Journal of Data Science **5**(2007), 289-296

Nonparametric Estimation of the Incubation Period of AIDS with Left Truncation and Right Censoring

Swami Onkar Shivraj[1], Debasish Chattopadhya[2] and Gurprit Grover[1]
[1] *University of Delhi and* [2] *National Institute of Communicable Diseases*

Abstract: In the natural history of Human Immunodeficiency Virus Type-1 (HIV-1) infection, many studies included the participants who were seropositive at time of enrollment. Estimation of the unknown times since exposure to HIV-1 in the prevalent cohorts is of primary importance for estimation of the incubation period of Acquired Immunodeficiency Syndrome (AIDS). To estimate incubation period of AIDS we used prior distribution of incubation times, based on a external data as suggested by Bacchetti and Jewell (1991, Biometrics, 47,947-960). In the present study, our estimate was nonparametric based on a method proposed by Wang, Jewell and Tsai (1986, Annals of Statistics, 14, 1597-1605).

Key words: AIDS, incubation period, left truncation, nonparametric, right censoring.

1. Introduction

The incubation period of the disease Acquired Immunodeficiency Syndrome (AIDS) is considered to be the time between infection with Human Immunodeficiency Virus type-1 (HIV-1) and diagnosis of AIDS. Understanding the nature of incubation period in the natural history of AIDS will facilitate improvements in timing and evaluation of therapies as well as in projecting the course of the epidemic. In the context of AIDS cohort studies, a great deal of effort has been made to provide accurate estimate of the incubation distribution, using both prospective and retroprospective data (Brookmeyer and Gail, 1988, Brookmeyer and Goedert, 1989; Munoz *et al.*, 1989; DeGruttola and Lagakos, 1989; Bacchetti, 1990; Kuo, Taylor and Detels, 1991; Munoz *et al.*, 1992; and Medley *et al.*, 1988; Kalbfleisch and Lawless, 1989). The limitation of these follow-up studies was that these cohort included subjects who were already HIV seropositive at the time of infection, commonly known as prevalent cohort. The prevalent cohorts have limited use in elucidating the natural history of AIDS because crucial information on the duration of infection is absent (Brookmeyer and Gail, 1987). However, this difficulty can be overcome if a prior infection distribution can be estimated

from external data. HIV antibody tested on stored sera can provide information on the distribution of seroconversion times, which are thought to be close to infection times (Horshburg *et al.*, 1989). Therefore, incubation time from seroconversion to AIDS diagnosis can be nonparametrcally estimated by using external information on seroconversion (Bacchetti and Jewell, 1991). An appropriate Kaplan-Meier estimator for the left-truncated data have been described by Turnbull (1976), Woodroofe (1985) and Wang, Jewell and Tasai (1986).

The aim of the present study was to estimate the incubation time of AIDS for a prevalent cohort by method proposed by Wang, Jewell and Tasai (1986).

2. Study Population

The study enrollment was initiated in April 1987 (Sricant *et al.*, 1994) and followed through December 1994 at National AIDS reference center, Vellore, Tamil Nadu (India). In this study a total of 32 patients were followed up from the date of earliest positive Western blot test to the development of AIDS as shown in figure 1. All the patients were seropositive at enrollment.

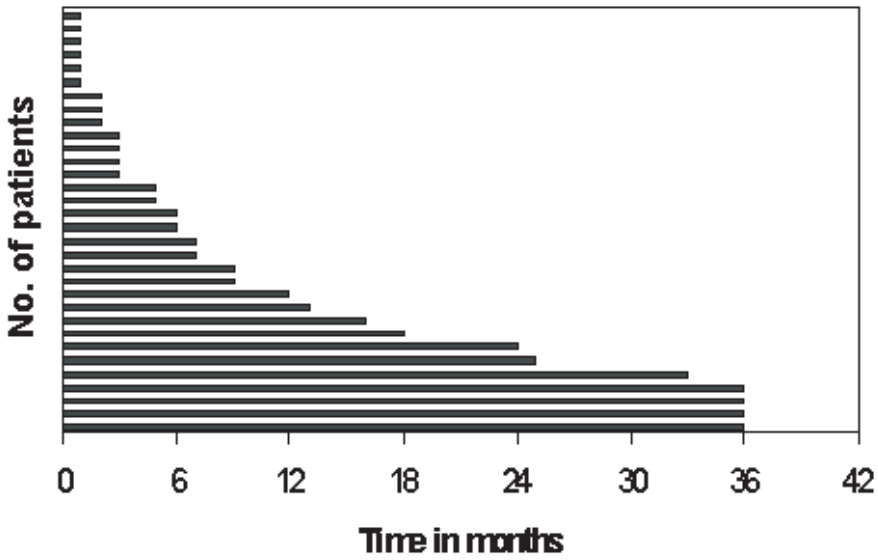

Figure 1: Time interval for the earliest diagnosis of HIV-1 infection to the development of AIDS in 32 patients.

