full dynamics of transmission in the population. The former are arithmetic exercises in extrapolation ignoring whatever is known of the epidemiology of the disease, and cannot incorporate information that one might have on changing patterns of transmission. They are reliable in the very short term, but are also likely to be overtaken by events; even in the medium term, the predictions may prove wildly inaccurate. Methods that model the full dynamics of the epidemic can provide much insight into the qualitative evolution of the epidemic, and identify the key variables that determine in the short and medium terms the number of cases that occur. Accurate information on these variables is, however, sparse and very difficult to obtain. Intermediate between these two approaches to prediction is the back projection method (Brookmeyer & Davison 1988; Zeger *et al.* 1989). The advantage of this method for short and medium term projections is that it can utilize the available information of particular importance for predicting the number of AIDS cases, namely the number of new HIV infections up to the present. Precise information on HIV infection is lacking, and a major epidemiological challenge is to construct ways of acquiring it in the future. Some data are available and these can be used to assess the credibility of different possible projections.

Back projection is based on the underlying relation between the number of new cases of AIDS in time t to $t+\mathrm{d}t$, which we designate $a(t)$, and the number of new HIV infections $h(s)$ at each time s since the start of the epidemic $(s=0)$ through the incubation period distribution $f(u)$, u being the time spent between initial infection and diagnosis of AIDS:

$$a(t) = \int_0^t h(s)\,f(t-s)\,\mathrm{d}s. \qquad (1)$$

If one knew $f(u)$, then this relation could be inverted to express $h(s)$ for all $0 \leqslant s \leqslant t$

as a function of $a(s)$, $0 \leqslant s \leqslant t$. More generally we can construct a family of values for $h(s)$, $0 \leqslant s \leqslant t$, which are consistent with a particular realization of $a(s)$, $0 \leqslant s \leqslant t$.

For practical purposes, the time axis will be grouped into intervals $(t_0(=0), t_1)$, (t_1, t_2) to (t_{n-1}, t_n). We designate by a_i and h_i the number of new AIDS cases and HIV infections respectively in the time interval (t_{i-1}, t_i), $i = 1, \ldots, n$.

We then have an approximate discrete version of relation (1),

$$a_i = \sum_{j=1}^{i} h_j f_{i-j}, \tag{2}$$

where
$$f_{i-j} = \int_{t_{i-j}}^{t_{i-j+1}} f(u)\,\mathrm{d}u,$$

or
$$\boldsymbol{a}_i = F_i \boldsymbol{h}_i,$$

where
$$\boldsymbol{a}_i = (a_1, \ldots, a_i), \boldsymbol{h}_i = (h_1, \ldots, h_i)$$

and F_i is the appropriate matrix.

The k,jth element of F_i will be given by

$$\begin{aligned} f_{kj} &= f_{k-j} \quad (k \geqslant j) \\ &= 0 \quad\quad (k < j). \end{aligned}$$

We then have
$$\boldsymbol{h}_i = F_i^{-1} \boldsymbol{a}_i$$

where F_i^{-1} has the same form as F_i, i.e.

$$\begin{aligned} f_{kj}^{(-1)} &= f_{k-j}^{(-1)} \quad (k \geqslant j) \\ &= 0 \quad\quad (k < j) \end{aligned}$$

with the $f_{k-j}^{(-1)}$ given straight forwardly by

$$f_0 f_k^{(-1)} = -\sum_{m=1}^{k} f_m f_{k-m}^{(-1)} \quad (k > 0)$$

and
$$f_0 f_0^{(-1)} = 1.$$

2. Materials and methods

We assume a certain form for the incubation period distribution $f(t)$. Then on the basis of observed values of $a(t)$ up to the present, we calculate the range of values of $h(t)$ consistent with the observed $a(t)$. This range of values of $h(t)$ is then examined in the light of available knowledge on HIV infection in the population, limited and mostly indirect, but nevertheless sufficient to indicate that some of the range of values for $h(t)$ are implausible.

Predictions of future AIDS cases are made by projecting forwards from the HIV estimates, by using the incubation period distribution. Assumptions have to be made regarding future HIV infections. The importance of these assumptions for future AIDS cases is examined, and their plausibility assessed in terms of the weak available data on trends in HIV infection.

