Electronic Journal of Applied Statistical Analysis
EJASA (2011), Electron. J. App. Stat. Anal., Vol. 4, Issue 2, 155 – 163
e-ISSN 2070-5948, DOI 10.1285/i20705948v4n2p155

ESTIMATION OF SURVIVAL TIMES OF HIV-1 INFECTED CHILDREN FOR DOUBLY AND INTERVAL CENSORED DATA

Gurprit Grover, Tanushree Banerjee*

Department of Statistics, University of Delhi, India

Received 24 August 2010; Accepted 23 May 2011
Available online 14 October 2011

***Abstract**: International studies have established the association of antiretroviral (ARV) therapy and survival in HIV-infected patients. Our study was carried out to confirm the influence of ARV therapy on survival in a group of vertically transmitted HIV-1 infected children. We employed Turnbull's methods based on double censoring (1974) and interval censoring (1976) to estimate the survival time distribution for children receiving and not receiving ARV therapy. It was noted that children undergoing ARV therapy had a longer survival than those not under treatment. In our study group, children under treatment survived for 15 years with a probability of almost 95% while children not under treatment had a survival of 11 years with a probability of 75%. Further, the estimates of survival times as obtained by the interval censoring mechanism were found to be more precise than those obtained by the double censoring mechanism. We even utilized imputation approach which showed that the width of the band of estimates was narrower for interval censored data as compared to double censored data.*

***Keywords**: HIV/AIDS, double censoring, imputation technique, interval censoring, self-consistency.*

1. Introduction

Survival analysis or time-to-event data analysis is predominately in biomedical science where the interest is in observing the length of time to death of patients. For their analysis, what interests' researcher most is the determination of the distribution of the survival time, taken for an event of interest to occur. Turnbull [17] proposed a non-parametric estimation of the distribution function

* Corresponding author. E-mail: tanushree@aidsresearcher.org

of a survival time when observations are doubly censored. In this method, we treat the estimation problem where left censoring occurs when some individuals have experienced the event of interest but the exact time is not known. Right censored observations are those which do not realize their events even on the termination of the study. In both situations, some individuals with exact event times are observed.

