Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 3051–3055, May 1986
Applied Mathematical Sciences

A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome

(length-biased sampling/truncated distribution)

KUNG-JONG LUI*†, DALE N. LAWRENCE‡, W. MEADE MORGAN*, THOMAS A. PETERMAN*, HARRY W. HAVERKOS*, AND DENNIS J. BREGMAN*

*Division of Viral Diseases and ‡Division of Host Factors, Center for Infectious Diseases, Centers for Disease Control, U.S. Department of Health and Human Services, Public Health Services, Atlanta, GA 30333

Communicated by Oscar D. Ratnoff, January 2, 1986

ABSTRACT **The incubation period, representing the interval between the date of exposure and the date of diagnosis, can be firmly ascertained in transfusion-associated cases of acquired immunodeficiency syndrome (AIDS). However, because the observation period of all transfusion-infected persons may be short compared with the average incubation period for AIDS, many cases with long incubation periods have not yet been diagnosed. Thus, the simple average of 2.6 years tends to underestimate the true mean. To correct for this underestimation bias, we assumed that the underlying distribution of the incubation periods is a member of a broad class of probability densities. Then, by maximum likelihood techniques, the mean incubation period for transfusion-associated AIDS was estimated to be 4.5 years, with the 90% confidence interval ranging from 2.6 to 14.2 years. The long incubation period has important consequences for infected individuals and implications for public health intervention and prevention policy.**

Transfusion-associated acquired immunodeficiency syndrome (AIDS) provides a unique opportunity to study the incubation period of this disease because the date of exposure to the virus can be accurately determined through epidemiologic investigations that identify which blood donor, and donation date, was associated with transmission of infection (1). We define the incubation period as the period from transfusion to the diagnosis of the first opportunistic disease associated with AIDS. The date of diagnosis is used as a closure for this period, rather than the more conventional date of onset of symptoms, because the earliest symptoms are typically nonspecific and their dates of onset are poorly remembered.

During the initial phase of a new disease epidemic, use of the simple average of the incubation periods of the observed cases is subject to two types of bias due to length-biased sampling. One bias can result from any omission of patients who had such short incubation periods that their illnesses went undiagnosed as AIDS before the existence of transfusion-associated AIDS was first recognized; this constitutes left censoring in our sample. As a consequence of this bias, the traditional estimate, the mean incubation periods of the observed cases, might tend to overestimate the true mean. On the other hand, another source of bias can arise from the fact that our sample does not include those exposed persons who will have such long incubation periods that they have not yet been diagnosed; this constitutes right censoring, which might cause the traditional estimate to underestimate the true mean.

Our estimate would now suggest that the right censoring exerts a greater impact on the estimate and the simple average tends to represent an underestimation. Moreover, immediately following a period when the rate of exposure is increasing, as certainly was occurring with AIDS during 1978–1984, the degree of underestimation due to the right censoring is even greater, since the absolute number of patients with short incubation periods is contributing proportionately more to the rising incidence number for diagnosed cases.

To correct for the biases of length-biased sampling in the traditional estimator, a general family of probability densities to describe the underlying distribution of incubation periods for transfusion-associated AIDS is assumed. Then, under the model assumptions, we obtain maximum likelihood estimates, which are asymptotically efficient and unbiased estimators of the mean incubation period (2, 3).

The influence of length-biased sampling upon other types of industrial and medical problems, such as the efficiency of estimation of mean life for electrical components, selection bias in a follow-up study, or a characteristic related to etiology, which would be subject to misleading conclusions without special considerations, is a subject of increasing interest (4–8).

Because a variety of chronic and neoplastic diseases are characterized by long intervals between documented exposure to an environmental agent and subsequent disease diagnosis, the proposed model-based approach may be applicable to other biomedical investigations.

METHODS

The 100 records of patients with transfusion-associated AIDS aged 13 years or older reported to the Centers for Disease Control (CDC) as of April 1, 1985, and diagnosed by December 31, 1984, were reviewed. The December cut-off point was used to ensure that reporting of cases was complete. Each of these cases met the CDC AIDS surveillance definition (1) and had no apparent risk factor other than their transfusion with one or more units of whole blood or blood components.

Cases with unknown dates of transfusion or dates of diagnosis were excluded from the analysis. As of the time of our analysis, 83 cases had undergone a sufficiently complete epidemiologic investigation to identify the date(s) of transfusion. Of these, 76 (92%) had developed AIDS following transfusion on a single date. Only 7 had received transfusions on more than one transfusion date prior to the development

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: AIDS, acquired immunodeficiency syndrome; HTLV-III/LAV, human T-lymphotropic virus type III/lymphadenopathy-associated virus.

†To whom reprint requests should be addressed at: Centers for Disease Control, Building 7, Room 109, 1600 Clifton Road, N.E., Atlanta, GA 30333.

of AIDS. For the patient who is known to have received blood from a suspect donor on only one of two occasions, we used that date, the more recent one.

In the remaining six cases in which there was no information as yet to indicate which date of transfusion was associated with the receipt of blood from a high risk donor, we used the date on which the larger number of blood units were transfused into the respective patient. We chose to use the transfusion date associated with the greater number of blood units for these six, although we realized that the risk associated with a transfused unit of blood was increasing, but at an unknown rate over the interval 1978–1984. Moreover, in only two cases was this choice associated with the earlier, rather than more recent, date and might result in a negligibly lengthened incubation period estimate. However, the consequence of the effect of using the earlier dates for these two and the overall effect of the six cases with multiple transfusions will still be examined as a part of our analysis.

Let t_i be the incubation period in months for the ith reported case. With this notation, the traditional estimator of the mean incubation period u is the simple average:

$$u = \sum_{i=1}^{n} t_i/n, \qquad [1]$$

where n is the total number of reported cases.

In the model-based approach, we assume that t_i follows an as-yet-unspecified probability density $f(t_i, \lambda)$. Let TR_i be the number of months between the transfusion and December 31, 1984, the last date of diagnosis being considered for this study. Let TL_i be the number of months between the transfusion and June 1, 1982, the month when the first recognized case of transfusion-associated AIDS was diagnosed. Then, the corresponding truncated density (9, 10), $f(t_i, \lambda)/F(T_i, \lambda)$, can be used to describe the distribution of the ith case incubation period, where

$$F(T_i, \lambda) = \int_{\delta_i TL_i}^{TR_i} f(t, \lambda)dt, \text{ where } T_i = (TL_i, TR_i, \delta_i),$$

$$\text{where } \delta_i = \begin{cases} 1, \text{ if } i\text{th case with transfusion date before June 1982, and} \\ 0, \text{ otherwise.} \end{cases}$$

Table 1. Cases of transfusion-associated AIDS with diagnosis on or before December 31, 1984

Parameter	Incubation period
n	83
Mean	30.7
SD	14.6
Range	1–65
Median	29

The incubation period is expressed in months from transfusion to diagnosis.

The likelihood of the observations is then

$$\prod_{i=1}^{n} f(t_i, \lambda)/F(T_i, \lambda), \qquad [2]$$

from which the maximum likelihood estimates of λ can be obtained.

In the likelihood given by Eq. 2, the distribution of incubation period $f(t_i, \lambda)$ can be taken to be any appropriate probability density. However, without the specific information to suggest a particular density, the Weibull family of densities, which are rich in shapes, has been focused, because this family has been successfully applied to describe waiting times or durations of epidemics and covers a variety of epidemic curves (9–13). The results for some other parametric models are presented in the *Appendix*. The form of the Weibull family and the maximum likelihood procedure of estimating its parameters can be found in the texts (9, 14) and are not given here.

Finally, to check the fit of the Weibull model, a χ^2 goodness-of-fit test was used.

RESULTS

The summary descriptions of the characteristics of the 83 cases for which diagnosis and transfusion dates were known in the study are summarized in Table 1. The incubation periods for these cases are plotted by year of transfusion (Fig. 1). Note that the mean of the incubation periods becomes progressively shorter as the year of transfusion becomes more recent, from 63 months for 1978 to 5.5 months for 1984. This trend illustrates the bias inherent in the traditional

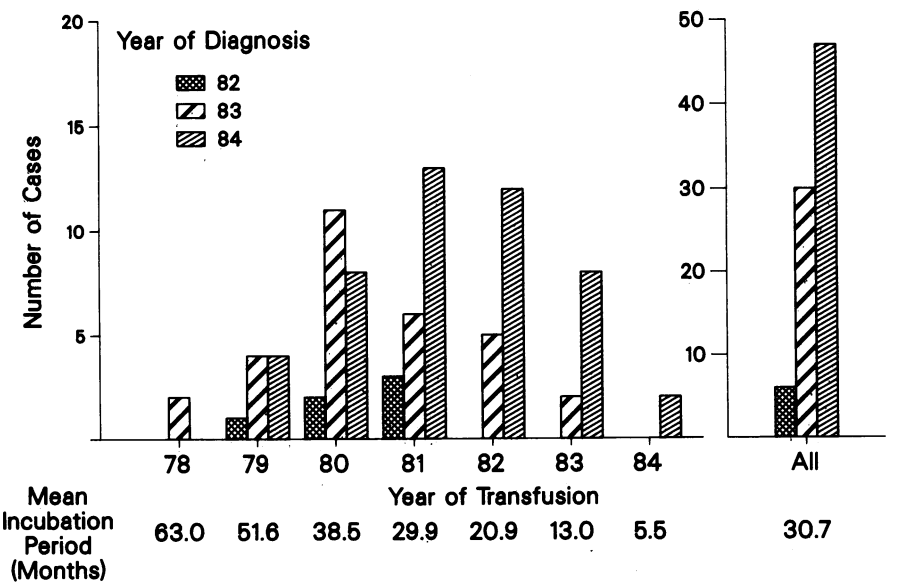

FIG. 1. Cases of transfusion-associated AIDS: Year of diagnosis by year of transfusion.

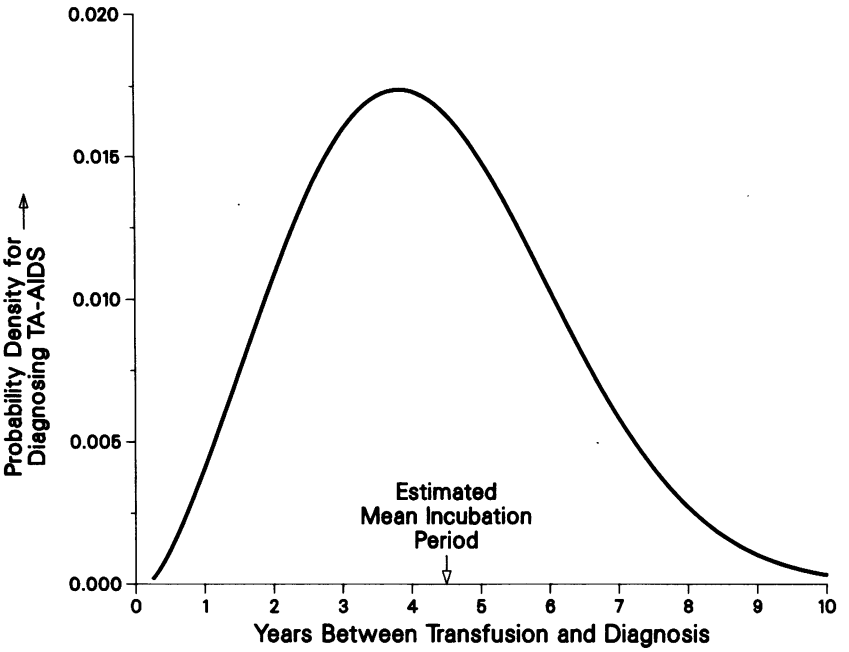

FIG. 2. Estimated distribution of the incubation time for transfusion-associated AIDS (TA-AIDS) based on transfusion cases diagnosed through December 1984.

estimator. The traditional mean estimate for the incubation period is 31 months, and the range is from 1 to 65 months.

In the model-based approach, the maximum likelihood estimate of the mean incubation periods gave 54 months (4.5 years) with an approximate lower 90% confidence bound of 2.6 years and an upper bound of 14.2 years. In the resulting distribution derived from the model-based approach (Fig. 2), the median is equal to 52 months. Based on this distribution, we infer that even 4 years after population exposure to human T-lymphotropic virus type III/lymphadenopathy-associated virus (HTLV-III/LAV), there is still a high proportion of cases that have not been diagnosed.

If we had excluded the two cases with more units of blood transfused on the more distant dates in our calculation, then the maximum likelihood estimate was 53 months (4.4 years) with a 90% confidence interval (2.6 years, 13.7 years). The results were very close to the previous estimates. Furthermore, excluding all of the six cases with multiple transfusions led to a modest increase in the maximum likelihood estimate of the mean incubation period to 60 months (5 years) with a 90% confidence interval (2.3 years, 19.6 years).

Since very few cases were transfused in the same month, we collapsed the data into annual exposure before applying the goodness-of-fit test (Fig. 1). The observed number of cases and expected number under the assumed model agree quite well (Table 2). The χ^2 goodness-of-fit test indicates that the Weibull model adequately describes the observed data.

DISCUSSION

AIDS has several biologically and epidemiologically relevant intervals following exposure to the agent. There may be an acute illness after which a clinically quiescent state ensues (15). The virus may be cultured from the infected host during this time, in spite of concurrent seropositivity to HTLV-III/LAV (16). The onset of symptoms represents another milestone but is typically very vague. Each of these intervals is of considerable importance and warrants a specific term.

The exposure-to-diagnosis interval is the most definitive one available at present and transfusion-associated AIDS patients provide the largest data base for analysis for such an interval. Rather than employ the term latency, or prepatent period, we chose to specify our intent and use incubation period, to conform with other applications in infectious disease epidemiology (John Last, personal communication).

The number of asymptomatic, infected potential blood donors was exponentially increasing early in the epidemic, when the number of actual AIDS cases was small. When studied an average of 28 months after donation, most of the donors implicated in transmission of transfusion-associated AIDS were HTLV-III/LAV culture-positive (16). Undoubtedly, until specific recommendations for the donation of blood were formulated in 1983 (17), the likelihood of receiving a HTLV-III/LAV-infected unit of blood was increasing

Table 2. Diagnosis distribution of observed cases reported as of April 1, 1985, and the expected cases based on the resulting Weibull model

Treatment year	Diagnosis year		
	1982*	1983	1984
1978			
Observed	0	2	0
Expected	0	1	1
1979			
Observed	1	4	4
Expected	2	4	3
1980			
Observed	2	11	8
Expected	4	9	8
1981			
Observed	3	6	13
Expected	3	8	11
1982			
Observed	0	5	12
Expected	1	6	10
1983			
Observed		2	8
Expected		2	8
1984			
Observed			2
Expected			2

*There were only 6 months of observation period in 1982.

annually, and the size of the infected, transfusion recipient cohort each year (1978–1983) was also increasing.

However, the proportion of the recipients who will eventually develop AIDS cannot be easily predicted because of the small number who have had time to develop symptoms and be diagnosed. Recent serologic testing of various U.S. populations to detect antibody to HTLV-III/LAV has resulted in widely varying estimates for the prevalence of infection (18). Even if only a small percentage of these individuals donated blood during a period when the blood was infectious, the total number of transfusion-associated AIDS cases could increase considerably. The validity of the proposed model-based approach is not influenced by changes in incidence of transfusion-associated AIDS. A better estimate will be obtained, of course, as more cases are identified in the future.

The maximum likelihood approach to estimating the mean incubation period overcomes the underestimation biases inherent in the traditional estimator when there is only right censorship. In our model-based approach, we truncate the right side of the incubation distribution to remove the observation bias resulting from cases with incubation periods so long that they have not yet been diagnosed. Similarly, we truncate the left side of the incubation distribution, because there may also have been some transfusion-associated AIDS cases with transfusions in the late 1970s and the early 1980s that were not properly diagnosed before transfusion-associated AIDS was reported in 1982. However, our estimate of the mean incubation period assumes that the degree of recognition of transfusion-associated AIDS has been uniform over the observation period June 1982–December 1984. Although we believe this to be a reasonable assumption, if recognition of cases has increased substantially over this period due to increased awareness of the risk of contracting AIDS from blood and blood products, then our estimate may tend to be large. Further analysis can be done as more data are collected.

If the exposure-to-diagnosis interval were only 2.6 years, as given by the traditional estimator, the reported incidence during early 1985 of cases linked to transfusions on or before December 31, 1981, would now be declining, and the incidence of AIDS cases linked to transfusions in 1982 should now be reaching a peak. There is no evidence of such a decline or peak (James Allen, personal communication). Furthermore, the width of the confidence interval associated with our estimate is an additional indication that the true mean incubation period is much larger than the observed mean, 2.6 years. Otherwise, the confidence interval would be much narrower, since about 50% of the cases in our data set had incubation periods longer than 2.6 years.

Several other sources of information indicate that our model-based estimate is reasonable. Based on a 6-year follow-up of a group of homosexual men enrolled in 1978 in a hepatitis B vaccine study, the median interval between the first collection of serum that was shown to be serologically positive for HTLV-III/LAV and the subsequent diagnosis of AIDS in the study participants was shown to be 43 months (19). This is a minimum estimate, since several persons were undoubtedly already seropositive for an unknown period prior to the first collection.

Recent analysis of HTLV-III/LAV RNA sequence homology and morphologic characteristics indicates a similarity to visna virus, a recognized member of the slow virus (lentivirus) family (20). Cases of the unconventional viral disease of humans, kuru, have been diagnosed more than two decades following the last likely exposure (21). Although our results cannot be used to establish a maximum incubation period for transfusion-associated AIDS, they do suggest that the chances of becoming a case subsequent to a known exposure will peak and eventually decline. This conclusion is obtained by testing the hypothesis, the shape parameter, $r = 1$ in the Weibull model. Weibull curves with $r = 1$, known as negative exponential densities, would describe the incubation period if the probability of developing AIDS at some point following exposure to the virus was subject to random event (22). However, our data indicate that r is significantly larger than 1 ($z = 7.6$; $P < 0.0001$), suggesting that the incidence of AIDS does peak some years following infection.

Our estimate of the mean incubation period must be understood as potentially generalizable at present only to adult patients with blood transfusion-transmitted cases. Not all patients exposed to HTLV-III/LAV infection through transfusion live long enough to develop overt AIDS, given the 4.5-year average described here. Many will die in the interim from the disease that required the transfusion. No ideal population for the study of the incubation period of AIDS is available since there is no population that is a cohort of healthy adults, infected on a documented date, for whom exquisite medical care eliminates deaths due to any competing cause. Our estimate is still valid as long as we are only concerned with the adult transfusion recipients who live long enough to develop AIDS.

The long incubation period also has important implications for public health intervention and for necessary planning to accommodate the medical and community resources needed. We cannot await decreases in the number of new cases to evaluate the success of the prevention procedures and educational programs instituted. Instead, serologic testing to detect the changing prevalence of HTLV-III/LAV antibody might offer a much faster way of evaluating the effectiveness of the programs in curtailing the acquisition of new infections. Furthermore, seropositive patients should be followed closely, because a substantial number would not be expected to develop disease until many years after exposure. Serious consideration may need to be given to therapeutic intervention before the overt onset of disease.

APPENDIX

If the incubation period of transfusion-associated AIDS follows a log-logistic density—i.e.,

$$f(t, \boldsymbol{\lambda}) = (\lambda p)(\lambda t)^{p-1}/[1 + (\lambda t)^p]^2,$$

then the median is 77 months and the mode is 52 months. By the delta method, the mean can also be approximated by 77 months. The model is a little inferior to the Weibull model, because the log-likelihood under the former model is smaller than that under the latter one.

Furthermore, if the incubation period follows a Gompertz density—i.e.,

$$f(t, \boldsymbol{\lambda}) = \lambda \exp(rt)\exp\{\lambda[1 - \exp(rt)]/r\},$$

then the median is 46 months and the mode is 51 months. The mean is calculated as 45 months. The log-likelihood of the Gompertz model is even smaller than that of the log-logistic model.

We thank J. Curran, H. Jaffe, J. Allen, E. Hall, M. Haber, J. Jason, B. Evatt, D. Culver, and A. Kendal for their assistance in providing data, helpful comments, and administrative support.

1. Curran, J. W., Lawrence, D. N., Jaffe, H. W., Kaplan, J. E., Zyla, L. D., Chamberland, M., Weinstein, R., Lui, K. J., Schonberger, L. B., Spira, T. J., Alexander, W. J., Swinger, G., Ammann, A., Solomon, S., Auerbach, D., Mildvan, D., Stoneburner, R., Jason, J. M., Haverkos, H. W. & Evatt, B. L. (1984) *N. Engl. J. Med.* **310**, 69–75.
2. Rao, C. R. (1965) *Linear Statistical Inferences and Its Application* (Wiley, New York), pp. 353–366.

3. Hoel, P. G., Port, S. C. & Stone, C. J. (1971) *Introduction to Statistical Theory* (Houghton Mifflin, Boston), pp. 13–26.
4. Blumenthal, S. (1967) *Technometrics* **9**, 205–218.
5. Singh, S. N., Bhattacharya, B. N. & Yadava, R. C. (1979) *J. Am. Stat. Assoc.* **74**, 917–920.
6. Zelen, M. & Feinleib, M. (1969) *Biometrika* **56**, 601–614.
7. Zelen, M. (1974) in *Reliability and Biometry, Statistical Analysis of Life Length*, eds. Proschan, F. & Serfling, R. J. (Society for Industrial and Applied Mathematics, Philadelphia), pp. 701–726.
8. Simon, R. (1980) *Am. J. Epidemiol.* **111**, 444–452.
9. Gross, A. J. & Clark, V. A. (1975) *Survival Distribution: Reliability Applications in the Biomedical Science* (Wiley, New York), pp. 97–132.
10. Cox, D. R. & Oakes, D. (1983) *Analysis of Survival Data* (Chapman and Hall, London), pp. 177–179.
11. Kalbfleish, J. D. & Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data* (Wiley, New York).
12. Johnson, N. L. & Katz, S. (1970) *Continuous Univariate Distribution-1 Distribution in Statistics* (Wiley, New York), pp. 250–271.
13. Olkin, I., Gleser, L. J. & Derman, C. (1980) *Probability Model and Application* (Wiley, New York), pp. 276–285.
14. Jennrich, R. J. & Ralston, M. L. (1978) *Fitting Nonlinear Models to Data* (Univ. of Calif. Health Sci. Computing Facility, Los Angeles), BMDP Tech. Rep. 46.
15. Cooper, D. A., Gold, J., MacLean, P., Donovan, B., Finlayson, R., Barnes, T. G., Michelmore, H. M., Brooke, P. & Penny, R. (1985) *Lancet* **i**, 537–540.
16. Feorino, P. M., Jaffe, H. W., Palmer E., Peterman, T. A., Francis, D. P., Kalyanaraman, V. S., Weinstein, R. A., Stoneburner, R. L., Alexander, W. J., Raevsky, C., Getchell, J. P., Warfield, D., Haverkos, H. W., Kilbourne, B. W., Nicholson, J. K. A. & Curran, J. W. (1985) *N. Engl. J. Med.* **312**, 1293–1296.
17. Centers for Disease Control (1983) *Morbid. Mortal. Wkly. Rep.* **32**, 101–104.
18. Landesman, S. H., Ginzburg, H. M. & Weiss, S. H. (1985) *N. Engl. J. Med.* **312**, 521–525.
19. Jaffe, H. W., Darrow, W. W., Echenberg, D. F., O'Malley, P. M., Getchell, J. P., Kalyanaraman, V. S., Byers, R. H., Drennan, D. P., Braff, E. H., Curran, J. W. & Francis, D. P. (1985) *Ann. Intern. Med.* **103**, 210–214.
20. Gonda, M. A., Wong-Staal, F., Gallo, R. C., Clements, J. E., Narayan, O. & Gilden, R. V. (1985) *Science* **227**, 173–177.
21. Prusiner, S. B., Gajdusek, D. C. & Alpers, M. P. (1982) *Ann. Neurol.* **12**, 1–9.
22. Bailey, N. T. J. (1964) *The Elements of Stochastic Processes with Applications to the Natural Sciences* (Wiley, New York), pp. 66–69.