Jpn. J. Infect. Dis., 57, 2004

Laboratory and Epidemiology Communications

Simple Approximate Backcalculation Method Applied to Estimate HIV Prevalence in Japan

Hiroshi Nishiura[1,2,3]*, Hideki Yanai[2], Takashi Yoshiyama[2] and Masayuki Kakehashi[3]

[1]*Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, London, England,*
[2]*The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo 204-8533 and*
[3]*Graduate School of Health Sciences, Hiroshima University, Hiroshima 734-8551*

Communicated by Naoki Yamamoto

(Accepted June 2, 2004)

Since accurate monitoring of HIV prevalence is critical to the planning and evaluation of its treatment, care, and prevention programs, HIV/AIDS surveillance has been initiated in Japan since 1984 (1). However, there remains a question regarding the reliability of estimates of HIV in Japan. For example, curiously enough, the relative index of HIV infection rate among blood donors in Japan was thirteen times greater than that of western countries (2). The non-obligatory reporting of "progression to AIDS" and reporting delay (3) has hindered us from grasping the reliable number of the HIV-infected in Japan.

Recently, Yoshikura (4) performed an arithmetic estimation of the number of "HIV-infected" subjects in order to calculate the figure crudely and simply. In order to increase the accuracy of his approach, we took the knowledge of disease progression into account. Our aim in this paper is not only to estimate the crude number of HIV infections but also to make estimation easier by using a reasonable method which extends the backcalculation method.

We consider the point process of new HIV infections and assume that the distribution of the lengths of the incubation periods (from infection to AIDS diagnosis) is independent, identically distributed variables with probability density function (PDF) $w(\tau)$. Let τ be the time between infection and AIDS diagnosis and $w(\tau)$ be the PDF of this incubation time τ. We always describe the PDF of the incubation period of HIV infection with the use of Weibull distributions (5). This could be written as follows:

$$w(\tau,\alpha,\beta)=\frac{\alpha}{\beta^{\alpha}}\tau^{\alpha-1}e^{-\left(\frac{\tau}{\beta}\right)^{\alpha}}, \quad [1]$$

where α and β are shape parameter and scale parameter, respectively (Fig. 1). The discrete time data in Table 1 is produced from Fig. 1 generated from equation [1] where $\alpha=2.286$ and $\beta=10$. Here, the cumulative fraction of AIDS progression in the initial 3 years after an infection (i.e., 0.09, 0.10, and 0.12 in the first, second, and third years, respectively) is higher than in the actual disease progression. This deviation was chosen for the purpose of mathematical convenience in order to avoid too large an effect of $w(t)$ when t comes close to 0 (6). Backcalculation is based on the

*Corresponding author: Mailing address: The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Matsuyama 3-1-24, Kiyose, Tokyo 204-8533, Japan. Tel: +81-424-93-3090, Fax: +81-424-92-8258, E-mail: h.nishiura@imperial.ac.uk

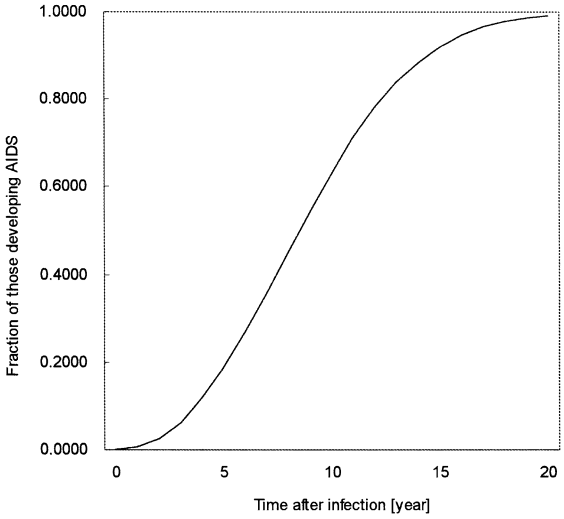

Fig. 1. Progression to AIDS after an infection.
Incubation period of AIDS was assumed to follow the probability density function of the Weibull distribution.

