

# Rank Estimation of Treatment Differences Based on Repeated Measurements Subject to Dependent Censoring

*D.V. Glidden and L.J. Wei*

In comparing the effectiveness of two treatments, suppose that for each patient repeated measurements of an outcome variable are taken at prespecified time points. Some observations, however, may be missing due to the patient's dependent right censoring. In this article, a simple rank estimation procedure, constructed based on an artificial censoring technique, is proposed for the treatment differences over time without imposing a parametric structure on the dependence between the outcome measures and the censoring variable. Our method can be easily implemented and is illustrated by a data set from an AIDS clinical trial.

**KEY WORDS:** Cumulative hazard function; Estimating function; Informative censoring; Longitudinal data; Resampling method; Wilcoxon statistic

### **Authors' Footnote**

D.V. Glidden is assistant adjunct professor, Department of Epidemiology and Biostatistics, University of California at San Francisco, 500 Parnassus Avenue, San Francisco, CA 94143. L.J. Wei is professor, Department of Biostatistics, Harvard University, 677 Huntington Ave., Boston, MA 02115. This work was partially supported by Grants AI07358, AI38855 and CA56844 from the National Institutes of Health. The authors are grateful to two referees, the Editor, the Associate Editor, Michael Hughes and James Robins for insightful comments on the paper.

## 1. INTRODUCTION

In comparing the effectiveness of two treatments, suppose that repeated measurements of an outcome variable for each patient are obtained at prespecified time points during the course of the study. One of the primary goals for collecting such repeated measurements is to study the treatment effect over time based on this particular variable. The analysis of this type of data, however, is complicated by the presence of missing observations. When missing values are missing completely at random, non-parametric and semi-parametric methods have been developed to analyze such data (Wei and Lachin 1984; Wei and Johnson 1985; Liang and Zeger 1986). If missing observations are missing at random (Rubin 1976), likelihood based inference procedures and semi-parametric methods have been studied, for example, by Laird and Ware (1982), Little and Rubin (1987), and Robins, Rotnitzky and Zhao (1995). If the patient's follow-up time depends on the underlying slope of the repeated measurements of the outcome variable, various parametric models have been utilized to model such informative censoring in the analysis (Wu and Carroll 1988; Wu and Bailey, 1988, 1989; Mori, Woodworth and Woolson 1992; DeGruttola and Tu 1994). Recently, Dawson and Lagakos (1993) have examined the properties of a simple method which summarizes each subject's repeated measures with a single statistic, and then performs distribution-free comparisons based on stratification according to the pattern of missing data.

In this article we propose a semi-parametric estimation procedure for treatment differences over time based on repeated measurements of an outcome variable when the patient's follow-up time may depend on observed and/or missing measurements. For example, in a recent clinical trial to examine the effect of switching to didanosine (ddI) from zidovudine (AZT) for HIV-infected patients who had tolerated AZT for at least 16 weeks, 304 patients were randomly chosen to continue AZT therapy and 298 patients were assigned to take ddI in a daily dose of 500 mg (Kahn et al., 1992). The patient's CD4 cell count, a proxy for the progression of the HIV infection, was periodically obtained during the trial. In addition to the usual clinical endpoint analy-

sis, it is also interesting and important to evaluate the group difference based on such repeated measurements. Most study patients, however, do not have complete sets of CD4 cell counts due to administrative censoring (from the patient's staggered entry) and withdrawal from the trial initiated by the patient or his/her physician. Table 1 gives the numbers of patients in the two treatment groups, who discontinued participation in the study for various reasons. Except for the administrative censoring, it seems difficult to justify the assumption that the patient's follow-up time is independent of the observed and/or missing CD4 cell counts. For instance, a withdrawal initiated by a patient from the study usually depends on his/her health status and is likely related to the underlying CD4 count. Figure 1 gives the Kaplan-Meier curves based on those possibly dependent follow-up times, which may be censored by administrative censoring, for the AZT and ddI groups. Patients in the AZT group tended to have shorter, response-dependent follow-up times than those from the ddI group. The method proposed in this paper adjusts for such unequal dependent censoring in assessing the treatment difference based on the incomplete repeated observations. Our method can be easily implemented and is illustrated by a data set from the previously discussed AIDS clinical trial.

## 2. ESTIMATING TREATMENT DIFFERENCES

### 2.1 The Model

Suppose that  $\mathbf{X}_i = (X_{1i}, \dots, X_{Ki})'$  ( $i = 1, \dots, n_x$ ) and  $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{Kj})'$  ( $j = 1, \dots, n_y$ ), measured at time points  $\{t_k, k = 1, \dots, K\}$ , are independent samples with completely unspecified (continuous) distribution functions  $F_x$  and  $F_y$  whose marginals are  $F_{xk}$  and  $F_{yk}$ , respectively ( $k = 1, \dots, K$ ). Furthermore, assume that there exists a vector of unknown constants  $\boldsymbol{\theta}_0 = (\theta_{10}, \dots, \theta_{K0})'$  such that  $F_y(\mathbf{s} - \boldsymbol{\theta}_0) = F_x(\mathbf{s})$ , for  $\mathbf{s} \in R^K$ . The location shift parameter  $\theta_{k0}$  can be interpreted as the treatment difference at time  $t_k, k = 1, \dots, K$ . We are interested in making inferences about  $\{\theta_{k0}\}$  based on  $\mathbf{X}_i$  and  $\mathbf{Y}_j$ . