

AIDS: predicting cases nationally and locally

B R Tennison, S Hagar

Abstract

Models for predicting the future course of the AIDS epidemic can be divided into five types: trend extrapolation models, compartment models, models based on the incubation period, comparison models, and models produced by expert committees. To predict the numbers of cases of AIDS in the United Kingdom and in East Anglia and Cambridge a two stage approach was chosen using trend extrapolation for the national case reports followed by reduction in scale to the two localities. The method predicted that about 2700 cases would be reported nationally during 1990 and about 6000 during 1992. The number of people with AIDS expected to present for treatment in East Anglia during 1990 was 48, and during 1992 was 105; for Cambridge the corresponding figures were 20 and 43.

These figures with their estimated 95% confidence intervals will be used for planning local services for people with AIDS, and they emphasise the need for preventive action.

Introduction

AIDS is a crucial public health problem; the recent emergence of the disease and its special features mean that predicting the future course of the epidemic presents many problems. Such predictions are needed at national and regional levels for health service planning; and at district level to put the local preventive programmes into perspective.

The objective of this work was to provide predictions of the number of people who would present with AIDS in the East Anglian region and the Cambridge district over the five years up to 1992. The East Anglian region consists of the counties of Cambridgeshire, Norfolk, and Suffolk, with a total population of 2 million; the Cambridge district contains the cities of Cambridge and Ely together with the surrounding area and has a total population of 270 000. The methods provide national predictions for the United Kingdom and also apply to other regions and districts. Predictive models currently in use for the AIDS epidemic fall into five categories: trend extrapolation models,^{1,4} compartment models,^{5,8} models based on the distribution of the incubation period,^{9,10} comparison models,¹¹ and expert committee models.

Trend extrapolation models take a version of the epidemic curve and extend it into the future. The curve is usually given by the number of cases of AIDS (cumulated or not) plotted against time, as data are sparse on the extent of infection with human immunodeficiency virus (HIV). The number of cases is usually derived from surveillance data, and a number of associated dates may be considered.⁴ As date of infection and date of onset are seldom known the usual choice is between date of diagnosis and date of report. Although date of diagnosis is more natural, there is a disadvantage in that the curve appears to fall off, as there is a time delay in reporting the most recent cases.¹⁸ As the most recent data are the most crucial for prediction it is preferable to plot by date of report and if necessary make an adjustment for any delays in reporting. The curve extrapolation methods in use include logarithmic linear regression (global or piecewise),^{2,3} quadratic regression with Box-Cox transformation of the dependent variable,¹ and general linear modelling.⁴ The main inaccuracy in these pre-

dictions lies in the choice of extrapolation method or curve to be fitted rather than the statistical error in the values predicted by one chosen curve.⁴ For the United Kingdom at present, predictions by different methods give appreciably different results after one to two years.⁴ Straightforward logarithmic linear models assume a constant doubling time, which is not supported by the experience in the United States.¹

Compartment models divide the population into groups or compartments by infection state, such as susceptible, incubating, infectious, or diseased, or by risk factor, such as multiple sexual relationships or injecting drugs. Mathematical equations describe the spread of infection between the compartments. The equations include detailed variables such as the duration of infectivity, the proportion of those infected with HIV who will eventually develop AIDS, and the distribution of sexual activity.^{5,7} Values for these variables are, on the whole, poorly known, and the observed incidence of AIDS can be fitted by a large variety of values for them.¹² The predictions produced are very sensitive to these values⁸ and the form of the equations chosen. Given the data available at present, and in the foreseeable future, this sort of model is mainly useful for predicting the long term overall shape and time course of the epidemic, rather than for short term quantitative predictions.

Models based on the distribution of the incubation period attempt to predict the minimum size of the epidemic arising from people already infected with HIV by using assumptions about the distribution of the incubation period. Working backwards from the distribution in time of observed cases, the distribution in time of past infections is deduced. Working forwards, the incubation period distribution gives predictions for the number of cases which will arise from current HIV infections.^{9,10} The method is very sensitive to the exact form assumed for the distribution of the incubation period.¹² This distribution is estimated from cases arising from blood transfusions,¹³ and it may not hold for those infected through sexual activity; it cannot therefore be estimated with great accuracy. The method ignores future transmission and is of little use other than to give an "at least" estimate of the numbers of cases that will be seen in the future.