Table 1. Weibull distribution for AIDS progression

τ (years after an infection)	$w(\tau)$
0	0.0000
1	0.0052
2	0.0249
3	0.0618
4	0.1158
5	0.1854
6	0.2673
7	0.3576
8	0.4514
9	0.5443
10	0.6321
11	0.7116
12	0.7806
13	0.8382
14	0.8844
15	0.9201
16	0.9465
17	0.9654
18	0.9784
19	0.9869
20	0.9924

underlying relation between the cumulative number of new AIDS cases in time t to t+dt which we will denote $a(t)$, and the number of new HIV infections $h(s)$ at time s since the start of the epidemic at time ($s=0$). Therefore, the formula is given by

$$a(t)=\int_0^t h(t-\tau)w(\tau)d\tau. \quad [2]$$

Without doubt, the most reliable data in HIV/AIDS surveillance is the number of AIDS cases because its apparent symptoms more likely to lead to hospital visits. Therefore, normally we estimate the number of HIV infections in a time series (by obtaining h) by inverting equation [2] (deconvolution) with a and w (7). However, since this method has been relatively impractical (or unfriendly) for epidemiologists as well as clinicians, it is fruitful to keep matters simple. We therefore extend simple approximation into this equation by transforming it into a discrete time model.

That is, 1 year after the onset of an epidemic, $h(1)w(1)$ cases of AIDS would be produced according to $a(1)=h(1)w(1)$. Here, $h(1)$ denotes the number of HIV infections during the first year. After 2 years, we would get $h(2)$ from $a(2)=h(1)w(2)+h(2)w(1)$. After the next year, in the same way, we could derive $h(3)$ from $a(3)=h(1)w(3)+h(2)w(2)+h(3)w(1)$. Thus, the cumulative number of HIV infections, after 3 years, is $h(1)+h(2)+h(3)$. Mathematically, the estimated number of HIV infections at t years after the onset of an epidemic would be given by

$$h(t)=\frac{a(t)-[h(1)w(\mathrm{t})+h(2)w(t-1)+\mathrm{L}+h(t-1)w(2)]}{w(1)}. \quad [3]$$

As seen in equation [3] (and as can be seen in the usual backcalculation method), the impact of $w(1)$ on the estimated trend of HIV infections becomes high in our method. With the use of the widely distributed statistical data regarding AIDS (8), we estimated the cumulative number of HIV infections within our model. Here, we had to exclude patients after 2000 from our subjects, because the progression to AIDS of patients previously reported as "HIV-infected" has been non-compulsory since 1999.

Applying our model using the simple assumptions mentioned above, the cumulative number of HIV infections has been obtained (Fig. 2). There could be approximately 5,453 cases in 1999. The growth trend is nearly exponential, which is consistent with the basic concept of the mathematical (epidemiological) modeling of an epidemic: the growth in a "linear" system would normally be seen at the exponential phase (early phase) of an epidemic. If the exponential growth trend continued (that is, if our methodology could be applied to recent epidemiology), the number of HIV-infected would have been approximately 15,135 cases in 2003. However, this must be an overestimation because of our assumptions (for PDF in first 3 years after an infection), and because of theoretical reasons as follows.

The goodness of fit could be seen especially at the early phase of an epidemic, becoming rougher later on. This indicates a limitation of this model. After definite recognition (or announcement) of an epidemic, the population would change their behavior (i.e., avoid risky sexual intercourse), which would lead to a decrease in the susceptible population as well as in transmission potential (9). Due to the introduction of an effective anti-retroviral therapy (ART), the incubation period defined with approximately parameterized Weibull distribution could form a different shape. As a result, it is more effective to consider the growth of an epidemic within a non-linear system. Therefore, to a greater or lesser degree, the backcalculation method, including our approximate method, could be applied for short-term study (10). When a longer-term model is required, we should apply a dynamic system for estimation, evaluation, and future prediction (11). Although we could do this by fitting the early phase data shown in Fig. 2 with the use of a mathematical model for HIV, we would like to leave this for further study. Another limitation to be noted in our model is that the estimation itself could be very different according to the definition of PDF and its parameters (12).