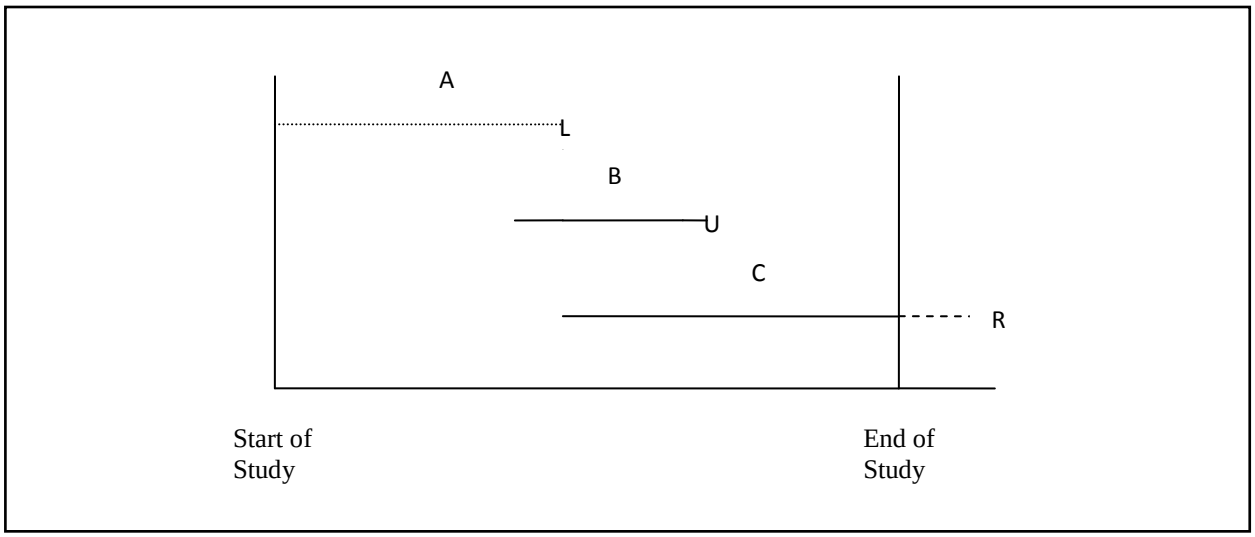

Figure 1. Representation of types of censoring among individuals in a study. U = Uncensored, L = Left censored, R = Right Censored.

Figure 1 focuses on the concept of censoring. In the figure individual A is left censored as exact time of the occurrence of event is not known although the event of interest has been realized. Individual B is an uncensored case since the exact time of the event of interest is known which falls within the span of the study. By right censoring or right censored failure time data, we mean that the failure time of interest is observed either exactly or to be greater than a censoring time. The individual C depicts the case of right censoring in the above figure.

Alternatively, in some applications the data may be interval-censored. Here the time-to-event variable X whose value is never observed is known to lie in the time interval between two consecutive inspection times L_i and R_i i.e. $X \in (L_i, R_i]$.

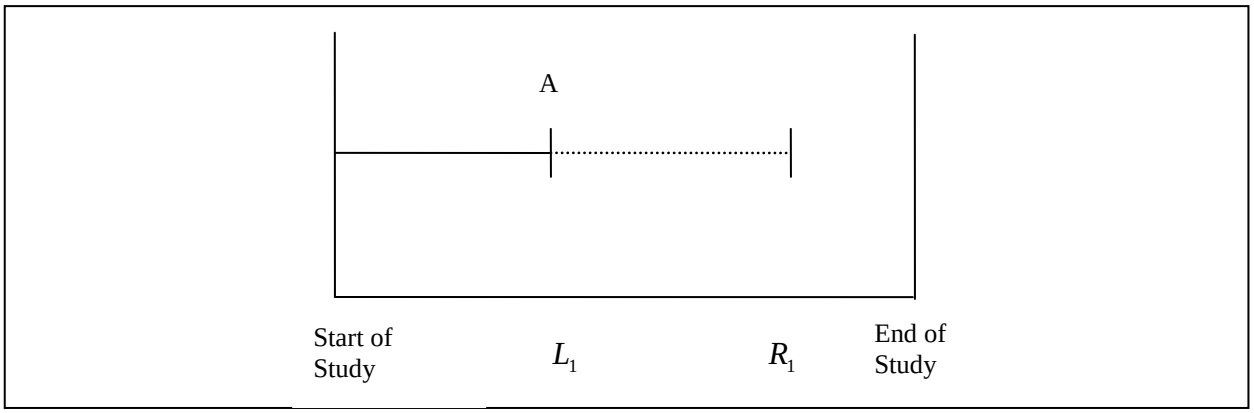

Figure 2. Representation of interval censoring in an individual of a study.

Such *interval censoring* occurs when patients in a clinical trial or longitudinal study have periodic follow-up and the patient's event time is only known to fall in an interval $(L_i, R_i]$ (L for left endpoint and R for right endpoint of the censoring interval). This type of censoring has been depicted by individual A in the second figure.