Comparison models attempt to match the epidemic curve with that from another place where the epidemic is more advanced, such as comparing the United Kingdom with the United States.¹¹ The use of this method for quantitative prediction is difficult because of the need to estimate simultaneously the exact time lag and the scaling factor to account for differences in population. The extent of each risk factor is not known accurately enough to contribute to the scaling factor. Differences between populations, such as in risk behaviour, may invalidate the comparison—for example, the epidemic in Italy continues to rise faster than in the United Kingdom and to affect proportionately more drug users who inject (G Marasca, personal communication).

Expert committee models entail debate and negotiation between knowledgeable parties. Despite the availability of protocols such as the Delphi method to improve task orientation with anonymity,¹⁴ committees have relied so far on the evidence of experts using one of the other types of model.^{15,16}

All these models lie within a range from complex structures requiring detailed data, sensitive to assump-

Department of Community
Medicine, Cambridge
Health Authority, Fulbourn
Hospital, Cambridge CB1
5EF

B R Tennison, PHD,
registrar

Health Education
Authority, London
WC1A 1AH
S Hagar, PHD, chief
executive

Correspondence to:
Dr Tennison.

tions and suited to long term projections, to more simple, robust structures requiring less detailed data that are more suitable for short term projections. Compartment models tend to lie in the first part of this range, trend extrapolation models in the last part, and the other types in between.

Methods and results

At present the only type of model offering the prospect of useful short term (five year) prediction is trend extrapolation. We lack the detailed data needed to use compartment models for this prediction; the distribution of incubation period is too uncertain and models based on it are too sensitive; and comparison models cannot be made accurate enough. The validity of using trend extrapolation for the United Kingdom at the present time is supported by two considerations: firstly, the epidemic in the United Kingdom is at an earlier stage than that in the United States, where it continues to follow a smooth increasing curve; and, secondly, the kinds of epidemic curves predicted by the best of the other models⁶ suggest that trend extrapolation will be valid over a five year period from 1987. On the other hand, in the areas outside London the numbers of cases of AIDS up to 1987 have been too small to form by themselves a basis for extrapolation.

To fulfil our objective we used a two stage approach: (a) extrapolation of the national data to predict numbers of cases nationally; followed by (b) scaling down the national predictions to give local predictions. Exploratory data analysis of the number of cases of AIDS by quarter in which the disease was reported (Public Health Laboratory Service Communicable Disease Surveillance Centre, unpublished) to the end of 1987 showed that the doubling time was increasing and therefore indicated the inadequacy of a simple logarithmic linear extrapolation. A Box-Cox analysis¹⁷ was carried out to find the most appropriate power transformation for the dependent variable of logarithmic linear weighted regression and of quadratic regression (as used by the Centers for Disease Control, Atlanta).¹ To emphasise the relative importance of recent data in making predictions, the weightings used exponentially decreased into the past with a common ratio (0.9).

Table I gives details of three models that were then used to calculate low, medium, and high predictions for the number of cases of AIDS up to 1992 (table II). An analysis of sensitivity showed that the main sources of bias and error in these predictions were the unknown degree of underreporting and the form of the extrapolation curve; the statistical error in any one of the regression extrapolations was much smaller. For each year of prediction a scaling factor was used to give a local prediction from the national one. The East Anglian Regional Health Authority has accurately assessed the number of cases of AIDS and has a detailed local surveillance scheme (C Duff, personal communication). The numbers of cases observed for East Anglia and Cambridge were too small to arise from a scaling factor simply derived by taking population ratios or by age standardisation. We lacked the data needed to standardise for size of risk group, and therefore the scaling factors used were the currently

TABLE II—Predicted number of case reports of AIDS in United Kingdom during each year to 1992 (non-cumulative) according to three trend extrapolation models

Year	Prediction level of model		
	Low	Medium	High
1987		656 (actual)	
1988	1030	1060	1120
1989	1620	1730	1920
1990	2300	2690	3180
1991	3040	4020	5100
1992	3740	5830	8010

TABLE III—Predicted number of new patients presenting for treatment for AIDS during each year to 1992 (non-cumulative) in East Anglia and Cambridge (approximate 95% confidence intervals)

Year	East Anglia	Cambridge
1987	9 (actual)	3 (actual)
1988	19 (10 to 29)	8 (2 to 14)
1989	31 (18 to 46)	13 (5 to 21)
1990	48 (29 to 72)	20 (9 to 33)
1991	72 (40 to 112)	29 (13 to 49)
1992	105 (47 to 182)	43 (17 to 79)

observed proportions of national cases—namely, 0.018 for the East Anglian region and 0.0073 for the Cambridge district. An analysis of sensitivity allowed these factors to rise or fall in the years to 1992 to an extent derived by judgment from the observed proportions in the past. In addition, as the predicted local numbers were small the Poisson error was large. Combining these sources of error gave rise to the predictions and approximate 95% confidence intervals shown in table III.