3. Assumption, Notation and Methods

Left-truncation results when only response-free subjects are recruited into a study. If W_i is the time from exposure to (potential) study entry for an individual i, it will be impossible to observe a response in individual i, if the incubation time is smaller than W_i. We refer W_i as truncation time. All follow-up studies recruit AIDS-free HIV-1 infected individuals. If E is study entry date and Sc_i is the HIV-1 infection date of subject i, as measured by seroconversion, then $W_i = E - Sc_i$. If subject i had developed AIDS fewer than W_i time units, then that subjects would not be AIDS-free and thus would be excluded (truncated) from the cohort.

Let X be the random variable representing the time between HIV-1 infection and AIDS diagnosis with associated survival function S^* which is assumed to be left continuous. Let (W, C) be the random variables describing the left-truncation and right censoring time respectively. Further, to estimate AIDS-free survival time, Wang, Jewell and Tasai (1986) have described appropriate Kaplan-Meier estimator for the left-truncated data. Therefore, we estimate the conditional probability of AIDS-free survival beyond time x (i.e., time between HIV infection and AIDS), given AIDS-free survival at time W, that is

$$\begin{aligned} \hat{S}^*(x) &= \prod_{W \le x_j \le x} (n_j - d_j)/n_j \quad \text{for } W \le x_j \le x \\ &= 0, \qquad\qquad\qquad\qquad \text{if there is no } x_j \text{ with } W \le x_j \le x. \end{aligned}$$

Here, d_j is the number of AIDS diagnosed cases at time x_j , and n_j is the number in the risk set at time x_j (i.e. number of HIV-1 positive at the time of enrollment in the study)

Similarly, for Greenwood's formula (1926) or Nelson-Aalen (1978) estimator, only AIDS cases that occur beyond W are considered.

The asymptotic normality of $\hat{S}^*(x)$ has been established by Wang, Jwell and Tasai (1986), who also derived the asymptotic variance of $\hat{S}^*(x)$. Further, the confidence interval for $\hat{S}^*(x)$ are obtained by Hall and Wellner (H-W band) (1980) or Nair (1984) methods. These results justify the use of the Kaplan-Meier estimator and its variance estimator with left-truncated data after the appropriate adjustment of the risk set.

4. Data Analysis

To estimate the time of HIV-1 infection in 32 patients, we made several assumptions, based on the unique dynamics of the HIV-1 infection in Tamil Nadu region (Kilmarx *et al.*, 2000).

In 1985, recipients of multiple blood transfusion and donors were investigated and were not found have any evidence of HIV-1 infection among them (Simoes

et al., 1985). In February 1986, the presence of HIV-1 antibody were detected among female prostitutes in the city of Chennai in Tamil Nadu (Simoes *et al.*, 1993). These evidences indicate that HIV-1 seroconversion among the heterosexual exposures prior to January 1983 were extremely rare. Consequently, the longest unobserved infection time for seroprevalent participants is unlikely to be greater than 52 months. Since most of the enrollment occurred during a one month period, we took W_i as 52 months for all i ($W_i < 52$). Further AIDS-free individuals were right censored as of April 31, 1990.

The median time from seroconversion to development of AIDS was 57 months (95% CI: 53-61) (Mean 64; 95% CI: 58-69). Kaplan-Meier product limit estimates of AIDS-free survival at 5, 6 and 7 years after seroconversion were 34%, 20% and 12% respectively.

Figure 2 gives $\hat{S}^*(x)$ for $W_i < 52$ together with 95% two sided EP and H-W confidence bands for $\hat{S}^*(x)$. The EP band shown corresponds to $a_L = 0.2$ (L stands for lower bound) and $a_U = 0.8$ (U stands for higher bound) with critical value $e_\alpha = 2.9674$. The H-W band shown was computed with value $h_\alpha = 1.2928$. From Figure 2, it can be seen that EP band is narrower in the tails, where as H-W bands performs narrower in the middle.