Estimation of survival distributions based on doubly censored data has been the focus of extensive research. Gehan [5], Mantel [10] and Peto [11] have given examples where double censoring might arise in medical applications. Another example occurred in a study of African infant precocity by Leiderman *et al.* [9]. Their purpose was to establish norms for infant development for a community in Kenya in order to make comparisons with known standards in the United States and the United Kingdom. The sample consisted of 65 children born between July 1 and December 31, 1969. Starting in January 1970, each child was tested monthly to see if he had learned to accomplish certain standard tasks. Here T would represent the time from birth to first learn to perform a particular task. However, relatively little research has been done with interval censored data compared to doubly censored data. De Gruttola and Lagakos [4] developed a method for the analysis of "doubly censored" survival data. It is essentially an analysis of two independent, interval censored endpoints (time to human immunodeficiency virus type 1 (HIV-1) infection and time from HIV-1 infection to acquired immunodeficiency syndrome (AIDS)). Hughes *et al.* [7] developed a non-parametric estimator to estimate the joint distribution of two interval censored dependent endpoints. Under these procedures nonparametric estimates can be obtained by the actuarial method [2] or by the product-limit (PL) or reduced-sample (RS) methods described in [8]. Geskus [6] has also developed methods for studies on HIV infection and AIDS when data on time from seroconversion to AIDS or death are doubly censored, both at the time of origin and at the endpoint of interest.

Our study is motivated by Turnbull's methods as it is the only method that allows estimation for left censored observations and right censored observations for the same event of interest. We begin with Turnbull's method on double censoring [17]. Here we focus on the estimation of the survival time distribution in both ARV and non-ARV groups for subjects whose exact time of death is not known. We have then used Turnbull's interval censoring method [18] to estimate the survival time distribution of the subjects considering the event of death to lie in an interval that is the last CD4 count when the subject is known to be last alive and the end of the study. We also compute estimates of the survival time distribution for both the groups by using imputation technique. Further, we compare the estimates obtained for the above mentioned censoring mechanisms. In the next section a brief explanation of Turnbull's double censoring and interval censoring mechanism has been discussed. Section 3 illustrates the results and Section 4 includes some concluding remarks.

2. Notations and Methods

In this section we examine the techniques for estimating the survival function for double censored data and interval censored data. Each sampling scheme provides different information about the survival function and requires a different technique for estimation. First, we examine a procedure for estimating survival function for double censored samples which include some individuals that are left-censored and some individuals that are right-censored. Turnbull [17] has suggested a modified Product-Limit estimator for dealing with problems which include both left

and right censoring. This estimator is based on an iterative procedure which extends the notion of a self-consistent estimator.

In some applications the data may be interval-censored. Here the only information we have for each individual is that their event time falls in an interval $(L_i, R_i], i = 1, 2, ..., n,$ but the exact time is unknown. In case of interval censored data an estimate of the survival function can be found by a modification of the above double censored techniques' iterative procedure as proposed by Turnbull (1976).

Imputation is a general approach for handling missing data problems [14]. But, interval censored data are really incomplete data, not exactly missing data. Nevertheless, one can still treat the underlying, unobserved true interval censored failure times as missing and replace them by using some imputed times conditional on the observed information.

We would like to estimate the survival time distribution for the ARV and Non-ARV therapy groups by using Turnbull's double censoring, interval censoring methods and imputation procedure [16]. Furthermore, comparison of the censoring mechanisms based on the estimates obtained by the two procedures would also be carried out. In this study, the estimates obtained for the therapy groups under the censoring mechanisms were used to obtain the log rank χ^2 value by Peto and Peto [12].

Using the left end point and right end point estimates under the double censoring and interval censoring mechanisms for the ARV and Non-ARV therapy groups we computed the bands for the survival estimates obtained by the two censoring mechanisms.

3. Data Analysis

The study population consisted of 130 HIV-1-infected children who were visiting the Paediatrics HIV Outpatient Clinic of Dr. Ram Manohar Lohia (RML) Hospital, New Delhi. It is a tertiary care centre catering to a large number of patients for complicated diagnoses and management from all parts of India. The eligibility criteria for this study was (i) vertically HIV-1-infected children born between 1992 and 2002 (ii) laboratory evidence of seroconversion by ELISA test for antibodies when they first visited the clinic on showing symptoms. The selection of cases was independent of the factor of progression to AIDS. The patients were followed up to December 31, 2007, the end of the study period and were diagnosed with HIV infection according to CDC classification [3]. They were submitted to clinical and immunological assessments at three months interval. Children reporting a CD4+T cell count below the age related CD4 count cut off of less than 20% or first AIDS defining illness were subjected to the ARV therapy. HIV seropositivity (3 times positive ELISA with a positive Western blot at $\geq$ 18 months of age) was used as the reference standard. Out of the 130 HIV-1-infected children, 105 were on ARV therapy. In the therapy group, 6 patients died during the follow-up while in the non-therapy group there were 10 deaths observed during the follow-up. SPSS 15.0 has been used for the programming in this study.