Discussion

The national predictions in table II are smaller than those derived from logarithmic linear regression, which assumes a constant doubling time. This agrees with the experience in the United States, where doubling time continues to lengthen. We can be more confident with predictions over the early years of the five year period. Extrapolation for more than about five years will probably be inaccurate because the epidemic curve may start to flatten over that length of time, corresponding to saturation of the higher risk populations (as in the curves predicted by some compartment models).⁶ As the median incubation period for AIDS in adults is probably at least five years¹³ preventive activities and changes in behaviour now and in the recent past will have little impact over the next five years. In the future it may become appropriate to fit different curves—for example, those with a plateau, such as logistic or Gompertz curves. In addition, as knowledge advances other sorts of model (such as compartment models) may become more refined and able to provide useful predictions.

Table II shows the predictions of the number of case reports to the Communicable Disease Surveillance Centre, as they are derived from observed reports. There is evidence to suggest that causes of death related to AIDS are underrecorded on death certificates.¹⁸ No study of underreporting has been published, but it is likely that the number of cases nationally is at least 20% more than the number reported. The serious nature of the disease means that almost all people with AIDS will make contact with the health service and require treatment. To obtain predictions of the number of new patients requiring treatment each year the figures in table II should be increased by a factor of at least 20% to compensate for underreporting. For a local area with a good surveillance scheme and little underreporting such compensation is not necessary;

TABLE I—Three trend extrapolation models for number of case reports of AIDS in United Kingdom

Prediction level of model	Formula	Optimal m	a	b	c
Low	$r^m = a + b + ct^2$	0.08	1.084	0.0274	-0.000278
Medium	$s^m = a + bt + ct^2$	0.31	1.621	0.276	0.00479
High	$(\ln x)^m = a + bt$	2.05	0.915	2.743	

r=Quarterly number of case reports; s=cumulated number to date; t=quarter number (l=1983, first quarter, 20=1987, fourth quarter); m=power chosen by the Box-Cox method; a, b, and c=regression coefficients.

in particular it is not needed for the predictions in table III.

Healy and Tillett showed that the median delay in reporting from date of diagnosis was one to two months with few delays of over six months.⁴ Hence to convert the predictions in tables II and III to predictions by date of diagnosis entails an insignificant adjustment corresponding to a time shift of one to two months. As at present the median survival time after diagnosis of AIDS is about one year,¹⁹ the number of people needing care during a given year can be estimated from table II or table III by adding the entry for that year to 50% of the entry for the previous year.

The method of scaling down to local populations can be applied to any region where accurate local case data are available. If the observed number of cases is too small the predictions will be unreliable. Other local factors, such as knowledge of specific risk factors, should be taken into account.

Predictions are an essential basis for planning local services for people with AIDS. Planning can proceed only once there is an estimate of the likely size of the case load. The form that the services take will depend on the philosophy of care, including the balance between institutional and community aspects. This balance should be made explicit, and as services need to be provided by health authorities, local authorities, and voluntary organisations, they should be planned by these agencies working together.

The long incubation period of AIDS means that preventive measures have a delayed impact. The need for prevention is emphasised, however, by the size of the predicted case load, the burden of care that this implies, and the fact that the incidence will continue to rise after the end of the five year prediction period. The local significance of these factors will be clarified by the local predictions and should encourage local

preventive activities. Monitoring of national and local case data will allow validation and updating of these predictions.

We thank Professor M J R Healy and Dr D R R Williams for their help.

