Listy Biometryczne – Biometrical Letters
Vol. 30 (1993), No. 2, 97-108

Backcalculation of constrained, flexible models of the human immunodeficiency virus infection curve

Antonino Salvaggio

Istituto di Igiene e Medicina Preventiva,
Università degli Studi di Milano,
Via F. Sforza, 35 - 20122 Milano, Italy

Summary

We present a simple approach to obtain constrained estimates of the infection curves, relative to the number of persons infected with the human immunodeficiency virus (HIV). Because expected HIV incidence must be expressed as a non-negative function, we use B-splines to represent the HIV infection curve, subject to the non-negativity constraint.

1. Introduction

The backcalculation is a method used to estimate the number of persons infected with the human immunodeficiency virus (HIV), and to project future acquired immune deficiency syndrome (AIDS) incidence (Brookmeyer and Gail, 1986, 1988; Center of Disease Control, 1987, 1990; Gail and Brookmeyer, 1988; Brookmeyer and Damiano, 1988; Taylor, 1989; Day and Gore, 1989; Isham, 1989; Rosemberg and Gail, 1991).

Brookmeyer and Gail (1986, 1988), Gail and Brookmeyer (1988), and Rosemberg and Gail (1991) discuss the backcalculation method, applied to estimating the numbers of individuals previously infected with the human immunodeficiency virus by flexible models. They base their backcalculation on step function models, as well as more general spline functions.

Although Rosemberg and Gail (1991) suggest the possibility of using constrained estimates to obtain nonnegative valued infection curves, their method,

Key words: backcalculation, human immunodeficiency virus, spline

introducing spline functions, admit the possibility of negative values at some time, and their (quadratic) spline model of the infection in the USA yields small negative values early in the epidemic. The integrated contributions from negative values were relatively small, and were judged to have little impact on the estimate of the total number of subjects infected, or on projections.

Rosemberg and Gail (1991) suggest however that each particular infection curve yielding negative values, as well as their relative families, should be reviewed critically, modified, or discarded. Thus, we present a simple approach to obtain constrained estimates of the infection curves.

We base our infection curve models on B-splines representations. Using B-splines of order one, step functions are included in our models (with discontinuity at knot points). Also, broken-lines are included as combinations of B-splines of order two (so called "hat functions"). Usual spline functions with continuous first or higher order derivatives are included in the representation proposed by us, if B-spline coefficients are estimated without constraints. Finally, we propose non-negativity constrained B-spline combinations as proper models for infection curves.

Thus, we present a regression approach to the backcalculation method, to estimate the HIV infection curve using flexible, nonparametric models.

Because expected HIV incidence must be expressed as a nonnegative valued function, we use B-splines (Boor, 1978) to represent HIV infection curves according to a non-negativity constraint.

2. The infection curve model

Linear B-spline combinations include step functions, picewise linear, as well as usual picewise polynomial functions (Boor, 1987, in particular chapter IX). Using B-spline representations, simple constrains in the curve values and derivatives are easily introduced. To be specific, infection curves, representing the number of persons infected per time unit (or "instantaneous" infection rates), must be non-negative valued.

Given a partition

$$a = \tau_1 = \ldots = \tau_k < \tau_{k+1} < \ldots < \tau_n < \tau_{n+1} = \ldots = \tau_{n+k} = b$$

of an interval $[a,b]$, a "spline" is a polynomial on each interval $[\tau_j, \tau_{j+1}]$ (we shall refer to the τ_i as "knots", introducing repeated knots at a and b to facilitate the introduction of the B-spline representation). In particular, parabolic (3rd order), and cubic splines (4th order), as well as higher order splines (in general, k-th order), are smooth functions, respectively C^1 and C^2 or smoother (C^{k-2}) .

Given a knot sequence, B-splines represent a convenient basis for the space of splines. In fact, any spline function $s(x)$ can be uniquely expressed as a linear combination of the B-splines $B_j(x)$, $j=1,\ldots,n$, in the form

$$s(x) = \sum_{j=1}^{n} \beta_j \cdot B_j(x) \ . \tag{1}$$

B-spline functions are non-negative and "local", i.e. only k B-splines, from $B_{j-k+1}(x)$ to $B_j(x)$, are nonzero on each particular interval $[\tau_j, \tau_{j+1}]$. Thus, the β_j sequence give a fair idea of the graph of $s(x)$. Moreover, if the β_j are non-negative, so is $s(x)$, and if β_j is a monotonic sequence, $s(x)$ is monotonic (Boor, 1987, in particular chapter IX and X).

3. From HIV infection to AIDS, the model

Being t_0 the start of the epidemic, we assume that AIDS incidence counts are available from t_0 to t_J, the ultimate date for which reliable data are available.

We assume that the number of AIDS cases is available in discrete form. Thus, having divided calendar time into J intervals $(t_0, t_1],\ldots,(t_{J-1}, t_J]$, we denote as Y_j the number of AIDS cases diagnosed in the j-th interval $(t_{j-1}, t_j]$.

We estimate the infection curve $s(x)$, which specifies the expected number infected in time interval $(s, s+ds)$, considering the general family of infection curves defined by a B-spline basis set $B = \{B_1(x), \ldots, B_n(x)\}$, where $B_i(x) \equiv B_{i,k}$ are B-spline functions of order k.

Thus, the infection function, $s(x)$, is defined according to (1), for real coefficients $\beta_i, \ldots, \beta_n$. Imposing the non-negativity restriction to β_i , we constrain $s(x)$ to a nonnegative valued function.

The total number of infected in an interval $(t_{k-1}, t_k]$ is

$$\int_{t_{k-1}}^{t_k} s(x)dx = \sum_{i=1}^{n} \beta_i \cdot \{ \int_{t_0}^{t_k} B_i(x)dx - \int_{t_0}^{t_{k-1}} B_i(x)dx \} \ . \tag{2}$$

Let $F(t)$ be the assumed incubation distribution, that describes the time from infection to clinical AIDS (equal to zero for $t<0$). The expected number of cases in interval $(t_{j-1}, t_j]$ is given by the convolution

$$E(Y_j) = \int_{t_0}^{t_j} s(x)\{F(t_j - x) - F(t_{j-1} - x)\}dx$$

$$= \sum_{i=1}^{n} \beta_i \int_{t_0}^{t_j} B_i(x) \{ F(t_j - x) - F(t_{j-1} - x) \} dx$$

$$= \sum_{i=1}^{n} \beta_i \cdot g_{ij} \ , \qquad (3)$$

where g_{ij} are obtainable from numerical integration. To obtain g_{ij} values we use an adaptative integration method (Rice, 1983), allowing for possible singularities at knot points.

Thus, β_j coefficients may be estimated by regression methods. It is possible to estimate the parameters β_i according to a constrained linear problem ($\beta_i \geq 0$). However, we prefer to constrain the parameters β_i introducing the transformations $\beta_i = \exp(\beta'_i)$, solving a nonlinear problem on β'_i. Further, we assume that the numbers of AIDS cases diagnosed on each time interval, Y_j, j=1,...,J, are independent Poisson variates. Thus we estimate parameters according to a Poisson, maximum likelihood method (McCullagh and Nelder, 1989).

Finally, interval estimates have been computed according to a bootstrap method (Hinkley, 1988, in particular pages 330-332). Bootstrap estimates have been obtained by sampling with replacement the residuals from the expected values obtained by the nonlinear regression that relate AIDS cases to the number of HIV infected. Each residual has been considered as a standardised residual, normalized with respect to the expected variance, proportional to the expected value, according to the Poisson model adopted for the stochastic component of our regression model. The confidence limits were obtained according to the percentile method (DiCiccio and Romano, 1988).

4. Examples

We illustrate the methods using AIDS incidence data for the United States, adjusted for reporting delays (Rosemberg, 1990), which have been utilised by Rosemberg and Gail (1991) to present their approach to the backcalculation.

