

## GAP TIME BIAS IN INCIDENT AND PREVALENT COHORTS

Mei-Cheng Wang

*Johns Hopkins University*

*Abstract:* Multiple event data are frequently encountered in incident and prevalent cohort studies when the multiple events are considered as the major outcomes. For incident cohorts, statistical analysis for the time to the first event, the first gap time, can be conducted using standard techniques in survival analysis under appropriate conditions. These techniques are, nevertheless, inappropriate for analyzing the second gap time because of the presence of induced informative censoring. For prevalent cohorts, because the sample is biased in general, standard methods do not apply to gap times of any order, but techniques for truncated data can be used for the analysis of the first gap time. It is shown that the combined incident and prevalent data form the usual survival data for analysis of the second gap time when certain stationarity conditions are satisfied. The problems are illustrated by a cohort example to study the natural history of Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS).

*Key words and phrases:* Gap time, informative censoring, longitudinal studies, multiple events, prevalent cohort.

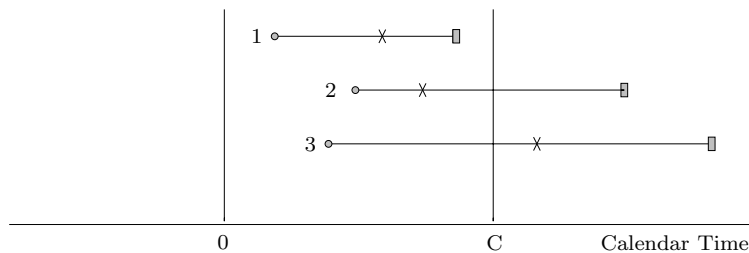
### 1. Introduction

Multiple event data are frequently encountered in medical follow-up studies when the events are considered as the major outcomes for the progression of a disease. To analyze multiple event data, either the time to an event or the time between successive events could be used as the variable of interest. The gap time, defined as the time between successive events, is generally the preferred variable when multiple events are chronologically ordered. Examples of such multiple event data include repeated hospitalizations due to a specific disease, multiple infections or tumors, and chronologically-ordered clinical events such as HIV (Human Immunodeficiency Virus), AIDS (Acquired Immunodeficiency Syndrome), and death in a natural history study of AIDS. Besides these examples, multiple events could also arise in problems unrelated to disease progression. An example is the promotion from assistant to associate professorship, and from associate to full professorship in university tenure-track positions.

An incident population in epidemiology is comprised of individuals who experience the incidence of the initiating event within a specified calendar time interval, say  $[0, C]$  where  $C$  is a positive constant. Suppose  $K$  events occur in

chronological order following the incidence of the initiating event. Denote by  $X$  the calendar time of the initiating event,  $Y$  the time from the initiating event to event  $j-1$ , and  $Z$  the time from event  $j-1$  to event  $j$ , where  $j$  is a fixed event index with value  $2, 3, \dots$ , or  $K$ . For simplicity, and without loss of generality, we set  $K = 2$  and consider  $Y$  as the first gap time, and  $Z$  the second gap time. The variables  $Y$  and  $Z$  are usually correlated because they come from the same subject. In this paper the main interest is focused on the outcome variable  $Z$ , the second gap time.

An incident cohort is a sample selected from the incident population. In general, standard methods in survival analysis serve as appropriate tools for analyzing the first gap time under an independent censoring assumption. Nevertheless, even if the observation of the multiple events is terminated subject to independent censoring, these methods are not applicable for analyzing the second gap time because of the presence of induced informative censoring (Gelber, Gelman and Goldhirsch (1989)). A prevalent cohort is defined as a group of individuals who have experienced the initiating event and have not experienced the first event before calendar time 0. For prevalent cohorts, standard methods do not apply to gap times of any order but, as will be discussed in Section 2.2, techniques for truncated data can be used for the analysis of the first gap time. According to the definition of incident and prevalent cohorts, it is understood that an incident cohort satisfies  $X \geq 0$  and a prevalent cohort satisfies  $X \leq 0$  and  $X + Y \geq 0$ . Figures 1 and 2 provide simple explanatory plots for incident and prevalent cohort data. In Figure 1, individuals 1, 2 and 3 are included in the incident cohort and the observation of their multiple events is terminated at calendar time  $C$ . In Figure 2, the individuals with dashed lines are excluded from the prevalent cohort by the sampling recruiting criterion, and only those who experience the initiating event but not the first event at the sampling time, 0, are recruited into the study. The observation of multiple events from the prevalent cohort individuals 4, 5 and 6 is also terminated at calendar time  $C$ .