- 1 Curran JW, Morgan WM, Hardy AM, Jaffe HW, Darrow WW, Dowdle WR. The epidemiology of AIDS: current status and future prospects. *Science* 1985;229:1352-7.
- 2 McEvoy M, Tillett HE. Reassessment of predicted numbers of AIDS cases in the UK. *Lancet* 1986;ii:1104.
- 3 Downs AM, Ancelle RA, Jager HJC, Brunet J-B. AIDS in Europe: current trends and short-term predictions estimated from surveillance data, January 1981-June 1986. *AIDS* 1987;1:53-7.
- 4 Healy MJR, Tillett HE. Short-term extrapolation of the AIDS epidemic. *Journal of the Royal Statistical Society (Series A)* (in press).
- 5 May RM, Anderson RM. The transmission dynamics of HIV infection. *Nature* 1987;326:137-42.
- 6 Anderson RM, Medley GF, May RM, Johnson AM. A preliminary study of the transmission dynamics of HIV, the causative agent of AIDS. *IMA Journal of Mathematics Applied in Medicine and Biology* 1986;3:229-63.
- 7 Knox EG. A transmission model for AIDS. *Eur J Epidemiol* 1986;2:165-77.
- 8 Pickering J, Wiley JA, Padian NS, Lieb LE, Echenberg DF, Walker J. Modeling the incidence of AIDS in San Francisco, Los Angeles and New York. *Mathematical Modelling* 1986;7:661-88.
- 9 Brookmeyer R, Gail MH. Minimum size of the AIDS epidemic in the United States. *Lancet* 1986;ii:1320-2.
- 10 Rees M. Describing the AIDS epidemic. *Lancet* 1987;ii:98-9.
- 11 Mortimer PP. Estimating AIDS, UK. *Lancet* 1985;ii:1065.
- 12 Anderson RM, Medley GF, Blythe SP, Johnson AM. Is it possible to predict the minimum size of the AIDS epidemic in the UK? *Lancet* 1987;i:1073-5.
- 13 Medley GF, Anderson RM, Cox DR, Billard L. Incubation period of AIDS in patients infected via blood transfusion. *Nature* 1987;328:719-21.
- 14 Linstone HA, Turoff M. *The Delphi method: techniques and applications*. Reading, Massachusetts: Addison-Wesley, 1977.
- 15 Anonymous. Coolfont report: a PHS plan for prevention and control of AIDS and the AIDS virus. *Public Health Rep* 1986;101:341-8.
- 16 House of Commons Social Services Committee. *Problems associated with AIDS*. London: HMSO, 1987.
- 17 Box GEP, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society (Series B)* 1964;26:211-43.
- 18 McCormick A. Trends in mortality statistics in England and Wales with particular reference to AIDS from 1984 to April 1987. *Br Med J* 1988;296:1289-92.
- 19 Marasca G, McEvoy M. Length of survival of patients with AIDS in the UK. *Br Med J* 1986;292:1727-9.

(Accepted 29 June 1988)

Body mass and prostatic cancer: a prospective study

Richard K Severson, John S Grove, Abraham M Y Nomura, Grant N Stemmermann

Abstract

Previous studies have suggested that increased body mass is associated with an increased risk of prostatic cancer, but these studies have been limited by the fact that they were based on a few simple measurements such as height and weight. Similar results were found in a prospective study of the incidence of prostatic cancer in a cohort of Japanese men born in 1900-19 and living in Hawaii. Further evaluation of the extensive anthropomorphic measurements made in this cohort suggested that the association between measures of body mass and prostatic cancer might be accounted for more by lean tissue than by fat tissue. There was a significant positive association of the risk of prostatic cancer with area of muscle in the arm but not with area of fat in the arm.

Further research is needed on the biological mechanisms of carcinogenesis that may be related to both lean and fat tissue and the development of prostatic cancer.

Introduction

Although several studies have attempted to identify factors that cause prostatic cancer, little has been established. Recently it was reported that the risk of prostatic cancer increased with increasing body mass index.¹ This finding, which was derived from a

hospital based case-control study of the incidence of prostatic cancer in northern Italy, is consistent with mortality data from prospective studies of Seventh Day Adventists² and volunteers from the American Cancer Society³ in the United States, which both found an increased risk of prostatic cancer in overweight men. These studies were restricted in that they related risk solely to variables based on height and weight. We report a prospective study of the incidence of prostatic cancer in a cohort of Japanese men living in Hawaii in which we recorded extensive anthropomorphic measurements.

Patients and methods

A cohort of 8006 Japanese men who had been born during 1900-19 and were living in the Hawaiian island of Oahu in 1965 were examined and interviewed from 1965 to 1968.⁴ Several anthropomorphic measurements were recorded during the clinical examination. Body mass index was calculated as weight (kg)/(height (m))², and the areas of muscle and fat in the upper arm were calculated according to the equations of Heymsfield *et al* and Frisancho, respectively.^{5,6} Seven men with prostatic cancer at the time of examination were eliminated from the analysis.

The time at risk of prostatic cancer was calculated for each subject as the time from examination to a

Japan-Hawaii Cancer Study, Kuakini Medical Center, Honolulu, Hawaii 96817, United States

Richard K Severson, PHD,

epidemiologist

John S Grove, PHD,

biostatistician

Abraham M Y Nomura,

MD, director

Grant N Stemmermann, MD,

pathologist

Correspondence to: Dr R K Severson, Program in Epidemiology W403, Fred Hutchinson Cancer Research Center, Seattle, WA 98104, United States.