Available data

(i) *Number of AIDS cases*

Our analysis is based on AIDS cases among U.K. residents. The biologically meaningful measure for the number of AIDS cases is the number of new diagnoses, in suitably chosen time intervals. The monthly published figures refer to reports, not diagnoses, and the delay between the two is variable. The Communicable Disease Surveillance Centre (CDSC) has made available a tape giving information on all reports to 30 June 1988 that includes date of diagnosis (table 1*a*). The number of reports, with no further information, is available for each month to the end of 1988 (table 1*c*). By using data on reports up to the end of 1987, and assuming an exponential increase in the epidemic for 1988 and 1989, the figures in column 1 of table 2 for the number of AIDS diagnoses in each year from 1982 to 1987 are produced (Healy 1988). We refer to these as the earlier Healy estimates for 1987. Data on reports to the end of 1988 make these estimates untenable (table 1*a* and 1*c*).

TABLE 1. YEARLY REPORTS AND DIAGNOSES OF AIDS

(*a*) Year of diagnosis and year of report for AIDS cases reported[a] to mid 1988.

		year of diagnosis							
year of report	total	1982	1983	1984	1985	1986	1987	1988	unknown
1982	2	2	—	—	—	—	—	—	0
1983	26[b]	4	19	—	—	—	—	—	0
1984	76[c]	3	9	61	—	—	—	—	2
1985	156	1	1	29	117	—	—	—	8
1986	298	1	0	10	63	216	—	—	8
1987	612	0	3	1	40	165	402	—	1
Jan.–Jun. 1988	362	0	0	1	2	23	129	202	5
total	1532	11	32	102	222	404	531	202	24
Imputed year of diagnosis for the 24 unknowns		—	—	6	10	4	3	1	—

(*b*) Distribution of reporting delays for the 362 cases reported in the first half of 1988.

report delay (months)	0	1–3	4–6	7–9	10–12	13–18	19–24	25–29
probability	0.194	0.497	0.126	0.070	0.020	0.045	0.037	0.011

(*c*) Empirical assignation of year of diagnosis, for cases reported from July 1988.

		Empirically assigned year of diagnosis		
year of report	total	1986	1987	1988
Jly–Dec. 1988	393	11.6	39.4	341.9
Jan.–Jun. 1989[d]	419	2.5	26.8	146.5
Jly–Dec. 1989[d]	445	—	13.0	44.5
Jan.–Jun. 1990[d]	471	—	2.8	30.1
Jly–Dec. 1990[d]	497	—	—	14.6
total	2225	14.1	82.0	577.6

[a] Excludes visitors to U.K.
[b] Includes three diagnosed in 1981.
[c] Includes one diagnosed in 1979.
[d] Imputation based on linear extrapolation of quarterly report totals for 1987 and 1988, namely 137, 138, 182, 155, 199, 163, 199 and 194.

We take a different approach. For the 18 months to the end of 1988, the number of new AIDS reports per quarter has remained nearly constant. The number of new reports in each quarter of 1989 and 1990 has been estimated by linear extrapolation from the results for July 1987 to December 1988. The observed delay distribution for cases reported in the first 6 months of 1988 is given in table 1*b*. The distribution is truncated at 30 months.

Applying the backward delay distribution of table 1*b* to these extrapolated figures yields the year of diagnosis as given in table 1*c*. We have also allocated proportionally a year of diagnosis for the 24 cases in table 1*a* for which year of report diagnosis was unknown. The final figures for the yearly number of AIDS diagnoses that we have used in the remainder of the paper are given in the second column of table 2; they incorporate the further data from CDSC to the end of 1988. Up to 1986 the first two columns of table 2 are, of course, nearly identical. The second column figure for 1987 depends little on our assumption for reports in 1989 and seems likely to be near the truth. The figure for 1988 is more speculative. Our principal predictions for AIDS cases per year up to 1992 are based on our current estimates of yearly AIDS diagnoses from 1982 to 1987 inclusive.