In this study, there are cases for which the exact time of event is not known. For each of the 130 subjects who were seropositive, the time of death, X_i was determined in two ways: (1) since exact time of event is not known and event has occurred, it was considered to be left censored and (2) event considered to lie in the interval $(L_i, R_i]$. The interval was defined by the dates of

the subject's last CD4 cell count being measured, which gave the last information about the individual's survival and end of the study wherein they were enrolled. The objective of the analysis is to compare the estimates of survival time for both the ARV and Non-ARV therapy groups obtained by: 1) Turnbull's double censoring mechanism 2) Turnbull's interval censoring mechanisms 3) imputation technique. In addition, we show comparison of the two censoring mechanisms for deciding the precision of the procedures.

Under the double censoring set up, subjects who had died before the end of the study i.e. 31st December, 2007 and whose exact time of death was unknown were treated as the left censored observations. As regards the right censored observations, they were the patients who continued to survive beyond the end of the study. The log-rank non parametric test was used for comparison of survival time to compare between the ARV therapy and Non-ARV therapy groups'. The non-parametric estimates for the therapy groups' as obtained by double censoring mechanism yielded a significant chi square value of $\chi^2 = 7.25$, df = 1 ($p < 0.001$). The results showed a significant difference between the survival estimates obtained for the ARV therapy and Non-ARV therapy groups.

Under the interval censoring set up the time of death for subjects whose exact time was not known was considered to occur in the interval between the last CD4 cell count measured for the patient and the end of the study. Interval censored estimates for both the ARV therapy and Non-ARV therapy groups produced a chi square value of $\chi^2 = 26.67$, df = 1 ($p < 0.001$). Our results confirmed that survival estimates obtained for the ARV therapy group were better than those obtained for the Non-ARV therapy group.

Since interval censoring is more challenging and computationally difficult one may use a general approach of imputation for handling incomplete information about failure variables of interest. We therefore obtained the midpoint estimates by imputation technique for the failure time data for both the ARV and Non-ARV therapy groups.

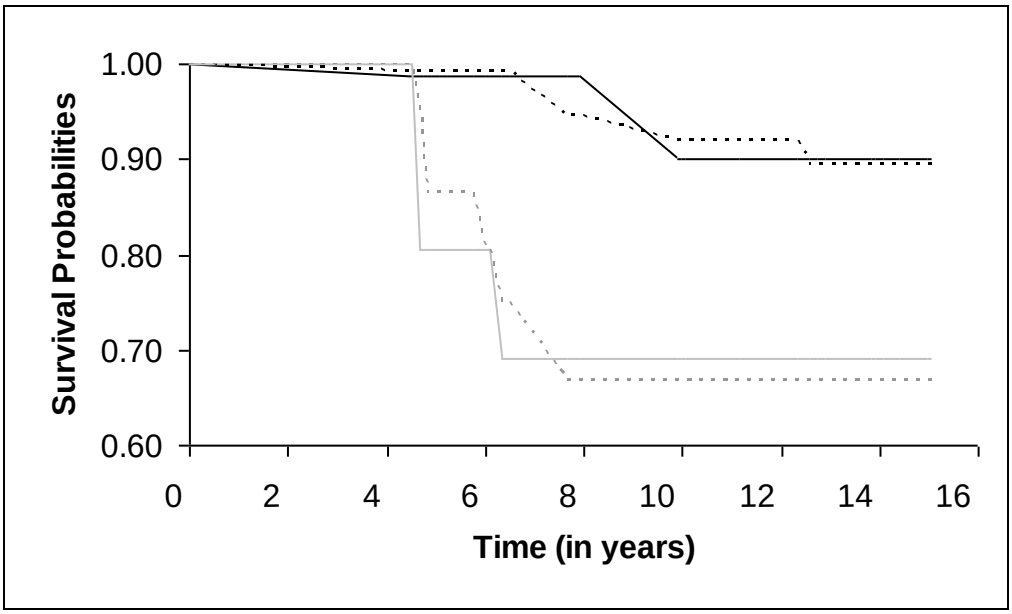

Figure 3. Estimated survival probabilities obtained by double censoring for both ARV therapy and Non-ARV therapy groups. ARV therapy group estimates (———), Midpoint estimates for ARV therapy group (– – –). Non-ARV therapy group estimates (——), Midpoint estimates for Non- ARV therapy group (– – – –).

Figure 3 show the survival probabilities for the ARV and Non-ARV therapy groups by using double censoring. The figure also show the midpoint estimates of survival time. Figure 3 did not show any significant differences between the midpoint estimate and Turnbull (1974) estimate in the ARV therapy group ($p = 0.098$). The two estimates obtained for the Non-ARV therapy group were also not found to be significantly different ($p = 0.115$).

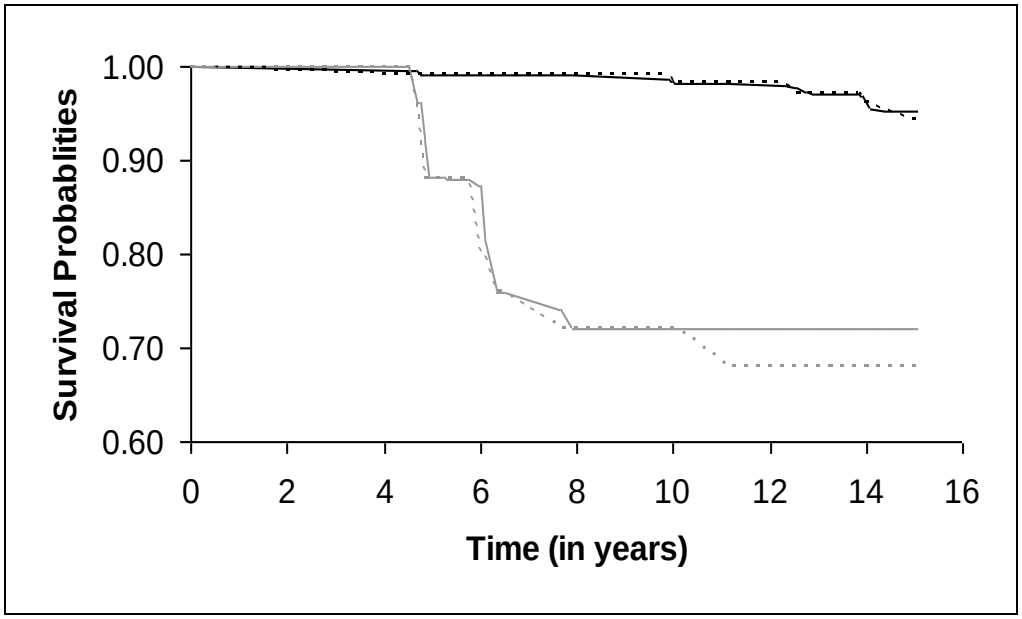

Figure 4. Estimated survival probabilities obtained by interval censoring for both ARV therapy and Non-ARV therapy groups. ARV therapy group estimates (———), Midpoint estimates for ARV therapy group (- - - -). Non-ARV therapy group estimates (——), Midpoint estimates for Non-ARV therapy group (- - -).

