MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **2**, Number **2**, April **2005**

http://www.mbejournal.org/

pp. **263–277**

INCUBATION-TIME DISTRIBUTION IN BACK-CALCULATION APPLIED TO HIV/AIDS DATA IN INDIA

Arni S.R. Srinivasa Rao [1]

Centre for Ecological Sciences, Indian Institute of Science
Bangalore, 560 012. INDIA

Masayuki Kakehashi

Institute of Health Science, Faculty of Medicine, Hiroshima University
Hiroshima, 734-8551, JAPAN

(Communicated by Patrick Nelson)

Abstract. In this article, HIV incidence density is estimated from prevalence data and then used together with reported cases of AIDS to estimate incubation-time distribution. We used the deconvolution technique and the maximum-likelihood method to estimate parameters. The effect of truncation in hazard was also examined. The mean and standard deviation obtained with the Weibull assumption were 12.9 and 3.0 years, respectively. The estimation seemed useful to investigate the distribution of time between report of HIV infection and that of AIDS development. If we assume truncation, the optimum truncating point was sensitive to the HIV growth assumed. This procedure was applied to US data for validating the results obtained from the Indian data. The results show that method works well.

1. **Introduction.** Information on accurate population sizes of HIV-infected persons and AIDS cases and the trend of these figures are requisite to the planning of preventive policies and public-health management. Sophisticated statistical models have been developed to facilitate provision of the information. Among the models, a simple extrapolation method is easy to apply and useful for summarizing the trend of the spread of infection, but it is difficult to clarify how long the obtained trend stays unchanged. By comparison, mathematical models of the spread of sexually transmitted diseases use information on sexual behavior in the population to investigate the effect of behavioral change caused by a preventive program. But mathematical models usually require detailed information on sexual behavior in the population, which is not always available. In contrast, the back-calculation method connects infection with HIV and the development of AIDS to incubation-time. Because of to the long incubation period, this method can provide a very reliable prediction of future AIDS development from present HIV data.

Traditionally, back-calculation method is applied to estimate past HIV trends and to predict future AIDS cases by using reported AIDS cases and the assumed

2000 *Mathematics Subject Classification.* 92A15,62A10.

Key words and phrases. Maximum likelihood method, back-calculation, incubation time distribution.

[1]Corresponding author's address: Department of Mathematics and Statistics, University of Guelph, Toronto, Canada.

incubation time distribution. Information on HIV incidence is not directly used in the attempt. This is to be expected where detailed information on AIDS cases and incubation time may be more easily obtained than figures about HIV incidence. But there is another situation in which information on HIV incidence is more available than information about incubation time. This is likely when HIV surveillance is started but medical treatment is not generally available or is inadequate. In any case it is useful and helpful to use all these data to obtain more reliable outcomes especially when the quality of each kind of data is insufficient. A recent attempted to effort to take advantage of the information on HIV in back-calculation is made [1], but it required more detailed information on reported HIV.

Although the application of sophisticated methods to HIV/AIDS data was delayed, recently it has begun. The projection of AIDS [2] is useful, if all India level transmission probabilities are available. However, in a population where the dependable data as mentioned above are not accessible, researchers could assume a reasonable set of scenarios for the behavioral and epidemiological parameters, so that the scope of the epidemic could be determined. Modeling of this kind may not be explicit, but it is important to note that such a model guides one to predict the scope of the epidemic in the future, until dependable data become available. The popular back-calculation approach [3, 4] that assumes the distribution of incubation time of AIDS to be known and then through convolution projects the AIDS cases is extensively used by researchers. Longitudinal studies on HIV-infected individuals with reliable infection dates are necessary to ascertain the incubation period of AIDS. Unavailability of infection dates in India causes problems of left censoring; methods to deal with such situations were developed and discussed in [5].

In such a situation, application of mathematical models is difficult to carry out. Dynamic transmission models suggested so far have focused on the rate of new infection in a population [6], and these also emphasized mixing patterns of the uninfected and infected individuals in a population [7]; for review see Valerie Isham [8]. Masayuki Kakehashi used a novel application of such models by incorporating realistic epidemiological parameters [9, 10], and features of epidemiological models in a basic and lucid way [11]. Network models studied the spread of sexually transmitted infections and methods were developed to estimate basic reproductive rates [12, 13]. A recently proposed deterministic model that recently proposed [14] could be a dynamic way of estimating growth of disease in specific situations where information on sexual behavior is available. Moreover, the Indian population is relatively free of AIDS therapy, which reduces the complexity of the analysis.

Thus our analysis uses HIV/AIDS data from India and focuses on estimating the incubation-time distribution from the reported AIDS cases and externally estimated HIV incidence density, in contrast to the traditional back-calculation method. Considerable differences in incubation-time distributions have been estimated in different countries and regions. Progression from HIV to death is slower in developed countries than in developing countries [15]. This work indicates that the median time from HIV to AIDS is about ten years in developing countries. In the absence of accurate country-specific progression rates from the date of HIV sero-conversion, the above research outcomes could be both better and reliable alternative sources of information (i.e. better in the absence of large scale clinical findings) to frame the range of incubation distribution parameters. Thus it is worthwhile to estimate the incubation-time distribution. In back-calculation, researchers have pointed out considerable sensitivity of HIV incidence to incubation-time distribution[16]. Thus

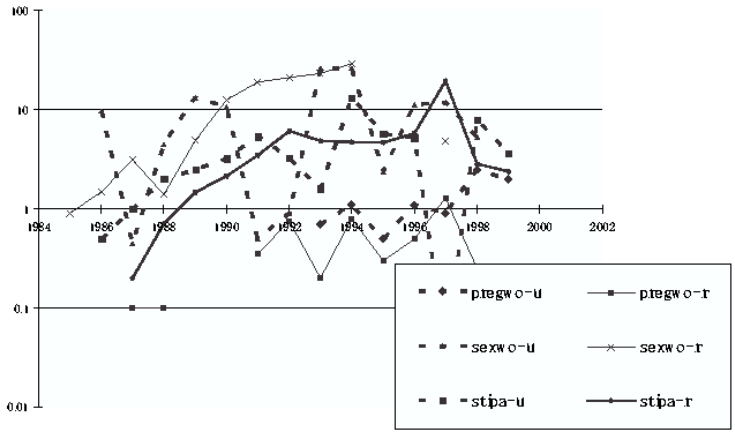

FIGURE 1. Prevalence of HIV in different groups in India. Pregnant women, sex workers, and STI patients in rural as well as urban areas, given in the UNAIDS for 2000 [18] are plotted in the log-scale over the period.

it can be expected that determining HIV incidence density could reduce the estimated range of plausible incubation-time distributions. The Weibull distribution, commonly used as an incubation distribution, has an unrealistically assumes that the hazard increases infinitely as time progresses. Hence, assuming the hazard of of contracting AIDS from an HIV-infected individual could stabilize after some time, we extended our study by estimating the parameters using a truncated incubation distribution with a set of truncation points. Second, we estimated the parameters of the non-truncated incubation distribution from the mean and truncated point of truncated distribution.

2. **HIV/AIDS surveillance in India and the data used.** In 1986, the year when first HIV was detected in India, the objective of HIV surveillance was to identify the geographical spread of HIV and the main modes of transmission through established HIV testing centers and reference centers. The main mode of transmission of HIV is through heterosexual contact, and the latest figures show 81 % of the reported AIDS cases in India acquired HIV through sexual contact. In 1992, the Indian government designed NACO to combat the epidemic, and accordingly, National AIDS Control Organisation (NACO) developed a sentinel surveillance system and extended its activity in HIV/AIDS surveillance. From the available surveillance data from almost all the states and union territories, NACO reports that the HIV growth rate is increasing. Consequently, it established more sentinel sites during 1997 and 1998, covering all states and union territories. By 2001 there were 320 such sites from where national level data has been pooled. Because AIDS cases are underreported, and the reported figure may be a fraction of the total AIDS incidence. HIV prevalence in the city of Mumbai reached 3 % in 1999, and this city has the highest percent (64%) of HIV prevalence among tested STD clinic patients. Other major urban centers of Calcutta, Chennai, and New Delhi also exhibit an increasing trend in HIV prevalence. Around 4% of the sero-positive individuals screened acquired HIV through infected syringes and needles. Figure 1 gives HIV prevalence in three selected populations over the period. Here, median prevalence

rates are plotted in log-scale. Sex workers in urban areas showed neither increasing nor a decreasing trend, but in the same population in rural areas, prevalence was increasing untill the mid-1990s. The situation is opposite for pregnant women in urban and rural areas. Interestingly, prevalence among the STI (sexually transmitted infections) population has been decreasing since the mid-1990s. We discussed this here briefly, and further details of this section can be seen elsewhere [17, 18].

Estimating India's past HIV trend using back-calculation is complex, since the incubation-time distribution is unknown. In addition, simultaneous estimation of parameters of HIV density and incubation distribution could lead to problems of nonidentifiability. NACO provides national level HIV estimates for the recent years [17]. Though there were HIV estimates from the Global Programme on AIDS (GPA), and the World Health Organization (WHO) during 1990 - 96 [17], neither provides annual HIV incidence. Also, the data provided do not cast much light on the incidence rates. The growth rate in the HIV numbers given by NACO [17] has not increased in the past four years. Also, prevalence of HIV in STI populations in rural areas has shown a declining trend since the mid-1990s (Fig. 1). The sample size of the sentinel surveillance centers from which these prevalence rates are derived is also large in comparison with that of the other groups [18]. Reported cases for each year from 1986 to 1989 were taken from the HIV/AIDS Surveillance Database 2000, US Bureau of the Census, and for each year from 1990 to 1997 they were taken from NACO [18] published data.

3. **Estimation of HIV incidence density.** As mentioned above, we estimate the incubation-time distribution from the reported data of AIDS cases and HIV incidence by the back-calculation method. In this section, we explain how we estimated the density function of HIV incidence used in our analysis. The explanation consists of two parts: parametric form of incidence density function and parameter estimation from available data.

In the beginning of the spread of infection, we theoretically expected that the number of infected persons increases exponentially except at the very beginning phase. This is based on the idea that the number of newly infected persons per-unit time should be proportional to the number of already infected persons. But, as infected persons increase, the growth rate will diminish because of the decrease of susceptible persons. Thus, we used a quadratic exponential function to represent the HIV incidence function

$$h(t) \quad = \quad exp\left(\gamma_0 + \gamma_1 t + \gamma_2 t^2\right), \tag{1}$$

where t represents time and γ_0, γ_1, and γ_2 are parameters to be estimated. To express the dependence on the parameters explicitly, dependence can be denoted as

$$h(\gamma, t) \quad = \quad exp\left(\gamma_0 + \gamma_1 t + \gamma_2 t^2\right) \tag{2}$$

where $\gamma = (\gamma_0, \gamma_1, \gamma_2)$. The parameter must be estimated from available HIV incidence data. Unfortunately, as we see in the previous section, there are no perfect data on HIV incidence in India. But what is essentially required here is the trend of HIV incidence rather than the details of it. Among the available data we noticed the median prevalence of HIV among STI patients outside major urban centers. It had the minimum level of fluctuations, because it was the largest sample

size. The incidence of HIV is presumably proportional to the prevalence rates (Fig. 1). Thus we estimated the parameters in (1) using standard regression analysis from these data. In addition to the quadratic exponential function, we also tested the sensitivity of the mean incubation period, assuming simple exponential growth.

4. **Estimation of incubation time distribution.** In this section, we briefly review the idea of back-calculation and then, in more detail, describe our extended method of estimating the incubation time distribution. Our notation is explained first together with the idea of back-calculation.

The back-calculation method is based on the idea that the number of AIDS cases can be calculated from the number of HIV-infected persons and the time schedule of developing AIDS among HIV infected persons. We express the HIV incidence density function as $h(\gamma, t)$, as shown in the previous section. The cumulative number of AIDS cases up to time t is denoted by $A(\gamma, \alpha, \beta, t)$. The parameters α and β are inherited from the incubation-time distribution as explained below. The incubation time distribution is denoted by $F(\alpha, \beta, t)$. Here, the parameters α and β specify the distribution function. This is a probability for the incubation time of less than t. The probability density function is given by $f\ (= dF/dt)$. As to the incubation-time distribution, a Weibull distribution is found to have a good fit and is widely used, although possibly that the incubation period of AIDS follows no simple parametric distribution function because of its long and variable nature [19]. Other researchers have also applied the same form to represent the incubation distribution [4, 20, 21]. The scope and the estimation of the incubation period widened over time [22, 23]. In this work, we also assume the incubation-time distribution follows a standard Weibull (3), as well as truncated version (4) and (5), given as

$$F(\alpha, \beta, t) \quad = \quad 1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\} \qquad if\ t \geq 0. \tag{3}$$

Here α and β are scale and shape parameters. In this case, the hazard function is assumed to increase infinitely as time passes, which may not be likely. Thus, we use another candidate distribution, the truncated Weibull distribution, assuming the hazard saturates at a constant level after a time point, say t_c. This is expressed as

$$F(\alpha, \beta, t_c, t) \quad = \quad \begin{cases} 1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\} & if\ 0 < t < t_c \\ 1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\} \exp\left\{-\left(\frac{\beta}{\alpha}\right)\left(\frac{t_c}{\alpha}\right)^{(\beta-1)(t-t_c)}\right\} & if\ t \geq t_c. \end{cases} \tag{4}$$

The corresponding probability density function for equation (4) is

$$f(t, \alpha, \beta, t_c) \quad = \quad \begin{cases} \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{(\beta-1)} \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\} & if\ 0 < t < t_c \\ \frac{\beta}{\alpha} \exp\left\{\frac{\beta}{\alpha}\left(\frac{t_c}{\alpha}\right)^{(\beta-1)}(t_c - t) - \left(\frac{t}{\alpha}\right)^{\beta}\right\}\left\{\left(\frac{t_c}{\alpha}\right)^{(\beta-1)} + \left(\frac{t}{\alpha}\right)^{(\beta-1)}\right\} & \\ & if\ t \geq t_c. \end{cases} \tag{5}$$

The above functional forms are used in the analysis, and results are discussed in section 5. If $t_c = 0$, then equation (4) is same as equation (3).

If the random variable T denotes the length of the incubation period, H the time of sero-conversion, and A the time of diagnosis of AIDS cases, then A can be treated as $A = H + T$. Assuming the independence of the two random variables H and T, the distribution of A, that is, $A(\gamma, \alpha, \beta, t)$, is given by the convolution of the functions that represent the density and the distribution functions of $H$and T as follows:

$$h(t) * F(t) \quad = \quad \int_{-\infty}^{\infty} h(\gamma, s) F(\alpha, \beta, t - s) ds \tag{6}$$

where the sign $'*'$ represents a convolution operation. If we assume there were no HIV infections prior to T_0 and T_R is the last time point of the AIDS reported case, then the above convolution (6) will be rewritten as an integration over a finite range $[0, T_R]$ and is given as follows:

$$A(\gamma, \alpha, \beta, t) \quad = \quad \int_{0}^{T_R} h(\gamma, s) F(\alpha, \beta, t - s) ds \tag{7}$$

Equation (7) is the fundamental relation between three important components in the method of back-calculation: cumulative number of AIDS, HIV incidence density, and incubation-time distribution.

Usually the number of AIDS cases are reported annually. The framework explained above uses time as a continuous variable. Some modification is required in the application to the available data. Let $X_1, X_2, \ldots, X_R$ be the reported number of AIDS cases at the corresponding calender time intervals $[T_0 - T_1), [T_1, T_2)$, ..., $[T_{R-1}, T_R)$. The conditional probability that an HIV infected individual who diagnosed with full blown AIDS before T_R actually developed AIDS in the time interval $[T_{i-1}, T_i)$ is

$$q_i \quad = \quad \frac{\int_0^{T_i} h(\gamma, s) F(\alpha, \beta, t_c, T_i - s) ds - \int_0^{T_{i-1}} h(\gamma, s) F(\alpha, \beta, t_c, T_{i-1} - s) ds}{\int_0^{T_R} h(\gamma, s) F(\alpha, \beta, t_c, T_R - s) ds}. \tag{8}$$

Let $X_i (i = 1, 2, ...R)$ follow a multinomial distribution with cell probabilities $q_i (i = 1, 2 \ldots, R)$, where $\sum_{i=1}^{R} q_i = 1$. The likelihood function, L, is $L(\alpha, \beta, t_c/x) = \prod_{i=1}^{R} q_i^{x_i}$. Here, α and β can be viewed such that $\alpha, \beta \in \mathbf{R}^+$. We apply the maximum-likelihood estimation technique, which is statistically robust and asymptotically efficient to estimate $\{\alpha,\beta\}$ in the log-likelihood (LL), $logL(\alpha, \beta, t_c/x) = \sum_{i=1}^{R} q_i log\, q_i$ for each truncation year between 8 and 25 and also without truncation (i.e., $t_c = \infty$), and hence, the shape of the incubation period curve is obtained.

Once the parameters of the Weibull distribution are estimated then by using the truncation point, one can estimate the mean incubation period and other moments of the Weibull distribution using (5). Other researchers have attempted to estimate the moment generating function for the truncated density function presented in this section [24]. The rth moment for this density functions when the shape parameters, $\beta = 1$ and 2, are

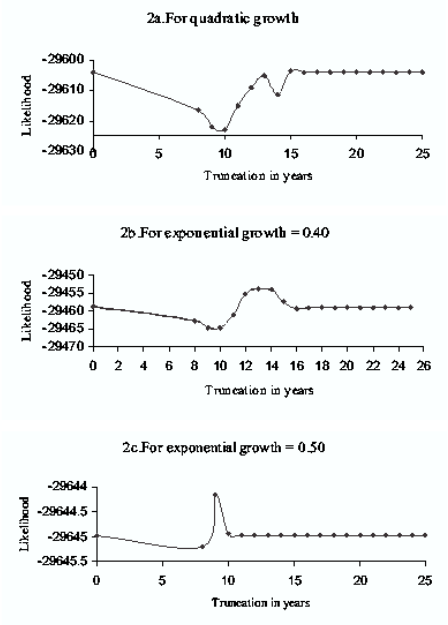

FIGURE 2. Likelihood values with respect to truncation. This is based on each truncation points from 8 to 25 years, corresponding to the maximum log-likelihood values plotted separately for (a) quadratic exponential, (b) exponential = 0.40, and (c) exponential = 0.50.

$$T_{r1} = \alpha^{r-1} 2 \left[\Gamma(r+1) - \gamma\left(r+1, \frac{\alpha\phi + t}{2\alpha}\right)\right], \tag{9}$$

$$T_{r2} = \alpha^{r} \gamma\left(\frac{r}{2}+1,\ (\sqrt{3}-1)^2 h_1\right) + 3.5\alpha^{-2}(\sqrt{3}-1)^{r+1}\,\Gamma(r+2,\ h). \tag{10}$$

See the appendix for a brief description of the derivation that led to expressions (9) and (10). A detailed mathematical analysis [24] is not in the scope of the present work. But, for prediction purposes, we are interested in these density functions for estimating parameters. One can calculate the non-truncated Weibull parameters corresponding to maximum LL explained in this section from these moments. To compare truncated versus nontruncated distributions, we finally give two sets of incubation distribution parameters corresponding to the explanations as follows:

$\{\alpha, \beta\}$: LL is maximum without truncation point

$\{\alpha_1, \beta_1\}$: LL is maximum with truncation point

As an example, we projected AIDS cases using $\{\alpha, \beta\}$. In addition to this estimate, we have assumed the shape of the HIV density in the past as a simple exponential with growth rates $\theta_0 = 0.40$ and $\theta_1 = 0.50$ [25]. Again using the densities in (7), we estimated the incubation-time distributions as explained in the procedure above. See [26] for the discussion on parameters of Weibull distribution.

5. **Results.** The estimated values of the parameters of HIV incidence are $\gamma_0 = -4.2787$, $\gamma_1 = 0.6938$, and $\gamma_2 = -0.0234$. According to these values, HIV incidence

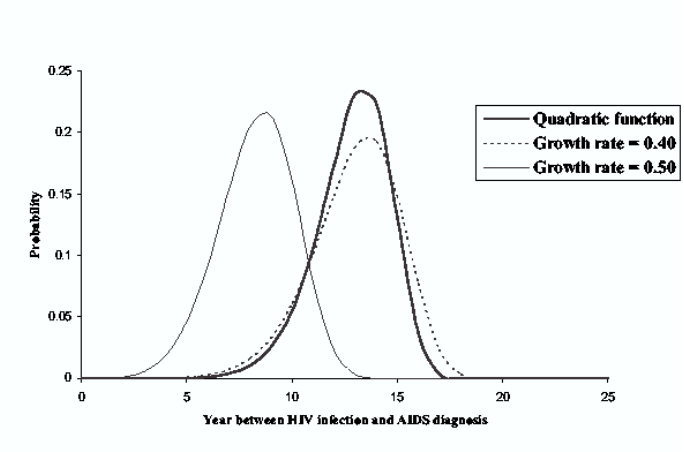

FIGURE 3. Estimated incubation-time densities of AIDS in India. This is from the shape and scale parameters for three scenarios given in Table 2 (also explained in section 4 in the text). Weibull densities are plotted.

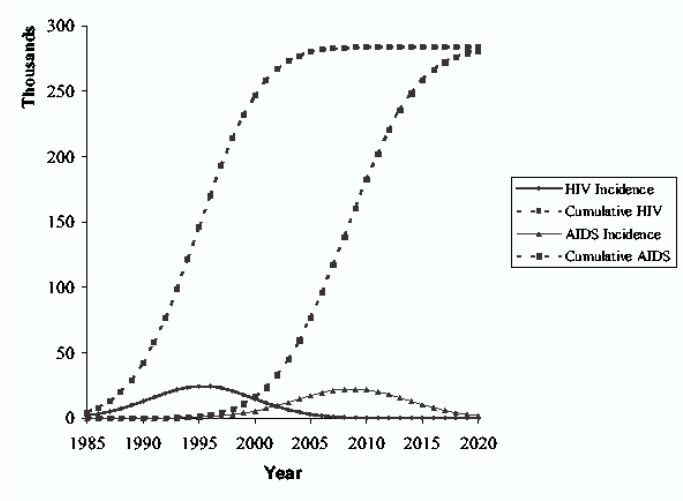

FIGURE 4. HIV/AIDS estimates in India. Cumulative numbers of AIDS and HIV are generated by $\int_0^t h(t)dt$ and the convolution equation $\int_0^t h(s)F(t-s)ds$ and then are plotted against the years. HIV and AIDS incidence is calculated by $h(t), t = 0, 1, 2\ldots.$ and $F(t) - F(t-1), t = 1, 2, 3\ldots.$ Where $h(t) = \exp\left(-4.2789 + 0.6938t^2 - 0.0234t^2\right)$ and $F(t) = 1 - \exp\left(-\frac{t}{13.6139}\right)^{8.7792}$

increased since the beginning of the epidemic, reaching a peak in 1995, and then started to decline. The estimated parameters of the incubation-time distribution without considering the truncation point are $\alpha = 13.6139$ and $\beta = 8.7792$ ($LL =$

TABLE 1. Prevalence of HIV among STI patients (out side major urban areas) in India

Year	Median Prevalence (%)	Year	Median Prevalence (%)
1986	-	1994	4.75
1987	0.2	1995	4.65
1988	0.71	1996	5.8
1989	1.47	1997	19.4
1990	2.14	1998	2.86
1991	3.48	199	2.4
1992	6.11	2000	-
1993	4.85		

(Source: UNAIDS/WHO Epidemiological Fact Sheet, India 2000 update [18])

-2960.4). (The mean is 12.87 years; SD = 3.07.) When we tested the impact of truncation on the shape of the incubation distribution, we found log-likelihood is maximum at truncation point 15 (i.e., at year 15 after the first year) with $LL = -2960.3$, and corresponding parameters are $\alpha_1 = 13.6145$ and $\beta = 8.7706$. Using these values, we found that the mean incubation period is 13.69 years and SD is 2.95. When the simple exponential densities with $\theta_0 = 0.40$ and $\theta_1 = 0.50$ were employed, mean incubation distribution periods were 12.77 and 8.30 years. Through these statistics we reestimated parameters for the nontruncated Weibull distribution as $\alpha_0 = 14.6407$ and $\beta_0 = 6.9981$. These three sets of situations (one quadratic exponential and two simple exponentials) indicate that the shape of the HIV curve also explains the variability in incubation-time distribution.

Log-likelihood values increased proportionally with year of truncation, and after reaching maximal value, they decrease with increasing year of truncation and again approach that of the nontruncated Weibull. This is obvious, as the delay in stabilization of the hazard of attaining full-blown AIDS will nullify the effect of year of truncation. This situation is well represented in Figure 2. It is clear that global minima and maxima varied in three situations.

Figure 3 shows the differences in the incubation-time densities that arose due to the three situations. We did not find the difference in the densities (i.e., in the disease progression process) when the hazard of AIDS is truncated and not truncated. Figure 3 shows how the values with and without truncation are close. The densities are close to the normal density curve, with the exception that the normal density also takes values on the negative axis. This property of the Weibull density approaching the normal density is also supported theoretically when the shape parameter β of the Weibull density exceeds 3.6. The proportion of individuals developing AIDS is flattened between 18 and 19 years, since the HIV infection when the HIV curve follows quadratic exponential as well as simple exponential with $\theta_0 = 0.40$. But in the case of $\theta_1 = 0.50$, the flattening started at year 14. This could be due to the fact that the high growth rate of HIV did not match with the trend in the diagnosed AIDS cases (see Fig. 4).

5.1. **Validation of the method with US data.** We have taken values from the Centre for Disease Control (CDC) and The United Nations Program on AIDS (UNAIDS) [18] sources for US data, and conducted similar analysis as explained in

TABLE 2. Estimated parameters and descriptive statistics

Growth rate	α	β	Mean	Variance	T_c	LL
Quadraric	13.6145	8.7706	12.86	2.95	15	-2960.4
	13.6139	8.7792	12.87	3.07	0	-2960.3
Exp (0.40)	14.1684	7.1572	12.98	3.74	13	-2945.4
	13.8914	7.3355	12.77	3.47	0	-2945.9
Exp(0.50)	8.8730	5.3745	8.06	2.55	9	-2964.4
	9.0131	5.2173	8.30	03.34	0	-2964.5

Note: T_c value indicates truncation values where LL is maximum corresponding to the α, β and '0' indicates no truncation.

section 4. The estimated parameters of the incubation-time distribution without taking a truncation point are $\alpha = 11.6713$ and $\beta = 4.812$ ($LL = -3050.2$). (The mean is 10.69 years; SD = 2.53). When we tested the effect of truncation on the shape of the incubation distribution, we found that log-likelihood is maximum at the truncation point 12 (i.e., at year 12 after the first year) with $LL = -3120.4$, and the corresponding parameters are $\alpha_1 = 11.9211$ and $\beta = 4.4500$. Using these values, we found the mean incubation period is 10.64 years and the SD is 2.50. Estimated parameters for the non-truncated Weibull distribution were $\alpha_0 = 14.6407$ and $\beta_0 = 6.9981$. These results show that the method worked well for the US data, which are known to be accurate with respect to diagnosis, reporting, and surveillance.

6. **Discussion.** HIV incidence density is estimated with the information on HIV prevalence among the STI patients residing outside major urban areas in India, unlike the regular back-calculation approach in arlier studies [4, 20, 21]. There could be a more sophisticated method to estimate HIV incidence density, but the degree of improvement will be small because the data contain significant ambiguity. The regression method of estimating parameters of the HIV density seems to have worked well. The HIV estimates given by NACO indicate that there was a decline in the new HIV infections from 1997 to 2000. Also the explanations given in sections 2 and 3 provide some basis for the reduction in new HIV cases from 1995. The peak of HIV incidence in the STI rural population occurred in 1995 (Fig. 1). Usually, the HIV peak in the general population occurs after the peak in the high-risk population, and NACO's HIV estimates support this typical epidemic behavior. These arguments justify the quadratic exponential as a representative of HIV epidemic in India. Also, there is a good possibility that new infections in India will decline in response to the significant prevention measures taken by the government. We have suggested using a quadratic exponential for the HIV growth in India [25]. During the past decade, back-calculation approach with suitable improvements was extensively applied in different populations throughout the world.