TABLE 2. ESTIMATED NUMBER OF AIDS DIAGNOSES BY YEAR

year	earlier estimate, based on reports to Dec. 1987	current estimate, based on reports to Dec. 1988
1982	10	11
1983	30	32
1984	109	108
1985	224	232
1986	425	426
1987	762	610
1988	—	777

(ii) *Incubation period distribution*

Three different distributions have been used, two based on fitting Weibull and Gamma distributions to observed data and the third, which we call the CDC distribution, based on observed empirical distribution (table 3). In the period of interest, the CDC distribution is close to twice the gamma; half the CDC distribution is shown in table 3 and has been used in further computation. In fact predictions up to 1992, based on reports to the end of 1988, are not sensitive to changes in the incubation period distribution. (The extent to which predictions further into the future are affected by this distribution is the subject of a separate paper.) Without this lack of sensitivity in short-term AIDS forecasts the method would be questionable since the progression of HIV disease is known to be heterogeneous across different subgroups of the population.

(iii) *Assumptions about the number of* HIV *infecteds in the population, by year*

Our aim is to explore a range of probabilistic evolutions of the number of cases of HIV infection over time that are consistent with, (i) the number of AIDS cases reported to the end of 1988, from which is derived the number of AIDS diagnoses to the end of 1987, and (ii) available data on HIV prevalence in previous years, albeit incomplete and from selected subsamples.

TABLE 3. PROGRESSION RATES FOR DIFFERENT INCUBATION PERIOD DISTRIBUTIONS

new percentage (cumulative %) progressing to AIDS after t years

incubation distribution, t =	1	2	3	4	5	6	7	8	9	10	11	12	total at 10 years
Gamma	0.3	0.9 (1.2)	1.6 (2.8)	2.2 (5.0)	2.7 (7.7)	3.0 (10.7)	3.3 (14.0)	3.5 (17.5)	3.6 (21.1)	3.7 (24.8)	3.7 (28.5)	3.7 (32.2)	24.8
$\frac{1}{2}\times$ CDC rates (extrapolated from Gamma)	0	1.0 (1.0)	1.5 (2.5)	2.5 (5.0)	2.5 (7.5)	4.5 (12.0)	3.0 (15.0)	3.0 (18.0)	3.0 (21.0)	2.5 (23.5)	5.0 (28.5)	3.7 (32.2)	23.5
CDC rates (extrapolated from $2\times$ Gamma)	0	2.0 (2.0)	3.0 (5.0)	5.0 (10.0)	5.0 (15.0)	9.0 (24.0)	6.0 (30.0)	6.0 (36.0)	6.0 (42.0)	5.0 (47.0)	10.0 (57.0)	7.4 (64.4)	47.0
Weibull	0.3	2.8 (3.1)	5.4 (8.5)	7.5 (16.0)	9.0 (25.0)	9.9 (34.9)	10.2 (45.1)	9.9 (55.0)	9.1 (64.1)	8.1 (72.2)	6.8 (79.0)	5.6 (84.6)	72.2

We have assumed that $h(t)$, the number of new infections in time t to $t+\mathrm{d}t$, can be described as

$$h(t) = a \exp(bt+ct^2), \text{ with } t = 0 \text{ at January 1981.}$$

In expression (2), we have taken as an approximation to h_i the weighted sum: $\frac{1}{4}h(i-1)+\frac{1}{2}h(i-\frac{1}{2})+\frac{1}{4}h(i)$. This functional form produces some improbable extrapolations to years beyond 1990; we have used it as a means of exploring different forms of the epidemic of HIV infection up to the present. In particular, we examined continuing exponential increase in the number of new cases of HIV infection per year in comparison to a slowing of the rate of increase, and a maximum reached somewhere between 1984 and 1988. There is considerable evidence that the risk behaviour of homosexual men has changed; changes, which may have begun in 1983, were widespread in 1985 and continued in 1986 and 1987 (Johnson & Gill 1988). This suggests that large increases in the yearly number of HIV-positives did not occur from 1985 to 1988, and that there may well have been a fall.

(iv) *Model fitting*

For each assumed incubation period distribution, by computation over a three dimensional grid, the range of values of a, b and c was determined for which the predicted number of AIDS cases in years 1982 to 1987 were consistent with the observed, using as a consistency measure a goodness of fit X^2 on 6 degrees of freedom (90% point). Additional fits were made including the preliminary estimates of 1988 AIDS diagnoses, and also to the numbers (column 1 of table 2) assumed in earlier reports (referred to as the earlier estimates).

For each set of values for a, b and c, the expected number of AIDS cases, and new HIV infections, were calculated for each year to 1992. The set of values of a, b and c consistent with the 1982–87 AIDS data generates a range of values for predicted numbers, both of new cases in 1992 and the cumulative number of cases up to the end of 1992.

Many of the values for a, b and c that generate expected AIDS cases consistent with the data to the end of 1987 generate implausible numbers of new HIV infecteds in the years 1985 to 1988, implausible either because they show a rapid rise or too rapid a decline. The crucial value in this respect is given by the time to peak HIV incidence, t_p, where $t_p = -b/2c$. If t_p is less than 3, there are unrealistically few new infections in 1987 and 1988; if t_p is greater than 8, then 1988 new infections are largely in excess of those of 1985.

3. Results

The incubation period distributions shown in table 3 range from less than 25% progression in 10 years for the gamma and half-CDC to nearly 75% for the Weibull. The CDC distribution is intermediate, and predictions were initially based on it. By using the current estimates of yearly AIDS diagnoses to the end of 1987, the range of predicted AIDS cases in 1992 is from 737 to 3538. The parameter values that best fit the data to 1987 predict 1031 AIDS cases in 1992. Table 4 gives the predictions by year, for both new AIDS cases and new HIV infections. Both the extreme low and the extreme high seem unlikely. For the former, it is very improbable that so few new infections would have occurred in 1987 and 1988, and none in later years. It is interesting, however, that such results are consistent with the current data on AIDS cases. The 'extreme high' model proposes a doubling of the number of new HIV infecteds between 1985 and 1988; this conflicts with what is known of changing behaviour.

Table 4. Predicted new aids cases and new hiv infecteds by year to 1992

(aids cases, current estimates for 1987; assumed incubation distribution, cdc rates.)

	1987 current estimate	aids cases predictions to 1992 best fit	lowest	highest	hiv infecteds predictions to 1992 best fit	lowest	highest
1981	—	—	—	—	410	307	724
1982	11	8	6	14	1233	1349	1212
1983	32	37	36	46	2575	3097	1935
1984	108	109	118	111	3732	3727	2945
1985	232	234	250	214	3756	2353	4274
1986	426	414	409	396	2625	778	5914
1987	610	616	567	643	1273	134	7802
1988	—	809	708	985	428	12	9813
1989	—	955	781	1437	100	1	11767
1990	—	1021	763	2002	16	0	13453
1991	—	1031	726	2717	2	0	14664
1992	—	1031	737	3538	0	0	15239
predicted total to 1992		6265	5101	12103	16150	11758	89742
t_p = time to peak hiv incidence; $t = 0$ = Jan. 1981					4.02	3.29	11.81

Table 5 compares predictions based on the cdc incubation period distribution to those based on the Weibull (faster progression) and on the half-cdc (slower progression). There is a major difference in the predicted numbers of hiv infecteds, the approximately 3-fold difference between prediction based on the Weibull and the half-cdc reflecting the similar proportional difference between the two distributions, up to 12 years. Both the best estimate and the range of predictions for the yearly number of aids cases to 1992 are similar for the three distributions.

The range of predictions for the number of new aids cases in 1992 is related to the different forms of the hiv epidemic model from which the aids cases are derived, specifically in terms of the time t_p at which the hiv epidemic is assumed to have peaked. This relation is shown in table 6*a*; table 7 illustrates the form that the hiv epidemic is assumed to take for a range of values of t_p. This table demonstrates three points: first, that knowing the number of new infections, by year, greatly improves the precision of the predictions of new hiv cases; second, that crude and unsystematic information, as described by Johnson & Gill (1988), can be used at least to assign plausibilities to the different values in table 6; and third, that reliable information on new hiv infections would be valuable for estimating the incubation period distribution, even if all data were anonymous.