Figure 4 show the survival probabilities for the ARV and Non-ARV therapy groups by using interval censoring. The midpoint estimate and Turnbull [18] estimate have been shown in the figure. It was noted that the differences in the estimates were not statistically significant in both the ARV ($p = 0.082$) and Non-ARV therapy groups ($p = 0.094$).

The comparison between the non-parametric estimates obtained for both the censoring mechanisms was made using the log rank test. A formal test of the significance of the interval censored and double censored estimates in the ARV therapy group yielded a chi square value of $\chi^2 = 86.74$, df = 1 with a corresponding $p < 0.001$. Similarly for the Non-ARV therapy group the chi square obtained for the estimates of the two sampling schemes was significant with a $p < 0.005$. The results for both the groups revealed that there was a significant difference in the estimates obtained by the interval censoring and double censoring mechanism.

Figure 5 and Figure 6 show the non-parametric estimates of the survival time distribution obtained for the ARV therapy group and Non-ARV therapy group respectively by the two mechanisms along with the band of estimates based on the left end point and right end point imputation.

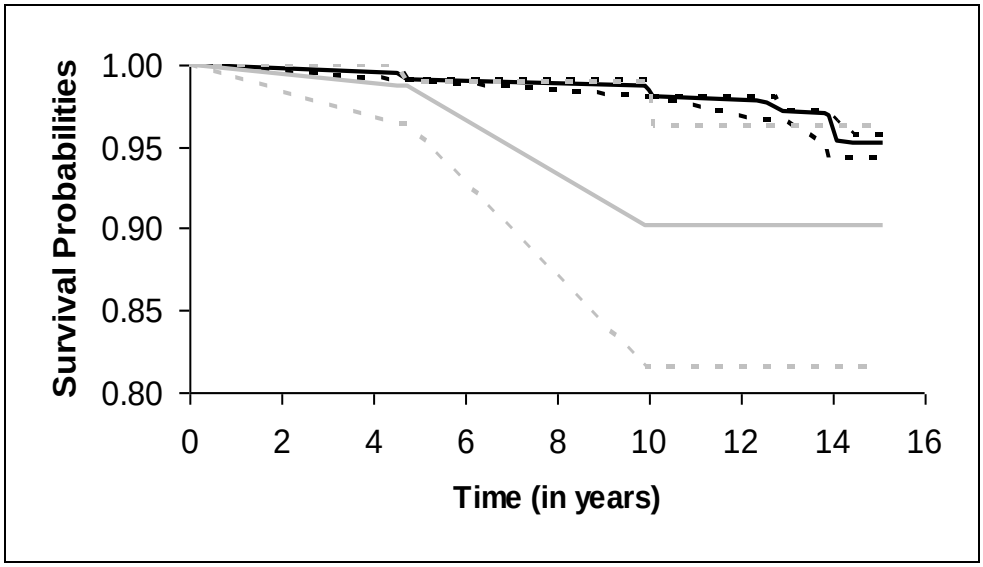

Figure 5. Survival function of HIV-1-infected children in the ARV therapy group, estimated by interval censoring scheme (——) and its band (– – – –) ; double censoring scheme (——) with its corresponding band (– – –).