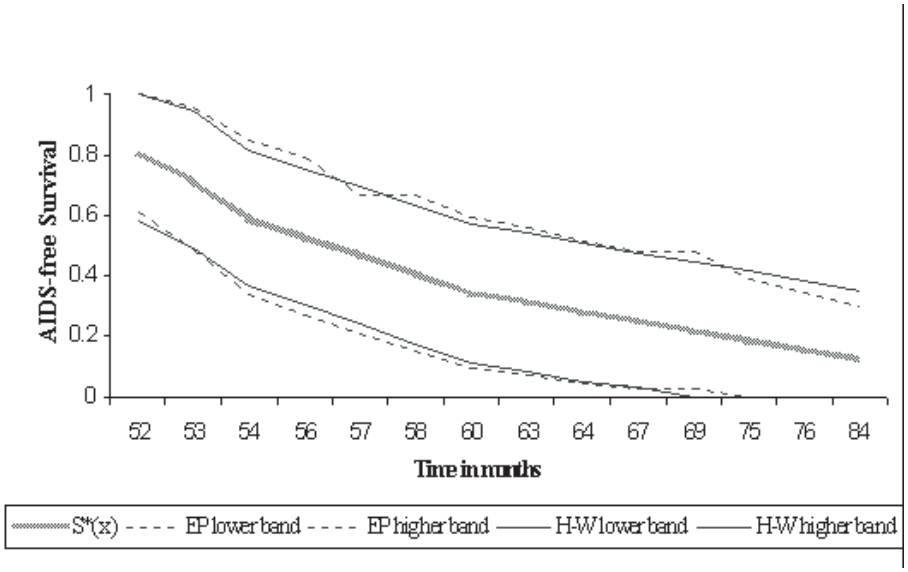

Figure 2: Cumulative AIDS-free survival of HIV-1 infected individuals after seroconversion with 95% EP and H-W confidence bands.

5. Discussion

Many prospective studies on natural history of HIV-1 infection are based on prevalent cohort which provide crucial information on the course of the disease progression. The members of the prevalent cohort are by definition at the risk of disease at the time of recruitment; hence the prevalent cohort will supply more at-risk person-time and disease events for the analysis than incident cohort (Muñoz *et al.*, 1992). Effective survival analysis of the incubation period, however, requires the establishment of onset of HIV-1 infection. When the distribution of measurement of origin of a survival time or the time of disease onset is unknown, it can be first estimated based on external data or the data concerning infection time (Bacchetti and Jewell, 1991). To complete the missing date of seroconversion in the seroprevalent cohort, we used the data concerning the infection time based on the unique dynamics of the HIV-1 infection in Tamil Nadu region.

The median time from seroconversion to the development AIDS was 57 months in the present study. Progression time in HIV-1 infected Indians were more rapid with only 34% remaining AIDS-free after five years of seroconversion which is similar to the reports from other developing countries like Uganda, Kenya and Haiti (Morgan *et al.*, 1997; Nagelkerke *et al.*, 1990; and Deschamps *et al.*, 2000). The developed countries have much longer progression times with 78%-85% of HIV-1 infected individuals remained AIDS-free after five years from seroconversion (Flegg, 1994; Lee *et al.*, 1991).

The EP and H-W bands are competitive, with the former being narrower in the tail and latter being narrower in the middle. The relative performance of EP bands to the H-W bands gets better with censoring and asymptotic critical values provide reasonable approximation in finite samples for both bands (Nair, 1984).

In conclusion, the progression of disease in patients infected with HIV-1 in India seems to be rapid. This is most likely to be due to late recognition of AIDS-related conditions in the early phase of an epidemic and delayed medical care for patients. To improve the progression times, greater attention to the early diagnosis and treatment of opportunistic infection is required.

References

Aalen, O. O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Statist.* **6**, 534-545.

Bacchetti, P. (1990). Estimating the incubation period of AIDS by comparing population infection and diagnosis patterns. *J. Amer. Statist. Assoc.* **85**, 530-539.

Bacchetti, P. and Jewell, N. P. (1991). Nonparametric estimation of the incubation

period of AIDS based on a prevalent cohort with unknown infection times. *Biometrics* **47**, 947-960.

Brookmeyer, R. and Gail, M. H. (1987). Biases in prevalent cohorts. *Biometrics* **43**, 739-749.

Brookmeyer, R. and Gail, M. H. (1988). A method of obtaining short-term projection and lower bounds on the size of AIDS epidemic. *J. Amer. Statist. Assoc.* **83**, 301-308.

Brookmeyer, R. and Goedert, J. J. (1989). Censoring in an epidemic with application to hemophilia-associated AIDS. *Biometrics* **45**, 325-333.

Cox, D. R. (1984). *Analysis of Survival Ddata*. Chapman and Hall.

DeGruttola, V. and Lagakos, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* **45**, 1-12.

Deschamps, M.-M., Fitzgerald, D. W., Pape, J. W., Johnson, W. D. (2000). HIV infection in Haiti: natural history and disease progression. *AIDS* **14**, 15-21.

Flegg, P. J. (1994). The natural history of HIV infection: a study in Edinburg drug users. *J. Infect.* **29**, 311-321.