The values in table 7 are for the cdc distribution. Values generated by the Weibull and the half-cdc distribution are proportionately almost exactly the same, but smaller or larger by a factor of 1.5 approximately in absolute value respectively. The form assumed for the course of the hiv epidemic has three parameters, of which t_p only fixes one. For given t_p, the range of values for the other two parameters, which are consistent with the observed aids data, is small.

Figure 1 displays graphically the information of tables 4–7. One can see the insensitivity of the 1992 aids prediction to the assumed form of the incubation period distribution, and the sensitivity of the hiv-infected estimates. The relation between the time point at which hiv infection is assumed to have peaked and the 1992 aids predictions is also clear.

In table 5, we also study the effect of accumulating additional information on aids diagnoses on short and medium term aids predictions. We consider the situation as it was in early 1988,

TABLE 5. CUMULATIVE HIV INFECTEDS BY THE END OF 1987 AND PREDICTED NEW AIDS CASES IN 1988 AND 1992

incubation distribution	observed AIDS data	best fit predictions 1981–1987 HIV	best fit predictions 1988 AIDS	best fit predictions 1992 AIDS	best fit predictions t_p	extremes consistent with data: low 1981–1987 HIV	low 1988 AIDS	low 1992 AIDS	low t_p	high 1981–1987 HIV	high 1988 AIDS	high 1992 AIDS	high t_p
$\frac{1}{2}$ CDC	current 1987	30911	805	1018	3.99	23325	703	735	3.17	49833	987	3413	10.51
CDC		15604	809	1031	4.02	11745	708	737	3.29	24806	985	3538	11.81
Weibull		10229	791	967	4.08	7752	671	683	3.38	16727	1023	3398	9.82
$\frac{1}{2}$ CDC	earlier 1987	58457	1166	2973	6.75	39217	966	1306	4.31	92418	1527	17928	α
CDC		29851	1178	3172	7.03	19406	956	1300	4.31	46308	1519	18610	α
Weibull		19323	1176	3071	7.07	13006	954	1276	4.43	29425	1550	20322	α
$\frac{1}{2}$ CDC	current 1988	29617	786	969	3.90	24444	733	763	3.32	36854	840	1514	5.31
CDC		15106	792	994	3.95	12221	729	764	3.35	18401	833	1548	5.45
Weibull		10229	791	967	4.08	8580	719	770	3.59	12130	832	1447	5.46

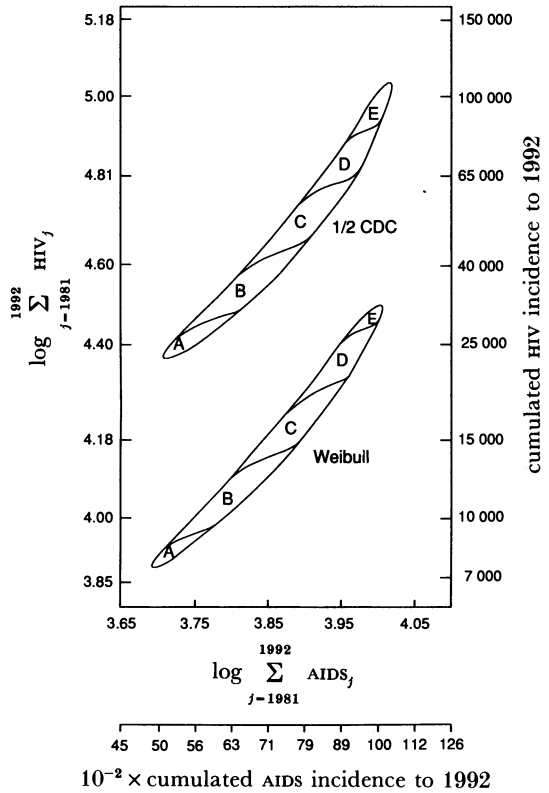

FIGURE 1. HIV incidence versus AIDS incidence: cumulated 1981–1992. Peak HIV incidence, pre 1984, A; 1984, B; 1985, C; after 1986, E.

TABLE 6. PREDICTED NEW AIDS CASES IN 1992: AS A FUNCTION OF PEAK ANNUAL HIV INCIDENCE DATE

(AIDS cases: current estimates for 1987.)

incubation distribution	lower/upper bound of predicted AIDS cases in 1992	time period in which incidence of HIV infection peaks: 1984	1985	1986	1987	1988	1989	1990
½ CDC	lower	735	958	1355	1707	2112	2369	2698
	upper	1087	1578	2153	2508	2928	3176	3389
CDC	lower	737	952	1353	1754	2078	2412	2658
	upper	1095	1627	2131	2604	2927	3147	3478
Weibull	lower	672	867	1241	1649	2030	2380	2942
	upper	992	1432	2026	2427	2797	3109	3398

as it is at the start of 1989 (on which the predictions of table 4 were based) and as it might be at the start of 1990. In early 1988, linear exponential extrapolation of the epidemic was still plausible, generating the estimates of AIDS diagnoses up to the end of 1987 as in column 1 of table 2. Predictions based on these estimates using Weibull, CDC and half-CDC models are given in table 5. They are markedly different from the predictions based on our 'current' estimates to the end of 1987.

In early 1990, we might expect the estimated number of AIDS diagnoses in 1988 to be 777

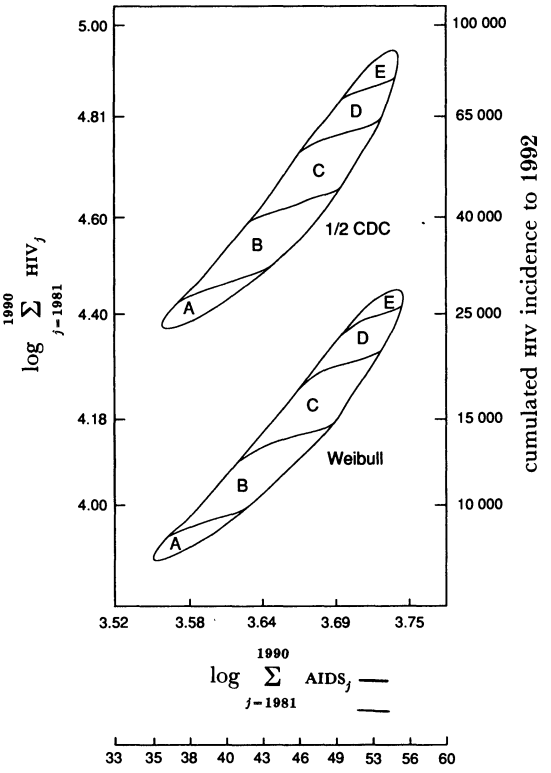

FIGURE 2. HIV incidence versus AIDS incidence: cumulated 1981–1990. Peak HIV incidence, pre 1984, A; 1984, B; 1985, C; 1986, D; after 1986, E.

TABLE 7. YEARLY NUMBER OF NEW HIV INFECTIONS: BY DIFFERENT PEAK ANNUAL HIV INCIDENCE DATE

(AIDS cases: current estimates for 1987; assumed incubation distribution: CDC rates.)

	peak annual HIV incidence date: Jan. 1981 + t_p where $t_p \approx$						
year	3.5	4.5	5.5	6.5	7.5	8.5	9.5
1981	364	493	595	541	592	643	670
1982	1349	1259	1211	1149	1145	1163	1176
1983	2960	2460	2106	2130	2001	1945	1933
1984	3846	3675	3125	3441	3158	3004	2975
1985	2962	4201	3959	4847	4505	4286	4285
1986	1351	3673	4282	5954	5808	5651	5777
1987	364	2457	3954	6377	6766	6882	7291
1988	58	1257	3116	5956	7122	7743	8614
1989	5	491	2097	4851	6775	8049	9526
1990	0	147	1205	3445	5824	7729	9861
1991	0	34	591	2133	4524	6857	9556
1992	0	6	247	1152	3176	5620	8669
$a\exp(bt+ct^2)$							
a	128	258	378	338	398	453	483
b	1.9638	1.248	0.8870	0.9063	0.7702	0.6791	0.6343
c	−0.2805	−0.1387	−0.0807	−0.0697	−0.0513	−0.0400	−0.0333