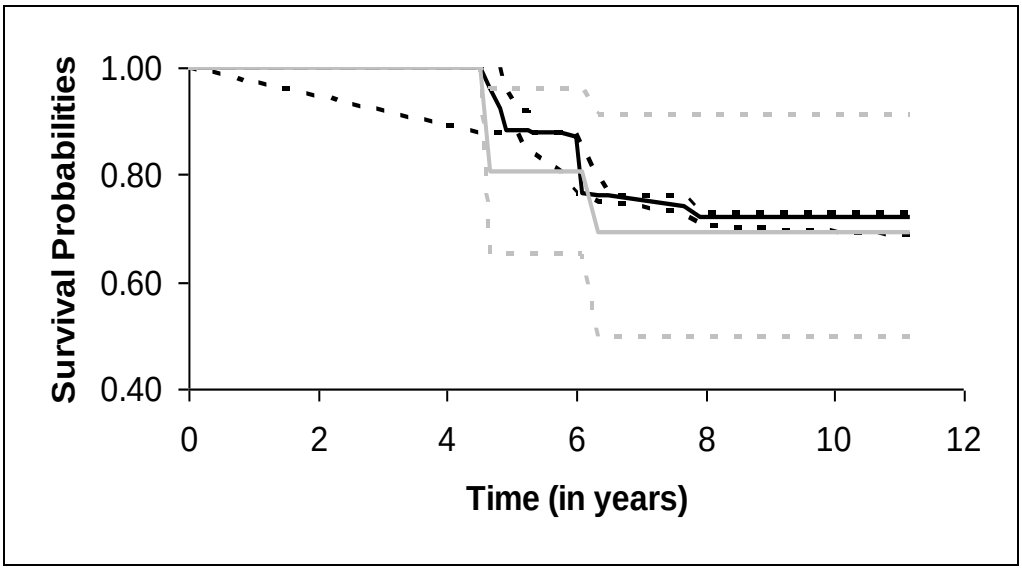

Figure 6. Survival function of HIV-1-infected children in the Non-ARV therapy group, estimated by interval censoring scheme (——) and its band (– – – –); double censoring scheme (——) and its band (– – –).

The band of estimates based on the left end point and right end point imputation obtained for interval censored data was narrower than the band obtained for double censored data for both the groups as can be seen from Figure 5 and Figure 6. It clearly shows that for the estimates obtained from interval censoring mechanism were more precise than those obtained from double censoring mechanism. Further, we can see that for interval censoring mechanism the estimates obtained based on the occurrence of the event at the right endpoint and left endpoint of each interval did not differ significantly ($p > 0.05$). Figure 5 and Figure 6 show that the estimates based on the endpoints nearly coincided with the actual estimates obtained by interval censoring

for both ARV and Non-ARV therapy groups. Therefore, imputation technique can be applied instead of complex interval censoring as the approach can provide simple approximation to the observed data.

4. Discussion

Our study shows that patients who were under therapy had a better survival prognosis compared to children not undergoing treatment. This was supported by the survival estimates obtained by interval censoring and double censoring techniques.

In this paper, we have described the two approaches- double censoring mechanism and interval censoring mechanism of Turnbull's self consistency algorithm along with imputation procedure- to estimate the distribution of survival time in a cohort of vertically transmitted HIV-1-infected children for those undergoing ARV therapy and those not under ARV therapy.

These algorithms are intuitively appealing to statisticians as well as non-statisticians. Our results corroborate the findings of previous studies [13,15] that children undergoing ARV therapy have a longer survival than those not under treatment. Children under treatment have survived for 15 years with a probability of almost 95% while children not undergoing ARV therapy have a survival of only 11 years with a probability of 75%. Therefore, it is recommended that more number of children should be put under treatment to improve their longevity since only about 6500 eligible children are receiving ARV paediatric drugs when more than 18,000 infected children have been identified in India [1].

The method discussed in section 2 is applicable for the situation where there is a discrete time scale or the data can be grouped naturally. The estimates obtained by the two mechanisms suggest that for data where the event of interest is unknown interval censoring method is a more appropriate approach than double censoring method. This finding is further supported by the width of the band of the estimates obtained by imputation approach which was narrower for interval censored data than double censored data. Owing to the lack of well-known statistical methodology and available software by non-statisticians and medical personnel, a common *ad hoc* approach is to assume that the event occurred at the end (or beginning or midpoint) of each interval, and then apply methods for standard time-to-event data. In this study, the estimates obtained by mid point imputation for both ARV and Non-ARV therapy groups were not statistically different (p-value > 0.05) with the estimates obtained by Turnbull's procedures. Furthermore, the estimates obtained by left end point and right end point imputation for interval-censored data were not statistically significant (p-value > 0.05) and coincided with the actual estimates obtained by interval censoring procedure.