◦: Initiating event    x: First event    ◻: Second event

Figure 1. Incident cohort.

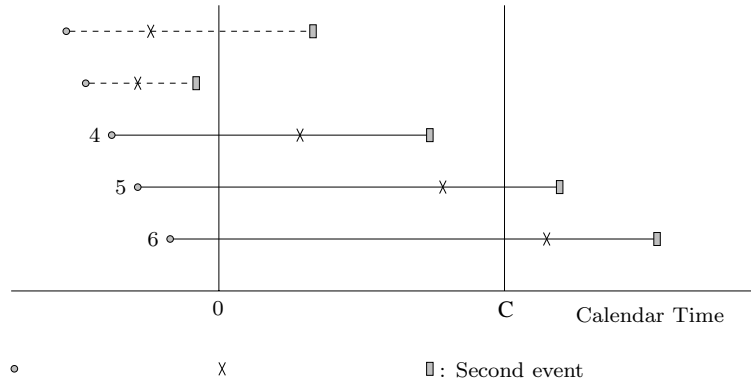


Figure 2. Prevalent cohort.

Assume that the initiating events occur at  $X$  with the bounded intensity function  $\phi_x(x)$ . The intensity function  $\phi_x$  can be interpreted as the disease occurrence rate function for  $X$ . Let  $f_{y,z}$  denote the joint density of  $(Y, Z)$ ,  $f_y$  the marginal density of  $Y$ , and  $f_z$  the marginal density of  $Z$ . Let  $S_y$  and  $S_z$  respectively denote the survival function of  $Y$  and  $Z$ . To reduce the mathematical complexity in the discussion, assume the failure time  $Y$  has finite support and let  $y^*$  denote the supremum of the support of  $Y$ . In order to study the potential bias in the prevalent cohort, we only consider the function  $\phi_x(x)$  within the range  $[-y^*, 0]$ . Let  $f_x$  be the incident population density of  $X$  in the interval  $[-y^*, 0]$ , derived as the normalized  $\phi_x$ :

$$f_x(x) = \phi_x(x)I(-y^* \leq x \leq 0) / \int_{-y^*}^0 \phi_x(u)du ;$$

and let  $S_x$  be the survival function of  $f_x$ . In this paper the question of how to estimate the survival function of the second gap time,  $S_z$ , will be studied and the bias resulting from the standard methods in survival analysis will be explored.

Consider the following two stationarity conditions:

- (S1) The joint distribution of  $(Y, Z)$  is independent of  $X$ .
- (S2) The occurrences of the initiating events started in the distant past and the rate of occurrence has been stabilized in the target population. Quantitatively, assume that the intensity function  $\phi_x(x)$  is constant for  $x \geq -y^*$ .

The two stationarity conditions serve as the fundamental assumptions for the estimation of distribution functions in one-sample models. The stationarity condition (S1) holds when the bivariate disease distribution of  $(Y, Z)$  is independent of when the disease is initiated. This condition will be assumed throughout the paper. The condition (S2) holds typically for stable diseases, such as some cancer and genetic diseases, where the number of the disease initiations remains

constant over calendar time. Assumption (S2) will be assumed only when indicated.

## 2. Bias in Cohorts

In studies of the natural history of a disease, both the incident and prevalent sampling schemes can be adopted for the collection of data, although the probability structure of the incident population is the target interest. This section studies the bias arising in these two different types of cohorts when estimating the survival function  $S_z$ .

### 2.1. Bias in incident cohort data

The incident cohort is a sample selected from the incident population where the individuals experience the initiating event within the fixed calendar time interval  $[0, C]$ . Because of loss to follow-up or end-of-study, survival data from the incident cohort are usually observed subject to right censoring. Let the censoring variable  $V$  be the time from the initiating event to the (potential) censoring point. Denote by  $G$  the survival function of  $V$ . The observed data consist of information about the history of the multiple events occurring prior to the minimum of the censoring time and the time of the second event. Suppose the independent censoring condition holds: the censoring time  $V$  is independent of  $(Y, Z)$ . Note that this independent censoring condition is implied by (S1) if the observation of multiple events is terminated only by the end of study at a fixed calendar time,  $C$ , since  $V = C - X$  is independent of  $(Y, Z)$ ; see Figure 1.

The initial approach that one might use for estimating  $S_z(t)$  is the Kaplan-Meier estimate (Kaplan and Meier (1958)) based on the observed gap times and the corresponding censoring indicator,  $(\min(z_i, v_i - y_i), I(z_i \leq v_i - y_i))$ . When  $Y$  and  $Z$  are not independent, this convenient estimate is generally inappropriate because the failure time  $Z$  and the censoring time  $V - Y$  are correlated. When (S1) holds, note that  $f_z(z|X > 0) = f_z(z)$ , and the effect of the first bias can be neglected. If we further assume the independent relationship of  $Y$  and  $Z$ , then the Kaplan-Meier estimator is appropriate and it can be calculated simply from  $\{(\min(z_i, v_i - y_i), I(z_i \leq v_i - y_i)) : v_i - y_i \geq 0\}$ , an observable data set.

Let  $v^*$  be the supremum of the possible values of  $V$ . The survival function of  $Z$  can be expressed as

$$\begin{aligned} S_z(z) &= Pr(Z > z) \\ &= Pr(Z > z, Y + z \leq v^*) + Pr(Z > z, Y + z > v^*). \end{aligned} \quad (2.1)$$

The second term in (2.1) is not estimable in the current nonparametric model. In addition, it can be shown that the marginal survival function  $S_z(z)$  is non-identifiable unless the second term in (2.1) equals 0. This constraint is satisfied

when the supremum of the values of  $Y + z$  is not greater than  $v^*$ . Practically, it means that the study period must be long enough to allow for the possibility to observe all the values of  $Y + z$ . Such a constraint, nevertheless, may not hold in many prospective follow-up studies. In the case that  $S_z(z)$  is not identifiable, what can be estimated is restricted to, for example, the conditional distribution function  $Pr(Z > z|Y \leq y)$  where  $y + z < v^*$ ; see Visser (1996), Lin, Sun and Ying (1997), Wang and Wells (1998) for various nonparametric ways to estimate the distribution functions.

## 2.2. Bias in prevalent cohort

Define the prevalent population as the class of individuals who have experienced the initiating event and have not experienced the first event before calendar time 0. A prevalent cohort is defined as a sample selected from the prevalent population. In an AIDS study, the prevalent cohort can be defined as a group of patients who have been HIV-infected but have not been diagnosed with AIDS at the time of recruitment (0).

Many articles consider that the observed  $y$  from a prevalent cohort can be treated as left-truncated and right-censored data; see Wang, Brookmeyer and Jewell (1993) and references therein. The presence of left truncation is due to the prevalent sampling, which tends to over sample individuals with larger  $y$ , and the presence of right censoring is due to the usual loss to follow-up or end of study in prospective studies. In the prevalent population, under (S1), the density of  $Y = y$  conditional on  $X = x$ ,  $x \leq 0$ , can be derived as

$$p_{y|x}(y|x) = \frac{f_y(y)I(x+y \geq 0)}{S_y(-x)},$$

and the marginal density of  $Y$  can be expressed as

$$p_y(y) = \frac{S_x(-y)f_y(y)}{\int S_x(-u)f_y(u)du};$$

see Wang (1991) for the derivation of  $p_y(y)$ . Based on the above density, it is clear that the distribution of  $Y$  systematically assigns more weight to larger values.

Although the bias associated with  $Y$  has been thoroughly studied (Brookmeyer and Gail (1987)), the bias of  $Z$  in the prevalent cohort has never been explored in the literature. Under assumption (S1), the joint density of  $(X, Y, Z)$  in the prevalent population can be derived as the density of  $(X, Y, Z)$  conditional on  $X \leq 0$  and  $X + Y \geq 0$ :

$$p_{x,y,z}(x, y, z) = \frac{f_x(x)f_{y,z}(y, z)I(-y^* \leq x \leq 0, x + y \geq 0)}{\int \int S_x(-u)f_{y,z}(u, v)dudv}.$$

By integrating out  $x$ , the density of  $(Y, Z)$  is

$$p_{y,z}(y, z) = \frac{S_x(-y)f_{y,z}(y, z)}{\int \int S_x(-u)f_{y,z}(u, v)dudv}. \quad (2.2)$$

Let  $\tau$  denote the denominator of the right side of (2.2), which is the probability for untruncated observation. The marginal density of  $Z$  in the prevalent population is thus

$$p_z(z) = \frac{w(z)f_z(z)}{\tau}, \quad (2.3)$$

where  $w(z)$  represents a selection-bias function and

$$w(z) = \int_0^{y^*} S_x(-y)f_{y|z}(y|z)dy.$$

In general, the prevalent population density  $p_z$  is connected to the incident population density  $f_z$  through (2.3), and the direction of bias can be identified from (2.3). There are two special cases worth mentioning.

- (i) Assume that  $Z$  is independent of  $Y$ . In this case, the two density functions  $p_z$  and  $f_z$  coincide because  $w(z) = \tau$ . Conditional on  $X \leq 0$  and  $X + Y \geq 0$ , it can be easily shown that the gap time  $Z$  is independent of  $X + Y$  (the calendar time of the first event); thus  $Z$  is independent of  $X + Y$  in the prevalent population. Let  $C$  be the calendar time of the censoring point, defined subject to the prevalent population, and suppose  $C$  is independent of  $Z$  for individuals in the prevalent population. Then, the censoring time for the observation of  $Z$  is  $C - (X + Y)$  which is independent of  $Z$ . It is not hard to conclude that the usual approaches in survival analysis are appropriate for deriving inferences associated with  $Z$ .
- (ii) Assume that both (S1) and (S2) hold. The validity of (S2) implies  $S_x(-y) = y/y^*$ , where  $0 \leq y \leq y^*$ . Thus, the weight function  $w(z)$  can be expressed as  $w(z) = E(Y|z)/y^*$  and the marginal density of  $Z$  in the prevalent population can be written as

$$p_z(z) = \frac{E(Y|z)f_z(z)}{\int E(Y|v)f_z(v)dv}.$$

In this case, the direction of bias in  $Z$  from the prevalent population is determined by the conditional mean of  $Y$  given  $Z = z$ .

### 3. Estimation from Combined Cohort Data

Given the presence of bias from both the incident and prevalent cohorts, a question of interest is the appropriateness of the use of combined cohort data for the estimation of  $S_z$ . In this section, based on the combined data, we study the probability structure of the second gap time and explore the conditions under

which the gap time data remain a representative sample from the target population. For simplicity of discussion, assume that the incident cohort recruits individuals who experience the initiating event within a calendar time interval  $[0, C]$ , where  $C > 0$  is a constant calendar time. The prevalent cohort is a sample selected from the prevalent population defined at calendar time 0. We first consider the case that the observation of multiple events from the incident and prevalent cohorts ends at calendar time  $C$ , and thus the constant  $C$  serves as the censoring time for the combined data. This simple censoring mechanism can be replaced by random censoring and we do so later in this section. The combined cohort includes the following four types of study individuals:

- (i) those prevalent individuals who have experienced the initiating event before calendar time 0, and the first event is observed in the calendar time interval  $[0, C]$ ;
- (ii) those prevalent individuals who have experienced the initiating event before calendar time 0, but the first event occurs after the calendar time  $C$ ;
- (iii) those incident individuals who experience both the initiating and first events in  $[0, C]$ ;
- (iv) those incident individuals who experience the initiating event during the time interval  $[0, C]$ , but the first event occurs after the calendar time  $C$ .

With  $Z$  as the focused variable, the sub-cohorts (i) and (iii) together form the cohort for observing the second gap times, termed the  $Z$ -cohort. Clearly, the population for the  $Z$ -cohort is a sub-population of the combined incident and prevalent population, termed the  $Z$ -population. The  $Z$ -population is essentially the population of those who develop the first event in the calendar interval  $[0, C]$ . In the  $Z$ -population, let  $W = X + Y$  denote the calendar time when the first event occurs. Let  $\{\min(Z, C - W), I(Z \leq C - W) : C \geq W\}$  be the observed second gap times and the corresponding censoring indicators, considered as the survival data from the  $Z$ -cohort for estimating  $S_z$ . A question of interest now is whether it is appropriate to apply standard methods to these survival data. In general, the appropriateness depends not only on the stationarity condition (S1) but also on (S2), as discussed below.