(table 2, column 2). Comparing predictions based on the 'best' estimates up to 1988 with those based on the 'current' estimates up to 1987, we can see a major reduction in the range, particularly at the upper end. (The predictions based on the best fitting model are virtually the same, because the 1988 estimate is close to the best fitting model using data up to 1987.) The main difference is that with the putative 1988 diagnoses, the peak in the epidemic of HIV infection can have occurred no later than mid-1986 ($t_p = 5.5$), whereas using only current 1987 data the peak could have occurred after 1988 i.e. new infections continually increasing in the period 1981 to 1988.

4. Discussion

The results of this analysis illustrate three points. First, that for predictions of AIDS cases four to five years into the future, the back projection method is largely insensitive to the assumption one makes for the incubation period distribution. The two extreme distributions considered represent the fast and slow extremes of incubation period distribution usually proposed; distributions that lie between these two give predictions within the range of predictions that the two generate. The estimated number of new HIV infections, however, is highly sensitive to the assumed incubation period distribution; prediction of AIDS cases in the long term will be similarly sensitive.

Second, that each prediction of AIDS cases within the range of consistent predictions corresponds to a particular form for the HIV epidemic. This correspondence enables one to use a variety of data, of varying degrees of reliability and direct relevance, to assess the plausibility of each prediction. This use of additional data is a major attraction of the back projection method, which would otherwise be just another form of extrapolation. Assuming a form for the HIV epidemic and for the incubation period distribution is equivalent to assuming a certain form for the yearly number of AIDS cases. Unless further information is used, the back projection method would be equivalent to straightforward extrapolation of the AIDS cases. In this straightforward extrapolation, however, there is no way to incorporate accumulating information on the incubation period, changing levels of transmission and rates of HIV infection. Dissection of the extrapolation process in the back projection method focuses attention on where this extra information can be used.

Third, that the AIDS reports in 1988 have made a major impact on short and medium term projections. The range of predictions is much narrower, and considerably lower, than it was based on data available at the start of 1988. Preliminary estimates for 1988 diagnoses, based on 1988 reports, would generate a range of only two fold for 1992 predictions, compared to the 15-fold range based on data to the end of 1987.

The functional form used for the evolving HIV epidemic covers, at least approximately, the plausible models of the epidemic up to the present. It is clear, however, that models that give a peak between 1985 and 1987, consistent with available information on changing homosexual behaviour, predict implausibly low numbers of HIV infecteds in the years 1989–92 (see table 7). However, even if one allowed the yearly number of HIV infecteds to plateau at, say, the 1987 level, few extra AIDS cases would be generated by 1992. Thus in table 5, the best fitting and low predictions should perhaps be increased by 100 to 150. One might also, as in the Cox report, add an extra 20% for underreporting. Even so, this would bring the best estimate of new AIDS cases in 1992 to at most 1500, and the upper bound on predictions to about 4200; if one limits the range to those consistent with the available data on HIV transmission and the number of HIV infecteds, then an upper bound of some 3000 looks appropriate. This range of uncertainty will be greatly reduced if the 1989 data follow current trends. By the end of 1989, the range of predictions for 1992 may well be approximately 1000–2000.

References

Brookmeyer, R. & Davison, A. 1989 Statistical methods for short term projections of AIDS incidence. *Statist. Med.* **8**, 23–34.

Department of Health Working Group 1988 Short term predictions of HIV infection and AIDS in England and Wales. London: HMSO.

Cox, D. R. 1988 Some remarks on the analysis of reporting delays. Appendix 8 of Department of Health Working Group Report.

Healy, M. J. R. 1988 Extrapolation forecasting. Appendix 6 of the Department of Health Working Group Report.

Johnson, A. M. & Gill, O. N. 1988 Evidence for recent changes in sexual behaviour in homosexual men in England and Wales. Appendix 5 of the Department of Health Working Group Report.

Zeger, S. L., See, L. C. & Diggle, P. J. 1989 Statistical methods for monitoring the AIDS epidemic. *Statist. Med.* **8**, 3–21.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES