A Model-Based Estimate of the Mean Incubation Period for AIDS in Homosexual Men

KUNG-JONG LUI, WILLIAM W. DARROW, GEORGE W. RUTHERFORD, III

Because of the difficulty in identifying the date of exposure to type 1 of the human immunodeficiency virus (HIV-1) infection in persons other than transfusion recipients, studies of the incubation periods for acquired immunodeficiency syndrome (AIDS) have been limited. When data from a cohort of 84 homosexual and bisexual men that provided the information to determine the years of conversion of sera infected with HIV-1 were analyzed, a model for the proportion likely to develop AIDS and the incubation period for AIDS in homosexual men could be derived. The maximum likelihood estimate for the proportion of infected homosexual men developing AIDS is 0.99 (90% confidence interval ranging from 0.38 to 1). Furthermore, the maximum likelihood estimate for the mean incubation period for AIDS in homosexual men is 7.8 years (90% confidence interval ranging from 4.2 years to 15.0 years), which is close to the estimate of 8.2 years for adults developing transfusion-associated AIDS.

IT HAS BEEN HYPOTHESIZED THAT THE incubation period, the time interval from infection with type 1 of the human immunodeficiency virus (HIV-1) to the date of diagnosis with acquired immunodeficiency syndrome (AIDS), could differ from one risk group to another (*1*). Because the infection date is difficult to determine in most AIDS risk groups other than those with transfusion-associated AIDS (TA-AIDS), the study of the incubation period has mainly been limited to TA-AIDS (*2–8*). Furthermore, no data are available on the number of persons infected through transfusions who have not yet developed AIDS. Therefore, the proportion of infected transfusion recipients who will ultimately develop AIDS cannot be estimated. However, because conversion of sera usually occurs within a few months of infection with HIV-1 (*9, 10*), data from a cohort of 84 infected homosexual and bisexual men who had a positive HIV-1 antibody test within 12 months of a negative antibody test can be used to estimate both the mean incubation period and the proportion of infected individuals who will develop AIDS.

The sample consists of 84 men drawn from a larger cohort of 6709 homosexual and bisexual men enrolled at San Francisco City Clinic between 1978 and 1980 for studies of hepatitis B (*11–13*). The 84 men include 83 men who were selected at random or returned for hepatitis B vaccine follow-up, could be located, and gave written consent for their stored sera to be tested for HIV-1 antibody, and one man who died from AIDS in 1982. For all of them, serum samples were available to document the date of seroconversion within 12 months of a negative HIV-1 antibody test. Antibody to HIV-1 was measured by enzyme-linked immunosorbent assay (ELISA). Equivocal samples were retested by ELISA, and weakly positive ELISA samples were further confirmed by a Western blot assay. Both the sensitivity and the specificity of ELISA test are higher than 0.99 (*14*). The year of HIV-1 seroconversion was estimated by the midpoint between dates of a negative and a subsequent positive HIV-1 antibody test. By January 1987, 21 (25%) infected men had been diagnosed with AIDS that met the Centers for Disease Control (CDC) surveillance definition (*1*). The San Francisco cohort has been more fully described elsewhere (*13*).

Notations are defined first for our model. Let t_i denote the time interval in years from seroconversion to diagnosis for the i_{th} reported case ($i = 1, 2, \ldots, n$). Similarly, let T_i be the number of years between the seroconversion and 1986, the last year of diagnosis under consideration for this study. We assume that t_i follows a probability density $f(t_i, \theta)$, where θ is an unknown vector of parameters to be estimated. Let p be the proportion of infected individuals who will eventually develop AIDS. The likelihood of the observations is then

$$\prod_{i=1}^{n} [p\, f(t_i, \theta)]^{\delta_i} [1 - p + p\, F(T_i, \theta)]^{1-\delta_i}$$

where

$$F(T_i, \theta) = 1 - \int_0^{T_i} f(t, \theta)dt$$

where δ_i is 1 if the i_{th} case was diagnosed by the end of 1986 and 0 otherwise. Here, $f(t_i, \theta)$ can be assumed to be any appropriate probability density function. Since the Weibull family has been widely applied to describe the incubation period for TA-AIDS (*2–6*), we focus $f(t_i, \theta)$ to have the form

$$f(t, \theta) = \lambda\gamma t^{\gamma-1} \exp(-\lambda t^{\gamma})$$

where λ and γ are greater than 0. A detailed description of the characteristics for this probability density can be found elsewhere (*15, 16*). However, the results for other distributions, such as the log-logistic have also been presented briefly. Based on the above likelihood, the maximum likelihood estimates of the parameters λ, γ, and p can be obtained by using a derivative-free procedure (*17*). For investigating the reliability of the estimate of p, a table that summarizes the log-likelihood and the corresponding estimates of the mean incubation period for different values of p is included. To check the fit of the Weibull model, we include a table that summarizes the observed and expected distributions of AIDS cases in the 84 men. The procedure for calculating the expected frequency can be found in Lui *et al.* (*4*).

The sample mean of the incubation periods for the 21 men who developed AIDS is 4.8 years (SD, 1.5 years), and the range is 1 to 8 years. The maximum likelihood estimate of p is 0.99 with a 90% confidence interval (0.38, 1.00), calculated by the likelihood-ratio test (*18–20*). However, as Table 1 shows, different values of p do not change the log-likelihood much. Therefore, it may still be too early to predict the exact proportion of infected homosexual and bisexual men who will develop AIDS. The maximum likelihood estimates for λ and γ are $\hat{\lambda} = 0.0038$ and $\hat{\gamma} = 2.571$, respectively. Substituting $\hat{\lambda}$ and $\hat{\gamma}$ for the corresponding

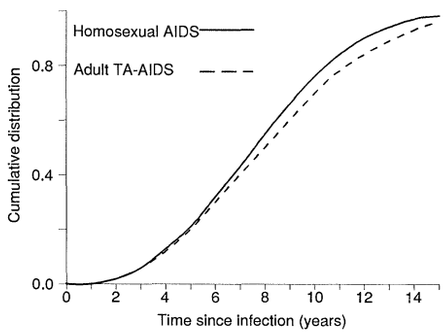

Fig. 1. The cumulative distribution of incubation periods for AIDS in homosexual and bisexual men (solid line) and transfusion-associated (TA) AIDS (dashed line). Both curves are based on the Weibull distribution $[1 - \exp(-\lambda t^{\gamma})]$, where λ and γ are parameters, and t is the time between infection and diagnosis. The maximum likelihood estimates for AIDS in homosexuals and TA-AIDS in adults (*2*) are, respectively, $\hat{\lambda} = 0.003807$, $\hat{\gamma} = 2.571$ and $\hat{\lambda} = 0.004799$, $\hat{\gamma} = 2.396$.

K.-J. Lui, Biometrics Branch, Division of Injury Epidemiology and Control, Centers for Disease Control, Atlanta, GA 30333.
W. W. Darrow, AIDS Program, Center for Infectious Diseases, Centers for Disease Control, Atlanta, GA 30333.
G. W. Rutherford, III, AIDS Office, Department of Public Health, San Francisco, CA 94103.

Table 1. The relations among log-likelihood, values of p, and the corresponding estimates of the mean incubation period.

Log-likelihood	Values of p	Estimated mean (years)
−74.103	1.00	7.8
−74.106	0.90	7.4
−74.119	0.80	6.9
−74.156	0.70	6.4
−74.258	0.60	5.9
−74.540	0.50	5.5
−75.304	0.40	5.1

parameters in the formula $\Gamma(1+1/\gamma)/\lambda^{1/\gamma}$ gives the estimate 7.8 years of the mean incubation period, with a 90% confidence interval of 4.2 years to 15.0 years. The cumulative distribution of the incubation period for the resulting model is shown in Fig. 1. This distribution is similar to that for TA-AIDS in adults (*2*). The observed and expected marginal frequencies in the last row of Table 2 agree reasonably well. The corresponding estimates of p and the mean incubation period for the log-logistic are presented in (*21*).

Since seroconversion usually occurs within a few months of infection (*9, 10*), using the year of seroconversion as a proxy for the year of infection would tend to produce an underestimate of the true incubation period. But since AIDS has such a long mean incubation period (of the order of 7 to 8 years), the magnitude of this underestimate is probably negligible.

The 84 men included in this analysis, compared with all other 515 HIV-1 seropositive men who were selected at random from the entire cohort of 6709 homosexual and bisexual men but about whom we had no information to determine the year of seroconversion, tended to be younger (mean = 31.6, SD = 5.2 versus mean = 34.3, SD = 5.6), to have fewer sex partners per year (mean = 49.5, SD = 60 versus mean = 99.3, SD = 164), and to be less likely to have been diagnosed with syphilis (odds ratio = 0.5; 95% confidence interval = 0.3 to 1). If younger men with fewer partners and fewer intercurrent infections with other sexually transmitted diseases are less likely to develop AIDS or to have longer incubation periods, the estimate of p may be conservative, but the estimate of the mean incubation period for these 84 individuals may be an overestimate. Further analysis can be done as more data are collected. Also, because AIDS was not recognized until 1981, some men with short incubation periods for AIDS could have been missed if they died before 1981. This sampling bias can cause us to overestimate the mean incubation period. However, since AIDS cases are extremely rare before 1981, and information on the 84 men is complete (even before 1981) in the data set, the bias resulting from misclassification should be minimal.

The maximum likelihood estimate of p is 0.99, which implies that almost all HIV-1–infected homosexual men will eventually develop AIDS if they do not die from other causes. This result is consistent with the argument of a close kinship between HIV-1 and animal lentiviruses (*22–24*). Penetrance and mortality for sheep infected by visnamaedi virus are found to be up to 1.00 (*25*). Therefore, if HIV-1 is comparable to related viruses in this respect, a good estimate of p should be close to 1. In fact, in a recent West German study, investigators made a similar inference and left little hope that the majority of HIV-1 carriers would not progress to AIDS or to other fatal manifestations (*26, 27*) in the absence of an effective therapy. On the other hand, since AIDS has such a long mean incubation period, approximately 8 years, which is longer than the follow-up time for most of our cases, infected men who have not yet developed AIDS cannot be physically distinguished from infected men who will never develop AIDS. Thus, the estimate of p should still be treated cautiously. Note also that the estimate of the mean incubation period decreases as the value p decreases (Table 1). This is because infected individuals who have not been diagnosed with AIDS by now are more likely to represent persons who will never develop AIDS when p is smaller. In fact, a similar discussion on the limitations of the application of limited-failure models to other topics has been noted elsewhere (*18*).

The maximum likelihood estimate of the mean incubation period is 7.8 years, which is similar to the estimate of the mean based on TA-AIDS for adults (*2*). Both distributions (Fig. 1) suggest that a very high proportion of men still have not been diagnosed even 5 years after infection with HIV-1. This result implies that the AIDS epidemic will continually increase for the next several years if no effective treatment becomes available to prevent the progression of HIV-1 infection from an asymptomatic stage to clinical AIDS. Note that the estimate of the shape parameter γ is greater than 1. This implies that the hazard rate for developing AIDS in homosexual men increases with the time interval after infection. Furthermore, the hazard rate increases at a faster rate over time, since $\hat{\gamma}$ is greater than 2. In contrast to the pediatric TA-AIDS (*4*), though the hazard rate also increases with time after infection, the slope of the hazard rate slows down.

Because of the possible difference in incubation periods among different groups at risk for AIDS, extrapolation of an incubation period from TA-AIDS to project the AIDS epidemic (*7*), in which the homosexual male risk group accounts for 74% of all AIDS cases in adults (*28*), is always questionable. However, our model provides no evidence to support the hypothesis of a major difference in incubation period between TA-AIDS in adults and AIDS in homosexual men: both have long mean incubation periods. The width of the confi-

Table 2. Observed and expected number of AIDS cases among 84 men by year of HIV-1 seroconversion and year of diagnosis.

Year of HIV-1 seroconversion		Year of diagnosis								Censored*	Total
		1979	1980	1981	1982	1983	1984	1985	1986		
1978	observed	0	0	0	1	0	1	1	0	3	6†
	expected	0.02	0.10	0.22	0.35	0.47	0.57	0.63	0.64		
1979	observed		0	0	0	0	0	0	1	7	8
	expected		0.01	0.04	0.09	0.15	0.20	0.24	0.27		
1980	observed			0	0	0	1	1	1	9	12
	expected			0.04	0.18	0.38	0.60	0.82	0.99		
1981	observed				0	2	2	1	5	19	29
	expected				0.18	0.88	1.87	3.00	4.07		
1982	observed					1	0	3	0	19	23
	expected					0.12	0.59	1.27	2.02		
1983	observed						0	0	0	2	2
	expected						0	0	0		
1984	observed							0	0	4	
	expected							0	0		4
Total	observed	0	0	0	1	3	4	6	7		21/84
	expected	0.02	0.11	0.30	0.80	2.00	3.83	5.96	7.99		

*Censored means that homosexual and bisexual men infected with HIV-1 had not been diagnosed with AIDS as of 1 January 1987. †For example, a total of six men showed seroconversion in 1978, one of the six was diagnosed in 1982, one in 1984, and one in 1985. The other three had not been diagnosed with AIDS by 1 January 1987.

dence interval indicates that our present data are not precise enough to pinpoint the mean. As additional data are collected, estimates of the characteristics of the incubation period for AIDS can be improved.

REFERENCES AND NOTES

1. J. W. Curran *et al.*, *N. Engl. J. Med.* **310**, 69 (1984).
2. C. F. Medley, R. M. Anderson, D. R. Cox, L. Billard, *Nature* **328**, 719 (1987).
3. K.-J. Lui *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3051 (1986).
4. K.-J. Lui, T. A. Peterman, D. N. Lawrence, J. R. Allen, *Stat. Med.* 7, 395 (1988).
5. K.-J. Lui, T. A. Peterman, D. N. Lawrence, *Nature* **329**, 207 (1987).
6. T. A. Peterman, K.-J., Lui, D. N. Lawrence, J. R. Allen, *Transfusion* 27, 371 (1987).
7. M. Rees, *Nature* **326**, 343 (1987).
8. D. E. Barton, *ibid.*, p. 734.
9. H. J. Alter *et al.*, *Science* **226**, 549 (1984).
10. D. P. Francis, *et al.*, *Lancet* **ii**, 1276 (1984).
11. M. T. Schreeder *et al.*, *J. Infect. Dis.* **146**, 7 (1982).
12. D. P. Francis *et al.*, *Ann. Intern. Med.* 97, 362 (1982).
13. H. W. Jaffe *et al.*, *ibid.* **103**, 210 (1985).
14. Centers for Disease Control, *Morbid. Mortal. Weekly Rep.* **36** (no. 2S), 13 (1987).
15. D. R. Cox and D. Oakes, *Analysis of Survival Data* (Chapman & Hall, London, 1984).
16. A. J. Gross, and V. A. Clark, *Survival Distribution: Reliability Applications in the Biomedical Science* (Wiley, New York, 1975).
17. M. Ralston, in *BMDP Statistical Software*, W. J. Dixon, Ed. (Univ. of California Press, Berkeley, 1981), pp. 305–329.
18. W. O. Meeker, *Technometrics* **29**, 51 (1987).
19. H. Chernoff, *Ann. Math. Stat.* **25**, 573 (1954).
20. P. I. Feder, *ibid.* **39**, 2044 (1968).
21. If the incubation period of AIDS in homosexual men follows a log-logistic density: $f(t, \theta) = (\lambda\gamma)(\lambda t)^{\gamma-1}/(1 + (\lambda t)\gamma)^2$, then the estimate $\hat{p}$ of the proportion of infected men who will eventually develop AIDS and the estimate of the median incubation period are 1.00 and 7.9 years, respectively. By the delta method, the mean can also be approximated by 7.9 years. The log-likelihood of this distribution is smaller than that for the Weibull distribution.
22. M. G. Koch, *AIDS—from Molecule to Pandemic* (Freeman, New York, 1987).
23. J. Seale, *J. R. Soc. Med.* **78**, 613 (1985).
24. M. G. Koch, *AIDS-Vår Framtid* (Svenska Carnegie Institutet, Stockholm, 1985).
25. P. A. Pålsson, in *Slow Virus Diseases in Animals and Man*, R. H. Kimberlin, Ed. (North-Holland, Amsterdam, 1976), pp. 17–43.
26. H. R. Brodt *et al.*, *Dtsch. Med. Wochenschr.* **111**, 1175 (1986).
27. J. J. González and M. G. Koch, *Am. J. Epidemiol.* **126**, 985 (1987).
28. *AIDS Weekly Surveillance Report—United States* (Centers for Disease Control, Atlanta, 1 September 1987).

29. We thank W. M. Morgan, H. Jaffe, R. Byers, J. Jason, D. McGee, E. Davis, and M. Kirk for their valuable comments, administrative support, and graphic assistance.

12 November 1988; accepted 5 April 1988

Location and Chemical Synthesis of a Binding Site for HIV-1 on the CD4 Protein

BRADFORD A. JAMESON, PATRICIA E. RAO, LILLY I. KONG, BEATRICE H. HAHN, GEORGE M. SHAW, LEROY E. HOOD, STEPHEN B. H. KENT

The human immunodeficiency virus type 1 (HIV-1) uses the CD4 protein as a receptor for infection of susceptible cells. A candidate structure for the HIV-1 binding site on the CD4 protein was identified by epitope mapping with a family of eight functionally distinct CD4-specific monoclonal antibodies in conjunction with a panel of large CD4-derived synthetic peptides. All of the seven epitopes that were located reside within two immunoglobulin-like disulfide loops situated between residues 1 and 168 of the CD4 protein. The CD4-specific monoclonal antibody OKT4A, a potent inhibitor of HIV-1 binding, recognized a site between residues 32 and 47 on the CD4 protein. By analogy to other members of the immunoglobulin superfamily of proteins, this particular region has been predicted to exist as a protruding loop. A synthetic analog of this loop (residues 25 to 58) showed a concentration-dependent inhibition of HIV-1–induced cell fusion. It is proposed that a loop extending from residues 37 to 53 of the CD4 protein is a binding site for the AIDS virus.

THE AIDS VIRUS, HIV-1, SHOWS A preferential tropism toward helper T cells. The gp120 envelope protein of HIV-1 directly interacts with the CD4 protein of helper T cells (*1, 2*). Cell surface expression of the CD4 protein on a nonpermissive cell line (HeLa cells) can enable the virus to bind and infect this cell type (*3*). Monoclonal antibodies to the CD4 protein can prevent HIV-1 from infecting permissive cells (*1, 2, 4, 5*), and soluble constructs of the CD4 protein are able to inhibit HIV-1–induced cell fusion (*6–10*).

The CD4 protein serves as a phenotypic marker for the separation of helper T cells from lymphocyte populations and plays an essential role in the recognition by helper T cells of foreign antigen in the context of the polymorphic class II major histocompatibility complex (MHC) (*11*). The CD4 protein is encoded by a single gene and is of invariant sequence, in contrast to the class II MHC molecules or the T cell antigen receptor. It is therefore thought that the structure or structures recognized by the CD4 protein on the MHC is conserved in spite of the polymorphic nature of the MHC. This argument appears to apply to the HIV-1 gp120 as well. The region (or regions) on the gp120 that have been implicated in binding to CD4 reside within highly variable portions of the gp120 (*12, 13*), yet all known isolates of HIV-1, as well as HIV-2, appear to use the CD4 protein as a receptor and show the same inhibition profiles in response to CD4-specific monoclonal antibodies (*14, 15*). Furthermore, CD4-specific monoclonal antibodies (mAbs) that block the binding of HIV-1 also inhibit the presentation of foreign antigen and the subsequent stimulation of helper T cells (*16–23*). This suggests that the same site on CD4 is used both for binding to the class II MHC and to the gp120 of HIV. Thus, identification and characterization of this binding domain on the CD4 is not only important for understanding the virus/receptor interaction, but will aid in the basic understanding of interactions of antigen-presenting cells with helper T cells.

In these studies we used a panel of eight CD4-specific mAbs (Table 1). Antibodies MT-151, OKT4F, and OKT4A are the most potent inhibitors of HIV-1 binding to the CD4 protein, while OKT4D and OKT4E show intermediate inhibitory effects, OKT4B shows only a weak effect, and OKT4 and OKT4C show no effect on virus binding (*4, 5, 24*). Sattentau *et al.* (*5*) suggested that two independent binding domains can be discerned by this panel of CD4-specific mAbs, one defined by OKT4A and the other by MT151.

To map the epitopes of the CD4-specific mAbs, we synthesized a panel of CD4-derived peptides (see Fig. 1). We have found that longer peptides (approximately 30 amino acids in length) give more consistent results in direct binding assays than do shorter peptides (6 to 12 amino acids in length) and that antibodies to the longer peptides cross-react better with the native target protein. We therefore used peptides of about 30 amino acids in length. The CD4-specific mAbs were directly assayed for their ability to bind to each synthetic peptide (Fig. 2). Only three of the mAbs consistently recognized any of the CD4-derived peptides in the solid-phase radioimmunoassay (RIA). OKT4C and OKT4F both

B. A. Jameson, L. E. Hood, S. B. H. Kent, Division of Biology, California Institute of Technology, Pasadena, CA 91125.
P. E. Rao, Biotech Division, ORTHO Pharmaceuticals, Raritan, NJ 08869.
L. I. Kong, B. H. Hahn, G. M. Shaw, Division of Hematology/Oncology and Departments of Biochemistry and Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294.

*To whom correspondence should be addressed.