Greenwood, M. (1926). The natural duration of cancer. In *Reports on public health and medical subjects 33*. London: His Majesty's Stationary Office. 1-26.

Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67**, 133-143.

Horshburg, E. R., Ou, C. Y., Jason, J., Holmberg, S. D., Longini, I. M., Schable, C., Mayer, K. H., Lifson, A. R., Schochetman, G., Ward, J. W., Rutherford, G. W., Evatt, B. L., Seage, G. R. and Jaffe, H. W. (1989). Duration of human immunodeficiency virus infection before detection of antibody. *Lancet* **2**, 637-640.

Kalbfleisch, J. D. and Lawless, J. F. J. (1989). *Inference based on retrospective ascertainment: Analysis of the data on transfusion-related AIDS. J. Amer. Statist. Assoc.* **84**, 360-372.

Kaplan, E. L., Meier P. (1958) Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Kilmarx, P. H., Limpakarnjanrat, K., Kaewkungwal, J., Srismith, R., Saisorn, S., Uthaivoravit, W., Young, N. L., Mastro, T. D. (2000). Disease progression and survival with human immunodeficiency virus type 1 subtype E infection among female sex workers in Thailand. *J. Infect. Disease* **181**, 1598-1606.

Kuo, J.-M., Taylor, J. M. G. and Detels, S. R. (1991). Esimation the AIDS incubation period from a prevalent cohort. *Amer. J. Epidemiol.* **133**, 1050-10578.

Lee, C. A., Philips, A. N., Elford, J., Jonossy, G., Griffiths, P., Kernoff, P. (1991). Progression of HIV disease in hemophiliac cohort for 11 years and the effect of treatment. *British Medical Journal* **203**, 1093-1096.

Medley, G. F., Billard, L., Cox, D. R. and Anderson, R. M. (1988). The distribution of incubation period for the acquired immunodeficiency syndrome (AIDS). *J. Royall Statsit. Soc. Ser. B* **233**, 367-377.

Morgan, D., Mandu, G. H., Malamba, S. S., Okongo, M. J., Wagner, H.-U., Mulder, D. W., Whiteworth, J. A. (1997). HIV-1 disease progression and AIDS defining disorder in rural Uganda. *Lancet* **350**, 240-250.

Munoz, A., Wang, M.-C., Bass, S., Taylor, J. M. G., Kingsley, L. A., Chemil, J. S. and Polak, B. F. (1989). Acquired immunodeficiency syndrome (AIDS)-free time after seroconversion in homosexual man. *Amer. J. Epidemiol.* **130**, 530-539.

Munoz, A., Carey, V., Taylor, J. M. G., Chemil, J. S., Kingsley, L. A., Raden, M. V. and Hoover, D .R. (1992). Estimation of time since exposure for a prevalent cohort. *Statist. in Medicine* **11**, 939-952.

Nair, V. N. (1984). Confidence bands for a survival functions with censored data. *Technometrics* **26**, 265-275.

Nagelkerke, N. J. D., Plummer, F. A., Hotton, D., Anzala A. O., Manji, F., Naugi, E. N. (1990). Transition dynamics of HIV disease in a cohort of African prostitutes: A Markov model approach. *AIDS* **4**, 743-747.

Simoes, E. A. F., Kirubakaran, M., John, T. J., Tzan, N., Madden, D. Sever, J.L. (1985). Absence of HTLV-III antibody in blood donors and recipients in India. *Lancet* **2**, 1284.

Simoes, E. A. F., Babu, P. G., Jeykumari, H. M., John, T. J. (1993). The initial detection of human immunodeficiency virus and its subsequent spread in prostitutes in Tamil Nadu. *India. J. AIDS* **6**, 1030-1034.

Sricant, P., John, T. J., Jeyakumari, H., Babu, P. G., Mithai, D., Kacob, M., Cherian, A. M., Ganesh, A., Zachiriah, A. (1994). Epidemiological features of Acquired immunodeficiency syndrome A. *Indian J. Med. Res.* **105**, 191-197.

Turnbul, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Royall Statsit. Soc. Ser. B* **38**, 290-295.

Wang, M.-C., Jewell, N. P. and Tasai, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.* **14**, 1597-1605.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163-167.

Received August 6, 2005; accepted December 6, 2005.

Swami Onkar Shivraj
Department of Statistics
University of Delhi
Delhi-11007, India
oswami@stats.du.ac.in

Debasish Chattopadhya
Division of AIDS
National Institute of Communicable Diseases
22-Sham Nath Marg, Delhi-110054, India
dchattopadhya@yahoo.co.in

Gurprit Grover
Department of Statistics
University of Delhi
Delhi-11007, India