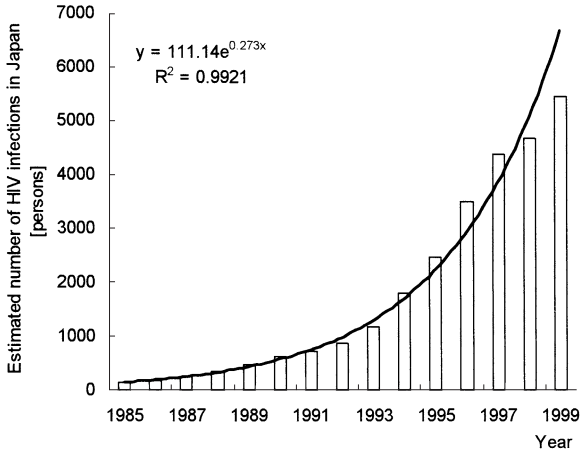

Fig. 2. Cumulative number of HIV infections generated by simple approximate backcalculation method.
Estimated number was obtained through equation [3]. Each bar indicates the accumulation of $h(t)$ up to the indicated year calculated by equation [3], and the curve is exponential fitted best to the bars.

The authors are grateful to Professor Roy M. Anderson of the Department of Infectious Disease Epidemiology, Imperial College London for his helpful advice and to Professor Hiroshi Yoshikura, Emeritus Member of the National Institute of Infectious Diseases, Japan for his discussions. HN is grateful to the Japanese Foundation for AIDS Prevention for supporting his stay and study in the UK. This work was supported in part by the Japanese Foundation for AIDS Prevention.

REFERENCES

1. Kihara, M., Ono-Kihara, M., Feldman, M. D., Ichikawa, S., Hashimoto, S., Eboshida, A., Yamamoto, T. and Kamakura, M. (2003): HIV/AIDS surveillance in Japan, 1984-2000. J. Acquir. Immune. Defic. Syndr., 32, S55-62.
2. Kihara, M., Imai, M. and Shimizu, M. (2000): Trends of HIV infections among blood donors. Infect. Agents Surveillance Rep., 21, 140-141 (in Japanese).
3. Matsuyama, Y., Hashimoto, S., Ichikawa, S., Nakamura, Y., Kidokoro, T., Umeda, T., Kamakura, M., Kimura, S., Fukutomi, K., Ikeda, C. and Kihara, M. (1999): Trends in HIV and AIDS based on HIV/AIDS surveillance data in Japan. Int. J. Epidemiol., 28, 1149-1155.
4. Yoshikura, H. (2003): Arithmetic estimation of the HIV-infected in Japan. Jpn. J. Infect. Dis., 56, 173-174.
5. Anderson, R. M., May, R. M., Boily, M. C., Garnett, G. P. and Rowley, J. T. (1991): The spread of HIV-1 in Africa: sexual contact patterns and the predicted demographic impact of AIDS. Nature, 352, 581-589.

6. Freund, H. P. and Book, D. L. (1990): Determination of the spread of HIV from the AIDS incidence history. Math. Biosci., 98, 227-241.
7. Brookmeyer, R. and Gail, M. H. (1988): A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. J. Am. Stat. Assoc., 83, 301-308.
8. Committee of AIDS Trends, Ministry of Health, Labor and Welfare (2003): Trend of AIDS incidence in Japan, Annual Report 2002. Ministry of Health, Labor and Welfare (in Japanese).
9. Inaba, H. (1995): On trends of AIDS and an estimate for the number of HIV infecteds in Japan. J. Popul. Probl., 50, 31-44 (in Japanese).
10. Inaba, H. (1994): Estimation of the number of HIV infecteds in the early stage of epidemic and control strategy. J. Popul. Probl., 49, 23-33 (in Japanese).
11. Anderson, R. M. and May, R. M. (1991): Infectious Disease of Humans. Oxford. Univ. Press, New York.
12. Brookmeyer, R. and Gail, M. H. (1994): AIDS Epidemiology: A Quantitative Approach. Oxford Univ. Press, New York.