Figure 1 presents estimates for the infection curve obtained according to unconstrained and constrained models. Left estimates were obtained assuming that the incubation distribution is Weibull with $F(t) = 1 - \exp(-0.0021\, t^{2.516})$, with median incubation period equal to 10 years (Brookmeyer and Goedert, 1989). Right estimates were obtained using the "Fast"–Weibull incubation function (lower 95% confidence limit for Weibull incubation curves), $F(t) = 1 - \exp(-0.0021\, t^{2.650})$, already adopted by Rosemberg, Gail, and Carroll

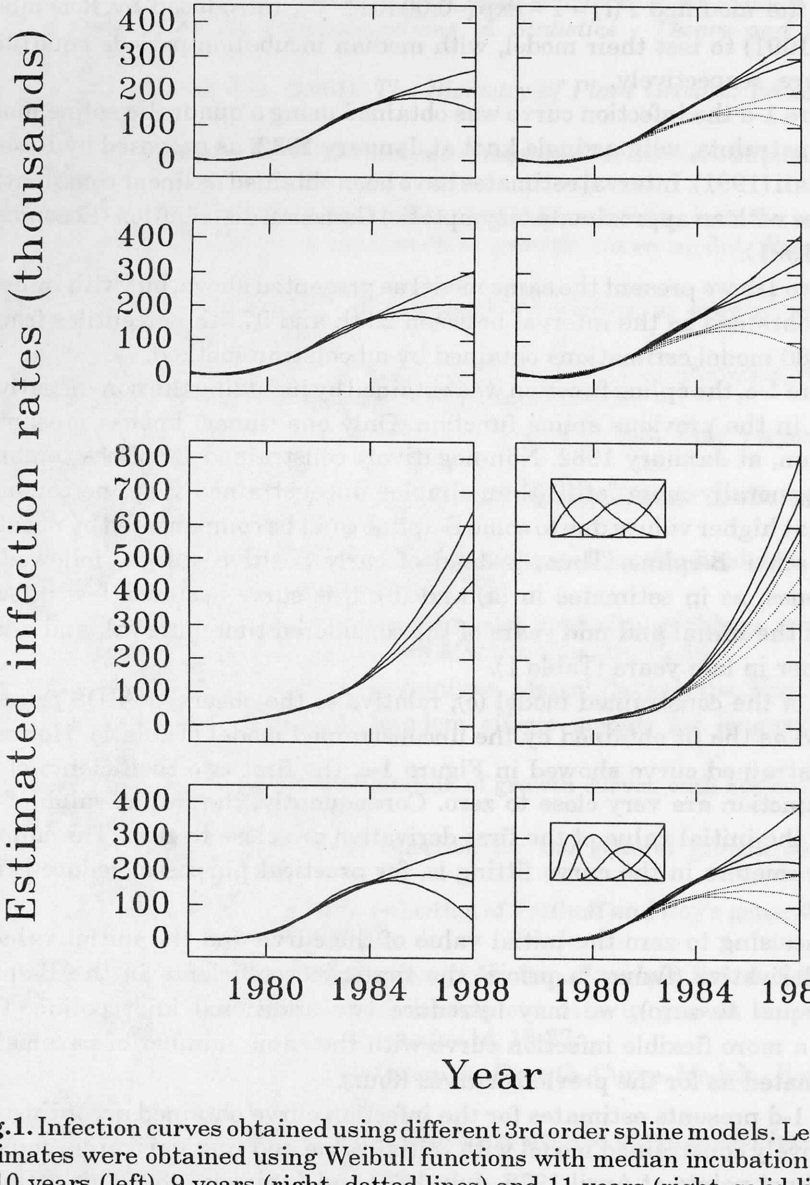

Fig.1. Infection curves obtained using different 3rd order spline models. Left and right estimates were obtained using Weibull functions with median incubation time equal to 10 years (left), 9 years (right, dotted lines) and 11 years (right, solid lines). In (a) we present the model without constraints, with a single knot at January 1982, and asymptotic interval estimates; in (b) the model (a) is presented with bootstrap interval estimates (percentile method); in (c) we present the infection curve obtained introducing the non-negativity constraint in the previous model (one knot at January 1982). Finally, graph (d) presents estimates of the infection curve obtained according a non negatively constrained model with initial value and first derivative equal to zero, and knot points at April 1978, July 1979, and January 1982. Insets show the adopted B-spline basis.

(1992), or the modified $F(t) = 1 - \exp(-0.0040\, t^{2.149})$, introduced by Rosemberg and Gail (1991) to test their model, with median incubation periods equal to 9 and 11 years, respectively.

In Figure 1-a the infection curve was obtained using a quadratic spline model without constraints, with a single knot at January 1982, as proposed by Rosemberg and Gail (1991). Interval estimates have been obtained as linear combination of estimates with an approximate (asymptotic) Gaussian distribution (Rosemberg and Gail, 1991).

In Figure 1-b we present the same model as presented above, but with interval estimates obtained as the interval between 2.5th and 97.5th percentiles from a sample of 80 model estimations obtained by a bootstrap method.

In Figure 1-c, the spline function was obtained by including the non–negativity constraint in the previous spline function. Only one (inner) knot is present in this function, at January 1982. Non-negatively constrained B-splines combinations are generally more "stiff" than simpler unconstrained B-spline combinations, in that higher values due to some B-spline can't be compensated by negative values of other B-spline. Thus, instead of early positive values, followed by negative ones, as in estimates in (a) and (b), this curve assumes few infected subjects in the initial and mid years of the considered time interval, and a very high number in late years (Table 1).

The fit of the constrained model (c), relative to the observed AIDS cases, is not as good as the fit obtained by the unconstrained model (Table 1). However, in the constrained curve showed in Figure 1-c, the first two coefficients of the B-spline function are very close to zero. Consequently, the initial value of the curve and the initial value of the first derivative are close to zero. The number of free parameters in the curve fitting is, for practical purposes, reduced from four to two.

Hypothesising to zero the initial value of the curve and the initial value of the first derivative (fixing "a priori" the first two coefficients in the B-spline function equal to zero), we may introduce two additional knot points, thus obtaining a more flexible infection curve with the same number of parameters to be estimated as for the previous curves (four).

Figure 1-d presents estimates for the infection curve obtained according to a non-negatively constrained model with initial value and first derivative equal to zero, and knot points at April 1978, July 1979 and January 1982. This estimate of infection curve is similar to the infection curve estimated by a simple, unconstrained spline. However, in this case the expected HIV infection is expressed by non-negative numbers all time. Also, the fit of the observed AIDS cases obtained by this constrained model appears as good as the fit obtained with the unconstrained model (Table 1).

It is relatively easy to introduce more complex models that accounts for therapy and changes in the surveillance definition of AIDS. In fact, since these

Table 1
Estimates of the number of persons infected with HIV before April 1988, according to various models, incubation curves and B-splines basis

Model characteristics (Figure 1 reference)		Numbers previously infected	(95% C.I.)[1]	Goodness of fit[2]
$F(t) = 1{-}\exp(-0.0021t^{2.516})$				
one (inner) knot at January 1982	(a)	1183614	(1091232-1275997)	100.6
+ bootstrap C.I.	(b)	1183614	(1043697-1354284)	100.6
+ non-negativity constraint	(c)	1682418	(1498514-1898480)	619.9
+ knots April 1978 and July 1979[3]	(d)	1139570	(969837-1321477)	101.3
$F(t) = 1{-}\exp(-0.0021t^{2.650})$				
one (inner) knot at January 1982	(a)	935533	(852976-1018091)	99.0
+ bootstrap C.I.	(b)	935533	(817908-1081937)	99.0
+ non-negativity constraint	(c)	1420887	(1261313-1629366)	536.4
+ knots April 1978 and July 1979[3]	(d)	903518	(751115-1015606)	100.1
$F(t) = 1{-}\exp(-0.0040t^{2.149})$				
one (inner) knot at January 1982	(a)	1420978	(1358257-1483699)	106.2
+ bootstrap C.I.	(b)	1420978	(1318889-1547120)	106.2
+ non-negativity constraint	(c)	1943734	(1757128-2156386)	858.8
+ knots April 1978 and July 1979[3]	(d)	1390466	(1269157-1518874)	104.7
$F_0(t) = 1{-}\exp(-0.0021t^{2.516})$, "natural incubation"				
+ time changes allowing for therapy and the 1987 revision of the AIDS definition, bootstrap C.I., (inner) knots at April 1978, July 1979, January 1982[3] gay : IVDU/heterosexuals, 2 : 1		913387	(808860-999552)	82.7
+ constant CV from October 1985		927410	(812207-1085601)	82.9[4]

[1] 95% confidence interval
[2] sum of the squared deviations divided by the expected values
[3] initial value and first derivative equal to zero
[4] weighted deviations, weights have been scaled to obtain results comparable to previous models

variables involve changes in the incubation distribution only, our B-spline representation of the infection curves is unchanged. Figure 2 presents infection curves estimates obtained according to a model allowing for therapy and the revision of the AIDS surveillance definition in 1987, according to the hypothesis formulated by Rosemberg, Gail, and Carrol (1992). In this example, we simly hypothesized that gay and IVDU/heterosexuals contribute to AIDS incidence in a 2:1 ratio. Also, a gradual onset of the effects of changes in AIDS definition was supposed, with main changes in "incubation" distribution occuring during 1 January 1985 – 31 December 1987 (Appendix).

Finally, to model the uncertainty introduced by correction for reporting delays, we may give up simple Poisson regression to model a constant coefficient of variation (CV) for adjusted AIDS counts according to Rosemberg's observations (Rosemberg, 1990). An example is showed in Figure 2.

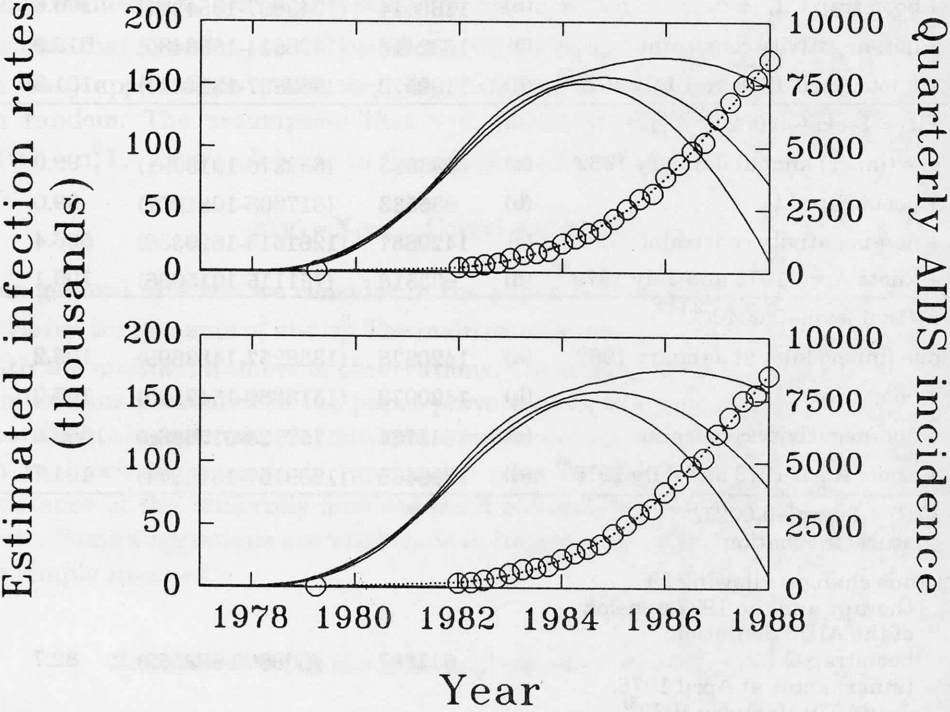

Fig.2. Observed (o) and estimated (dotted lines) AIDS cases, and infection curves (solid lines) according to a model allowing for therapy and revision of AIDS definition. B-splines (infection curve model) and interval estimates are as in Figure 1-d. Curve on the top was estimated according to a simple "Poisson" regression. Bottom curve represents the result of a joint model for expected values and dispersion, according to a quasi-likelihood approach in wich variance was hypothesized to be proportional to the expected values before 1 October 1985, and proportional to squared expected values – constant coefficient of variation (CV) – from 1 October 1985. CV was estimated by joint modelling expected values and dispersion, maximizing an extended quasi–likelihood, in which expected values have been modeled according to a (quasi) Poisson and gamma regression and a gamma distribution was used to model dispersion parameters (McCullagh and Nelder, 1989). Estimated CV from 1 October 1985 was 2.5%, (before 1 October 1985 we also estimate Poisson overdispersion, with a dispersion parameter equal to 3). Very similar results have been obtained fixing "a priori" CV to 3%, according to the values obtained by Rosemberg (1990) from Northeastern region of United States.

Estimates of previous numbers infected according to the various models proposed, and goodness of fit of the models, are resumed in Table 1.

5. Comments

In this paper we present flexible models which may be used to backcalculate the HIV infection curve subject to non-negativity constraints. Spline functions, according to a B-spline representation, are used.

Simpler models, as step functions and broken-lines, are included as special cases, but the inclusion of the non-negativity constraint introduce some calculation complexities, especially regarding to interval estimation. Thus, in our approach, the estimation of the infection curve was a nonlinear regression problem, and interval estimates have been calculated according to a bootstrap method. However, the biological relevance of the non-negativity constraint, and the importance of the data and results, justify our computational efforts.

Non-negatively constrained B-spline combinations appear less flexible than simpler, unconstrained B-spline combinations, or "classical" spline functions, in that higher values due to some B-spline can not be compensated by negative values of other B-spline. However, this is not a disadvantage, involving less wiggles in estimated curves, provided that underlying ties are recognised.

B-spline representation of classical spline functions involve multiple knots at extreme values. In particular, B-spline with multiple knots at time zero may have, at time zero, a positive value and a negative trend (1st B-spline; three knots) or a positive derivative (2nd B-spline; two knots). However, it may occur that positive, initial values (positive 1st B-spline coefficient) may be accepted only if associated with a very fast negative trend (negative 2nd B-spline coefficient) and eventual negative values, or a strong, initial (and extended) positive trend (positive 2nd B-spline coefficient) may appear inappropriate; thus, when fitting non-negatively constrained B-splines, one may force the first two coefficients to zero, with a possible important penalization in flexibility. In our case only four parameters were introduced; in constrained models, they were reduced, in practice, to two, with a great loss in flexibility. Introducing zero initial value and zero first derivative by hypothesis, allow us to introduce additional "inner" knots, thus restoring flexibility to our constrained curves without introducing too many parameters.

Received 28 September 1993

REFERENCES

de Boor C. (1978). *A practical guide to splines.* New York: Springer–Verlag.

Brookmeyer R., Damiano R. (1988). Statistical methods for short term projections of AIDS incidence. *Statistics in Medicine* **8**, 23-34.

Brookmeyer R., Gail M.H. (1986). Minimum size of the acquired immunodeficiency syndrome (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. *Journal of the American Statistical Association* **83**, 301-308.

Brookmeyer R., Goedert J.J. (1989). Censoring in an epidemic with application to hemophilia-associated AIDS. *Biometrics* **45**, 325-335.

Center of Disease Control (1987). Human immunodeficiency virus infection in the United States: a review of current knowledge. *Morbidity and Mortality Weekly Report* **36**, 1-48.

Center of Disease Control (1990). HIV prevalence estimates and AIDS case projections for the United States: report based upon a workshop. *Morbidity and Mortality Weekly Report* **39**, 1-31.

Day N.E., Gore S.M. (1989). Prediction of the number of new AIDS cases and the number of new persons infected with HIV up to 1992: the results of "back projection" methods. In *Short-term Prediction of HIV Infections and AIDS in England and Wales*. London: Her Majesty's Stationery Office.

DiCiccio T.J., Romano J.P. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society, Series B* **50**, 338-354.

Gail M.H., Brookmeyer R. (1988). Methods for projecting course of acquired immunodeficiency syndrome epidemic. *Journal of the National Cancer Institute* **80**, 900-911.

Hinkley D.V. (1988). Bootstrap methods. *Journal of the Royal Statistical Society, Series B* **50**, 321-337.

Isham V. (1989). Estimation of the incidence of HIV infection – the back projection method. In *Short-term Prediction of HIV Infections and AIDS in England and Wales*. London: Her Majesty's Stationery Office.

McCullagh P., Nelder J.A. (1989). *Generalized Linear Models*. 2nd edition. New York: Chapman and Hall.

Rice J.R. (1983). *Numerical methods, software, and analysis*. New York: McGraw-Hill, pp. 193-196.

Rosemberg P.S. (1990). A simple correction of AIDS surveillance data for reporting delays. *Journal of the Acquired Immunodeficiency Syndrome* **3**, 49-54.

Rosemberg P.S., Gail M.H. (1991). Backcalculation of flexible linear models of the human immunodeficiency virus infection curve. *Applied Statistics* **40**, 269-282.

Rosenberg P.S., Gail M.H., Carrol R.J. (1992). Estimating HIV prevalence and projecting AIDS incidence in the United States: a model that accounts for therapy and changes in the surveillance definition of AIDS. *Statistics in Medicine* **11**, 1633-1655.

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

Received 25 September 1993

Appendix

A model allowing for therapy and the revision of the AIDS definition

The expected number of cases in interval $(t_{j-1}, t_j]$ is given by the formula

$$\mathrm{E}(Y_j) \;=\; \int_{t_0}^{t_j} s(x)\{F(t_j-x\,|\,s)\,F(t_{j-1}-x\,|\,s)\}dx \;\;,$$

which differs from (3) since incubation depends on s, the time of HIV infection. The incubation function, u years after infection, is defined as

$$F(u\,|\,s) = \int_{\tau\geq s} F(u\,|\,s,\tau)\,dP(\tau\,|\,s) \;\;,$$

where

$$P(\tau\,|\,s) \;=\; \begin{cases} 0 & \text{for } \;\tau<s_0 \;\text{ or }\; \tau<s \\ p(\tau-s_0)\,/\,(s_f-s_0) & \text{for } \;s_0\leq\tau<s_f \;\text{ and }\; \tau\geq s \\ p & \text{for } \;\tau>s_f \;\text{ and }\; \tau\geq s \end{cases} \quad ,$$

is the treatment onset distribution, with $s_0\,/\,s_f$ equal to 1 April 1987/31 March 1990 in gay, 1 April 1988/31 March 1991 in IVDU and heterosexuals, and

$$F(u\,|\,s,\tau) = 1\text{–}\exp\{-\int_0^u h(u\,|\,s,\tau)du$$

is the incubation distribution according to the hazard function

$$h(u\,|\,s,\tau) = h_0(u)\{\,\theta(u)I(u{+}s \geq \tau) + I(u{+}s < \tau)\,\}\cdot\delta(u{+}s) \;\;,$$

where $h_0 = 0.0021\cdot 2.516\,t^{\,1.516}$ is the "natural" hazard, corresponding to the Weibull distribution $F(t) = 1\text{–}\exp(-0.0021\,t^{\,2.516})$, $I(\cdot)$ is an indicator function with value 1 when the argument is true and 0 otherwise, and τ is the calendar time at which people have access to treatment.

$\theta(u)$ is an efficacy function relative to treatment effects,

$$\theta(u) = 1 - 0.5\cdot\exp\{\,2(u\text{–}\alpha)\}\,/\,[1\text{+}\exp\{\,2(u\text{–}\alpha)\}]\;, \quad \alpha = 5.0 \;\text{ years.}$$

$\delta(u{+}s)$ includes the effects of AIDS definition changes in 1987: not excluding AIDS cases diagnosed before 1 October 1987 solely on the basis of newly-recognised conditions (Rosemberg, Gail and Carroll, 1992), we suppose a gradual increase in reporting (hazard), according to the function

$$\delta(u+s) = 1 + 0.1 \cdot \exp\{2(u+s-\beta)\} / [1+\exp\{2(u+s-\beta)\}] , \ \beta = 1 \text{ July 1986} .$$