The assumption of incubation-time distribution, such as the Weibull distribution, has been justified from previous articles [4, 19, 22]. So far, no rigorous estimation procedure of incubation distribution in India has been published. Hence, estimating indirectly through the back-calculation approach may provide the most reliable estimation to date. Our estimates give the maximum likelihood in the cross-section ranges between mean incubation period 8 and 15 with SD=3, 4, 5, 6. From a statistical point of view, theoretical limitations arise if one estimates the incubation

distribution from the prevalent cohort of HIV individuals in India [5]. Earlier, Peter Bacchetti, [16] estimated parameters of the incubation distribution by using AIDS incidence and HIV estimates in the population through deconvoluting the relationship given by back-calculation. This method will be direct if the AIDS and HIV incidence in the population is quite clear, but this is not feasible in the India due to a lack of infection data for a sufficient time period. We estimated parameters using the maximum-likelihood method, rather than with nonparametric estimation [27]. Back-calculation could also be used to estimate parameters of the incubation distribution and incidence density simultaneously, but this approach may lead to problems of nonidentifiability, as noted in section 3. If the actual population incidence rates matched those of the back-calculation method, then simultaneous estimation would not be objectionable. However, we found that these two estimates of incidence are not close, which led us to use deconvolution as a method to estimate parameters. Sensitivity analysis is conducted to see how the incubation time changes due to the HIV curve in the past. The assumption of a simple exponential for the future is not appropriate since HIV growth has shown moderate stability, and preventive strategies taken by the government could well arrest the incidence. Hence, we did not carry out the future analysis using a simple exponential. It was found that only 11 % to 16 % of the AIDS cases were being reported in India [25]; hence, by adjusting this to the current estimates, the actual AIDS cases in the future can be obtained. We are not sure whether this underreporting will also be constant in the future. The model may improve by using adjusted AIDS case numbers in the wake of reliable estimates of underreporting in the future. Even information on HIV clinical data along with reported AIDS cases can allow further modeling of the HIV density function [28]. Application of back-calculation to the Indian AIDS data in the future may lose reliability because of the expected increase in the therapy and alternative medicines that prolong the time between HIV infection and appearance of AIDS. In addition to this, information on CD4 count decay and its parameters have been estimated. The lengthier incubation period estimated in the Indian population is a good indication of longtime survival of HIV-positive individuals. The absolute magnitude of HIV cases is large, but the proportion of HIV-positive individuals is comparatively smaller than that of some African and Asian countries. Also, the expected attainment of stability in the first half of the current decade is a positive sign, in terms of HIV advocacy. If the estimated trend of HIV is continued in the future, it will further systematize the prevention policies undertaken by the government and nongovernmental bodies, and will avoid the disorder in the public-health scenario that arose when HIV appeared in the country. However, the decrease in new cases will not continue if governmental and nongovernmental authorities reduce the effort to prevent HIV/AIDS. It will take an additional fourteen to fifteen years to attain stability in the estimated cumulative AIDS cases. Health-care professionals need a special understanding of the rapid HIV/AIDS spread mechanism and hence must make better efforts to lessen the public-health burden. There must be efforts to manage AIDS deaths and other opportunistic-infection deaths that emerge due to AIDS.

No study in India so far gives the AIDS incubation period with some degree of variation. Our estimates of incubation distribution assuming three sets of HIV trends should provide an understanding of the nature of the AIDS incubation period, if there is reasonable matching between AIDS and HIV numbers applied.

Application of the back-calculation method on Indian data with theoretical observations and limitations provides a new look at a well-known old method. Our estimate of time of developing AIDS is longer than the general belief [15]. An increase of almost three years in the estimation could have occurred due to the following errors: incidences of HIV and AIDS might not have matched perfectly; diagnosed AIDS cases were not reported in time; and small delays could have caused some the cases to be reported in the next year. Even lack of facilities or awareness might delay diagnosis. Any combination of these causes might have stretched an increase between two and three years in the mean incubation period. We do not have any reliable information on the effect of Anti Retroviral Therapy (ART) and Highly Active ART, on survival rates among Indian patients, or in the best case, whether the population may naturally have a longer incubation period. Epidemiological studies are needed to obtain accurate estimates of the incubation period. It seems very important to establish the consistency of HIV/AIDS data, investigation of the incubation time, or time between report of HIV and that of AIDS.

Remark. The method demonstrated here uses data and information obtained from an Internet search in January 2002.

Acknowledgement. A. S. R. S. Rao is extremely thankful to Professors Subrata Lahiri and Sukumar Mukherji for their enthusiasm and their encouragement on AIDS modeling before collaborating this work with Masayuki Kakehashi. Many thanks to Professors Biswanath Bhattacharya, Jayanta Ghosh, Nilanjan Joshi for their critical comments at various stages in preparing this work. We thank Dr. Chris Bauch for assisting us with our English grammar. We are grateful to Professor Patrick Nelson, an associate editor, whose constructive suggestions helped us to arrive at this final version. The authors are grateful to Heiwa Nakajima Foundation for initiating this collaborative research through A. S. R. S. Rao's visit to Hiroshima University in 2000. We are also thankful for the partial support received through grants-in-aid for scientific research from the Japan Ministry of Education; Science and Culture (Kakehashi); and the Department of Science and Technology, New Delhi (Rao).

Appendix: The rth moment for shape parameters ≤ 2. Let T_r denote the r th moment for (5), then it is expressed as follows:

$$\begin{aligned} T_r &= \int_0^{\delta} y^r \frac{\beta}{\alpha}\left(\frac{y}{\alpha}\right)^{\beta-1} e^{-\left(\frac{y}{\alpha}\right)^{\beta}} dy + \int_{\delta}^{\infty} y^r \frac{\beta}{\alpha} e^{\left\{\frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}(t-y)-\left(\frac{y}{\alpha}\right)^{\beta}\right\}} \left(\frac{t}{\alpha}\right)^{\beta-1} dy \\ &\quad + \int_{\delta}^{\infty} y^r \frac{\beta}{\alpha} e^{\left\{\frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}(t-y)-\left(\frac{y}{\alpha}\right)^{\beta}\right\}} \left(\frac{y}{\alpha}\right)^{\beta-1} dy. \end{aligned} \tag{11}$$

Evaluating integral (5) is not easy.

Take $\beta = 1$ in (11), then we get the following expressions:

$$\begin{aligned} T_{r1} &= \frac{1}{\alpha}\int_0^{\delta} y^r e^{-\left(\frac{t}{\alpha}\right)} dy + \frac{2}{\alpha}\int_{\delta}^{\infty} y^r e^{\left\{\frac{1}{\alpha}(t-y)-\left(\frac{y}{\alpha}\right)\right\}} dy \\ &= \frac{1}{2^r\alpha}\int_{-t/\alpha}^{\phi} (\alpha w + t)^r \, e^{-\left(\frac{\alpha w + t}{2\alpha}\right)} dw + \frac{1}{2^{r-1}\alpha}\int_{\phi}^{\infty} y^r e^{-w} dw \\ &= \alpha^{r-1} 2\left[\Gamma(r+1) - \gamma\left(r+1, \frac{\alpha\phi + t}{2\alpha}\right)\right]. \end{aligned} \tag{12}$$

In 12 please note that $\left(\frac{y}{\alpha}\right) - \frac{1}{\alpha}(t-y) = w$, as $y \to 0 \Rightarrow w \to -t/\alpha$ and $y \to \delta \Rightarrow w \to \frac{2\delta - t}{\alpha}$ (say ϕ). Equation (12) can be used for getting the moments of (5) in this special condition.

The lower incomplete gamma function can be written in terms of confluent hypergeometric functions [29]. In such a case,

$$\gamma(r+1, \frac{t}{\alpha}) = (r+1)^{-1}\left(\frac{t}{\alpha}\right)^{r+1} e^{-\frac{t}{\alpha}} \mathbf{M}(1, r+2, \frac{t}{\alpha}). \tag{13}$$

In (13), $\mathbf{M}$ is called Kummer's function.

Take $\beta = 2$. If we substitute $y = -t + t\sqrt{3}$ in (11) and change the limits accordingly (as $y \to 0$ then $t \to 0$, as $y \to \infty$ then $t \to \infty$, and *as* $y \to \delta$ then $t \to 1.37\delta$ (say, h), then the terms of (11) can be written as follows:

$$\begin{aligned} &\int_0^{\delta} y^r \frac{\beta}{\alpha}\left(\frac{y}{\alpha}\right)^{\beta-1} e^{-\left(\frac{y}{\alpha}\right)^{\beta}} dy \\ &= \int_0^{h} t^r (\sqrt{3}-1)^r \frac{2}{\alpha}\left(\frac{(\sqrt{3}-1)\,t}{\alpha}\right) e^{-\left(\frac{(\sqrt{3}-1)t}{\alpha}\right)^2} dt.\,(\sqrt{3}-1) \\ &= \frac{2\,(\sqrt{3}-1)^{r+2}}{\alpha^2}\int_0^{h} t^{r+1} e^{-(\sqrt{3}-1)^2\left(\frac{t}{\alpha}\right)^2} dt. \end{aligned}$$

Taking $t^2/\alpha^2 = k$ and changing the limits and simplifying, then the above integral becomes $\big($as $t \to 0$ then $k \to 0$ and as $t \to h$ then $k \to h^2\delta^2/\alpha^2 = 1.88\delta^2/\alpha^2$ $(say, h_1)\big)$

$$\begin{aligned} &= \frac{(\sqrt{3}-1)^{r+2}}{\alpha^2}\int_0^{h_1} (k\alpha)^r \, e^{-(\sqrt{3}-1)^2 k}\alpha^2 dk \\ &= \alpha^r \gamma\left(\frac{r}{2}+1,\ (\sqrt{3}-1)^2 h_1\right) \end{aligned} \tag{14}$$

$$= \alpha^r (\sqrt{3}-1)^{r+2} h_1^{\frac{r}{2}+1}\left(\frac{r}{2}+1\right)^{-1} e^{-(\sqrt{3}-1)^2 h_1}\mathbf{M}(1,\ \frac{r}{2}+2,\ (\sqrt{3}-1)^2 h_1). \tag{15}$$

Now consider the second term in (11) and substitute $\beta = 2$, then we get,

$$\int_{\delta}^{\infty} y^r \frac{2}{\alpha} e^{\left\{\frac{2}{\alpha}\left(\frac{t}{\alpha}\right)(t-y)-\left(\frac{y}{\alpha}\right)^2\right\}}\left(\frac{t}{\alpha}\right) dy. \tag{16}$$

Substituting $y = t(\sqrt{3}-1)$ in (16), changing the limits accordingly, and simplifying, we get (17)

$$\frac{2\left(\sqrt{3}-1\right)^{r+1}}{\alpha^2}\int_h^\infty t^{r+1}dt \tag{17}$$

$$= \frac{2\left(\sqrt{3}-1\right)^{r+1}}{\alpha^2}\Gamma(r+2,\ h). \tag{18}$$

Now consider third term in (11) and substitute $\beta = 2$. Then we get,

$$\int_\delta^\infty y^r \frac{2}{\alpha} e^{\left\{\frac{2}{\alpha}\left(\frac{t}{\alpha}\right)(t-y)-\left(\frac{y}{\alpha}\right)^2\right\}}\left(\frac{y}{\alpha}\right)dy. \tag{19}$$

Substituting $y = t(\sqrt{3}-1)$ in (19), changing the limits accordingly, and simplifying, we get

$$\frac{2\left(\sqrt{3}-1\right)^{r+2}}{\alpha^2}\int_h^\infty t^{r+1}dt \tag{20}$$

$$= \frac{2\left(\sqrt{3}-1\right)^{r+2}}{\alpha^2}\Gamma(r+2,\ h). \tag{21}$$

Now the rth moment T_{r2} when $\beta = 2$ is the sum of (14),(18), and (21) is given by

$$T_{r2} = \alpha^r\gamma\left(\frac{r}{2}+1,\ \left(\sqrt{3}-1\right)^2 h_1\right) + 3.5\alpha^{-2}\left(\sqrt{3}-1\right)^{r+1}\Gamma(r+2,\ h). \tag{22}$$

Here $h = 1.37$ and $h_1 = (h\delta)^2/\alpha^2$. For β values higher than 2, it will be very complex to evaluate the rth moments.