The survival of HIV-1-infected children may be associated with different cofactors apart from ARV therapy. Opportunistic infections (OIs) have been reported to invade at the different stages of illness in HIV-1-infected paediatric patients. In this study, we have not considered the role of OIs on survival both in the ARV and Non ARV therapy groups. The wide difference in the survival times in the ARV and Non ARV therapy groups in our study could possibly be attributed to invasion of OIs along with ARV therapy. Moreover, paediatric HIV-infected patients with a CD4+T cell count below the age related cut off of less than 20% suffering from OIs particularly tuberculosis are subjected to treatment for tuberculosis solely. On recovery they may be put on ARV therapy if required. Future research can be carried out by considering the confounding effect of OIs on survival of the HIV-1-infected children.

References

[1]. AVERT (2009). *Overview of HIV and AIDS in India*. URL: http://www.avert.org/aidsindia.htm.

[2]. Berkson, J., Gage, R.P. (1950). Calculation of Survival Rates for Cancer, in *Staff Meetings of the Mayo Clinic*, 25, 270-286.

[3]. Centers for Disease Control (1994). 1994 revised classification system for human immunodeficiency virus infection in children less than 13 years of age. *Morbidity and Mortality Weekly Report*, No. RR-12, 1-10.

[4]. De Gruttola, V., Lagakos, S. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics*, 45, 1-11.

[5]. Gehan, E.A. (1965). A generalized Two-Sample Wilcoxon Test for Doubly Censored Data. *Biometrika*, 52 (3 and 4), 650-653.

[6]. Geskus, R.B. (2001). Methods for estimating the AIDS incubation time distribution when date of seroconversion is censored. *Statistics in Medicine*, 20 (5), 795-812.

[7]. Hughes, J.P., Richardson, B.A. (11 February 1999). Estimation of the Joint Distribution of Infection Time and Death Time for Vertical Transmission of HIV-1. *Technical Report*, 162, 12-24.

[8]. Kaplan, E.L., Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53, 457-481.

[9]. Leiderman, P.H., Babu, D., Kagia, J., Kraemer, H.C., Leiderman, G.F. (1973). African Infant Precocity and Some Social Influences during the First Year. *Nature*, 242, 247-249.

[10]. Mantel, N. (1967). Ranking Procedures for Arbitrarily Restricted Observations. *Biometrics*, 23, 65-78.

[11]. Peto, R. (1973). Experimental Survival Curves for Interval-Censored Data. *Applied Statistics*, 22 (1), 86-91.

[12]. Peto, R., Peto, J. (1972). Asymptotically Efficient Rank Invariant Procedures. *Journal of the Royal Statistical Society, Series A*, 135, 185-207.

[13]. Romanelli, R., Pinto, J.A., Melo, L.J., Vasconcelos, M.A., de Matos Pereira, R. (2006). Effectiveness of dual and triple antiretroviral therapy in the treatment of HIV-infected children. *Jornal de Pediatria*, 82 (4), 260-265.

[14]. Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.

[15]. Sanchez-Granados, J.M., Carreno, M.O., Gonzalez-Tome, M.I., Ferrando, P., Ruiz-Contreras, J., Ramos, J.T. (2002). Impact of HAART on the survival and disease progression in HIV-infected children, in *International Conference of AIDS*, Madrid, Spain, 7-12 July 2002, 14.

[16]. Sun, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer: New York.

[17]. Turnbull, B.W. (1974). Nonparametric Estimation of a Survivorship Function with Doubly Censored Data. *Journal of the American Statistical Association*, 69 (345), 169-173.

[18]. Turnbull, B.W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B*, 38, 290-295.