The fundamental requirement for the validity of the usual survival analysis is the independence between  $Z$  and  $C - W$ . This requirement is essential for both one-sample models and, conditional on covariates, testing and regression models. Given that  $C$  is a constant, independent censoring is equivalent to the independence of  $Z$  and  $W$ , or that the distribution of  $Z$  is independent of the calendar time of the first event in the  $Z$ -population. When (S1) holds, note that the density of  $Z$  conditional on  $W = w$  is

$$p(z|w) = \frac{\int_{w-y^*}^w f_{y,z}(w-x, z)\phi_x(x)dx}{\int \int_{w-y^*}^w f_{y,z}(w-x, v)\phi_x(x)dx dv}$$

$$= f_z(z) \left[ \frac{\int_{w-y^*}^w f_{y|z}(w-x|z)\phi_x(x)dx}{\int \int_{w-y^*}^w f_{y,z}(w-x,v)\phi_x(x)dx dv} \right]. \quad (3.1)$$

Also note that, for each  $z$ ,

$$\int_{w-y^*}^w f_{y|z}(w-x|z)dx = 1$$

and

$$\int \int_{w-y^*}^w f_{y,z}(w-x,v)dx dv = 1.$$

The bracketed term in (3.1) equals 1 when  $\phi_x(x)$  is constant for  $x \geq -y^*$ . Thus, when both (S1) and (S2) hold, the density  $p(z|w)$  in the Z-cohort is independent of  $w$  and equals the population density  $f_z(z)$  in the Z-population. It is important to note that the equality  $p(z|w) = f_z(z)$  does not generally hold if only (S1) is valid.

Next we study the case that the censoring time  $C$  is random. Assume that  $C$  is independent of  $(W, Z)$ ; that is, the censoring is independent of when the first event occurs and the second gap time. In the preceding discussion, when  $C$  is a positive constant, it is seen that  $Z$  is independent of  $W$ . When both (S1) and (S2) hold, note that the Z-cohort is a random sample from the Z-population conditional on  $C \geq W$ . Let  $p_c(z|w)$  be the density of  $Z$  conditional on  $W = w$  and  $C \geq W$ ; then clearly,  $p_c(z|w) = p(z|w)$  because  $C$  is independent of  $(W, Z)$ . Thus, when both (S1) and (S2) hold, the density  $p_c(z|w)$  in the Z-cohort equals  $f_z(z)$  and is independent of  $w$ . Further, the gap time  $Z$  is independent of the induced censoring time  $C - W$  because  $Z$  is independent of both  $W$  and  $C$  in the Z-cohort. Let  $z_i^* = \min(z_i, c_i - w_i)$  and  $\delta_i^* = I(z_i \leq c_i - w_i)$ . The survival data  $(z_i^*, \delta_i^*)$  from the Z-cohort, which are observable only for those satisfying  $c_i - w_i \geq 0$ , can be treated as the usual right-censored data for inferences of  $Z$ . The estimation or testing results based on such survival data can thus be equivalently derived from the class of observation  $\{(z_i^*, \delta_i^*) : c_i - w_i \geq 0\}$ .

An interesting question raised by a reviewer concerns the directions of incident and prevalent data bias in the probability structure of  $Z$ . As an example, suppose  $Y$  and  $Z$  are positively correlated. In an incident cohort the larger values of  $Z$  are likely to be censored because the corresponding censoring times  $V - Y$  tend to be small, while in a prevalent cohort the larger values of  $Z$  are likely to be included. Essentially, in the combined cohort, the different directions of bias are cancelled out under assumptions (S1) and (S2). The assumption (S2) does play an important role in the balance of two kinds of bias: for the subjects whose first event occurs in the calendar time interval  $[0, C]$ , the failure time  $Y$  is subject to left and right-truncation (i.e., double truncation). The left and right-truncation cannot be cancelled out when (S2) is violated but can be cancelled out when



both (S1) and (S2) are satisfied. This phenomenon was studied by Wang and Lu (1998). The balance of incident and prevalent data bias for the second gap time,  $Z$ , is achieved when study subjects in the combined cohort are sampled with both (S1) and (S2) satisfied.

#### 4. An Example: The MACS Study

In this section we present an example to illustrate the problems arising in the analysis of gap time data.

The Multicenter AIDS Cohort Study (MACS) is a cohort study of HIV-1 infection and AIDS among homosexual men. The study recruited 4,954 homosexual men without prior diagnosis of AIDS during an approximate one-year period from 1984 to 1985. Among the homosexual men, there were 1,745 HIV-1-infected individuals; the rest were not infected with HIV-1 virus. The data provide a source of laboratory measurements, AIDS medication, sociodemographic, behavioral and psychological information (Kaslow, Ostrow, Detels, Phair, Polk and Rinaldo (1987)).

Suppose the outcomes of interest are the chronologically-ordered events (HIV-1 infection, AIDS, death). Those homosexual men who were identified as infected with HIV-1 at entry are classified as the prevalent cohort. The incident cohort is formed by those who entered the study with sero-negative tests but became HIV-1-infected before the end of follow-up. In the literature, a considerable amount of statistical work has been conducted surrounding problems related to the analysis of the time from HIV-1 infection to the diagnosis of AIDS from the prevalent cohort. The observed data are recognized as survival data subject to left truncation and right censoring, where the left truncation time may not be observable. With the presence of left truncation in the data, the analysis of the time from AIDS to death — a gap time — is apparently non-trivial, as discussed earlier in this paper. The rate of the incidences of HIV-1 infections before the recruitment (1984-1985) is generally believed to be increasing over time (Brookmeyer and Gail (1994)). The incidence rate of HIV-1 infections after the recruitment (1984-1985) may not follow the same pattern of the population incidence rate because of the (potential) effect of prevention in a cohort study.

The density function corresponding to the incidence rate can be estimated using the observed right censored data. In Figure 3, a plot of the discrete version of the density function for the time to HIV-1 infections is presented. Given that the recruitment period is short, this density plot reveals the pattern of the incidence rate of HIV-1 infections, for the incident cohort, since 1984: the rate was low at entry to study, became high in a year, and gradually decreased over time. A possible explanation for such a pattern is that the cohort individuals were cautious at the beginning of the study, but reverted to their previous social

and sexual behaviors within a year after entering the study, and gradually learned or accepted methods to prevent infection through programs offered in the study.

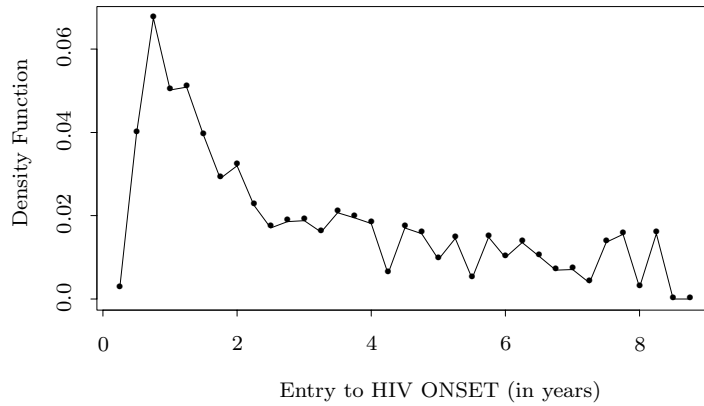


Figure 3.

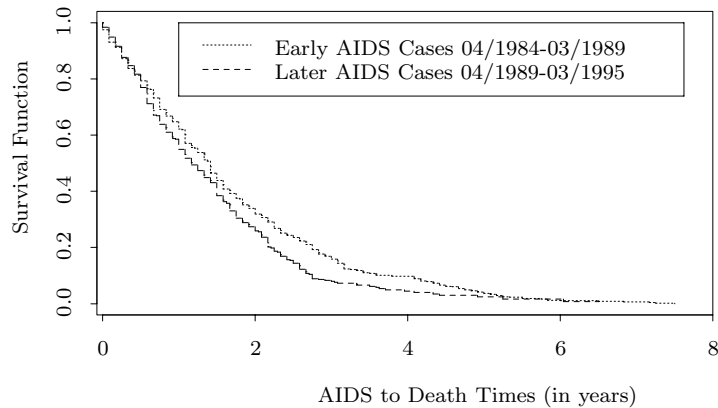


Figure 4.

Clearly the stationarity condition (S2) is seriously violated with the AIDS disease; thus the application of the usual methods in survival analysis might lead to biased results. In Figure 4, the Kaplan-Meier estimates for early AIDS cases (April 1, 1984 – March 31, 1989) and for later cases (April 1, 1989 – March 31, 1995) are calculated for the estimation of the survival function for the time from AIDS to death. It is seen, surprisingly and unexpectedly, that the early AIDS cases have a better chance for longer survival. The overall log relative hazard of the late v.s. early AIDS-onset indicator in the proportional hazards model  $\lambda_0(z) \exp\{x\beta\}$  (Cox (1972)) is  $\hat{\beta} = 0.6180$  (s.e.= 0.0805). To