REFERENCES

[1] Cui, J., and Becker, N. G. (2000). Estimating HIV incidence using dates of both HIV and AIDS diagnosis. Stat. Med. 19, 1165 - 77.

[2] Basu, A., Basu, S., Chakraborty, M. S., Dewanji, A., Ghosh, J. K., and Majumder, P. P. (1998). Projection of HIV infection in Calcutta. Ind. J. Med. Res. 107, 159 - 72.

[3] Brookmeyer, R., and Gail, M. H. (1986). Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. Lancet. 2, 1320 - 22.

[4] Brookmeyer, R., and Gail, M. H. (1988). A method for obtaining short term predictions and lower bounds on the size of the AIDS epidemic. J. Am. Stat. Assoc. 83, 301 - 08.

[5] Rao, A. S. R. S., Basu, S., Basu, A., and Ghosh, J.K. Parametric models for incubation distribution in presence of left and right censoring. (submitted to Ind. J. Pure Appl. Math.)

[6] Anderson, R. M., and May, R. M. (1991). Infectious Diseases of Humans: Dynamics and Control. Oxford: Oxford University Press.

[7] Anderson, R. M. (1988). The Epidemiology of HIV infection: Variable incubation plus infection periods and heterogeneity in sexual activity. J. Roy. Stat. Soc. A. 151, 66 - 93.

[8] Isham, V. (1988). Mathematical Modelling of the Transmission Dynamics of HIV infection and AIDS: A review. J. Roy. Stat. Soc. A., 151 (1), 5 - 30.

[9] Kakehashi, M. (1998). A mathematical analysis of the spread of HIV/AIDS in Japan. IMA J. of Math. Appl. Med. Biol. 15, 299 - 311.

[10] Kakehashi, M. (1999). Erratum: A mathematical analysis of the spread of HIV/AIDS in Japan. IMA J. of Math. Appl. Med. Biol., 16, 111 - 12.

[11] Hethcote, H. W. (1989). Three basic epidemiological Models. In Applied Mathematical Ecology (S. A. Levin, T. G. Hallam and L. J. Gross, editors). 119-144. Berlin: Springer-Verlag.

[12] Bauch, C. T and Rand, D. A. (2000). A moment closure model for sexually transmitted disease transmisssion through a concurrent partnership network. Proc. R. Soc. Lond. B, 267, 2019 - 27.

[13] Bauch, C. T. (2002). A versatile ODE approximation to a network model for the spread of sexually transmitted diseases. J. Math. Biol. 45, 375 - 95.

[14] Kakehashi, M. (2000). Validity of simple pair formation model for HIV spread with realistic parameter setting. Math. Pop. Stud., 8 (3), 279 - 92.
[15] Schwartlander, B., Stanecki, K. A., Brown, T., Way, P. O., Monasch R., Chin, J., Tarantola, D., and Walker, N. (1999). Country-specific estimates and models of HIV and AIDS: methods and limitations. AIDS. 13, 2445 - 58.
[16] Bacchetti, P., Segal, M. R., Hessol, N. A., and Jewell, N. P. (1993). Different AIDS incubation periods and their impacts on reconstructing human immunodeficiency virus epidemics and projecting AIDS incidence. Proc. Nat. Acad. Sci. USA, 90, 2194 - 96.
[17] NACO. Surveillance of HIV infection/AIDS cases in India. Indian AIDS Scenario. www.naco.nic.in (browsed in January 2002).
[18] UNAIDS. Epidemiological fact sheet on HIV/AIDS and sexually transmitted infections. http://www.unaids.org. (browsed in January 2002).
[19] Lui, K. J., Lawrence, D. N., Morgan, W. M., Peterman, T. A., Haverkos, H. W., and Bregman, D. J. (1986). A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome, Proc. Nat. Acad. Sci. USA. 83, 3051 - 55.
[20] PHLS. (1990). The incidence and prevalence of AIDS and other severe HIV disease in England and Wales for 1992 - 1997: Projections using data to the end of September 1989. Communicable Disease Report (suppl).
[21] PHLS. (1993). The incidence and prevalence of AIDS and other severe HIV disease in England and Wales for 1992 - 1997: projections using data to the end of June 1992. Communicable Disease Report 3, suppl 1.
[22] Medley, G. F., Anderson, R. M., Cox, D. R, and Billard, L. (1987). Incubation period of AIDS in patients infected via blood transfussion. Nature. 328, 719 - 21.
[23] Bacchetti, P., and Moss, A. R. (1989). Incubation period of AIDS in San Francisco. Nature. 338, 251 - 53.
[24] Rao, A. S. R. S. A note on derivation of generating function for the right truncated Rayleigh distribution (communicated).
[25] Rao, A. S. R. S., and Kakehashi, M. (2004). A combination of differential equations and convolution in understanding the spread of an epidemic. Sadhana. 29 (3), 305 - 13.
[26] Johnson, N. L., Kotz S., and Balakrishnan, N. (1994). Continuous Univariate Distributions. 2nd ed. Vol. 1. New York: Wiley and Sons.
[27] Bacchetti, P. (1990). Estimating the Incubation Period of AIDS by comparing Population Infection and Diagnostic Patterns. J. Am. Stat. Assoc. 85, 1002 - 08.
[28] Raab, G. M., Fielding, K. L., and Allardice, G. (1994). Incorporating HIV test data into forecasts of the AIDS epidemic in Scotland. Stat. Med. 13, 2009 - 20.
[29] Abramowitch, M., and Stegun, I. A. (Eds). (1972). Handbook of Mathematical Functions. Dover, New York.

Received on June 30, 2004. Revised on March 9, 2005.

E-mail address: arnirao@uoguelph.ca
E-mail address: mkake@hiroshima-u.ac.jp