adjust for possible confounding factors, a few covariates enter the proportional hazards model for further regression analysis. The covariates used in the model include the late v.s. early AIDS-onset indicator ( $x_1$ ), baseline age at the time of AIDS-onset ( $x_2$ ), and baseline CD4 count at the time of AIDS-onset ( $x_3$ ). When the model is adjusted for age ( $x_2$ ), the partial likelihood estimates report  $(\hat{\beta}_1, \hat{\beta}_2) = (0.5946, 0.0126)$  (s.e. = (0.0810, 0.0046)); the model adjusted for CD4 count ( $x_3$ ) reports  $(\hat{\beta}_1, \hat{\beta}_3) = (0.5268, -0.0014)$  (s.e. = (0.0817, 0.0002)); the model adjusted for both age ( $x_2$ ) and CD4 count ( $x_3$ ) reports  $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (0.4961, 0.0136, -0.0014)$  (s.e. = (0.0826, 0.0050, 0.0002)). The simple regression results suggest that the direction of hazards for the late v.s. early AIDS-onset indicator ( $x_1$ ) is not changed by adjusting for baseline age and CD4 count. To conduct a more satisfactory regression analysis, other potentially interesting covariates such as the incubation time from HIV infection to AIDS, or various treatment plans, could enter the model for adjustment. Nevertheless, these covariates may or may not be available in a study and, even if they enter the model, it is still questionable if the corresponding regression analysis can adjust the gap time bias.

In this section we emphasize that the analytical results presented could be misleading because of many complicated factors. One important factor might be that the early and late AIDS cohorts are in fact not comparable because the distributions of the time from AIDS to death defined at different calendar times are affected by the distribution of HIV-1 infection, as well as the distribution of the time from HIV-1 infection to AIDS. Statistical methods have thus far not been developed for analyzing the described data. The future development of such methods will be important for AIDS-related studies as well as for other multiple event problems.

## 5. Concluding Remarks

In this paper the gap time bias arising from the incident and prevalent cohorts is studied. For the incident cohort, the population probability structure of gap times is unbiased but the second gap time is observed subject to induced informative censoring. In contrast with the bias in the incident data, the gap time bias is present in the prevalent population since the cohort tends to recruit individuals with longer first gap time. The combined cohort data can be treated as standard right-censored data only when the stationarity conditions (S1) and (S2) hold. For future research, it will be interesting to explore statistical methods for gap times based solely on the incident data, solely on the prevalent data, or on the combined data with less restrictive model assumptions. Research in these directions remains relatively unreported in the literature. Related statistical methods will be important for the analysis of multiple event data where the gap times serve as the outcome variables of interest.

## Acknowledgement

This research was supported by the National Institute of Health grants R01 DA10184 and R01 MH56639.

## References

- Brookmeyer, R. and Gail, M. H. (1987). Biases in prevalent cohorts. *Biometrics* **43**, 739-749.
- Brookmeyer, R. and Gail, M. H. (1994). *AIDS Epidemiology: A Quantitative Approach*. Oxford University Press, Oxford, England.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc. B* **34**, 187-220.
- Gelber, R. D., Gelman, R. S. and Goldhirsch, A. (1989). A quality-of-life oriented endpoint for comparing therapies. *Biometrics* **45**, 781-795.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.
- Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F. and Rinaldo, Jr. C. R. (1987). The multicenter AIDS cohort study: rationale, organization and selected characteristics of patients. *Amer. J. Epidemiology* **126**, 310-318.
- Lin, D. Y., Sun, W. and Ying, Z. (1997). Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika* **86**, 59-70.
- Visser, M. (1996). Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS. *Biometrika* **71**, 507-518.
- Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Amer. Statist. Assoc.* **86**, 130-143.
- Wang, M.-C., Brookmeyer, R. and Jewell, N. P. (1993). Statistical models for prevalent data. *Biometrics* **49**, 1-11.
- Wang, M.-C. and Lu, S.-E. (1998). Analysis of doubly-truncated data with case-cohort designs. Technical Report, Department of Biostatistics, Johns Hopkins University.
- Wang, W.-J. and Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* **85**, 561-572.

Department of Biostatistics, School of Hygiene and Public Health, John Hopkins University, 615 North Wolfe Street, Baltimore MD 21205, U.S.A.

E-mail: mcwang@jhsph.edu

(Received September 1997; accepted May 1999)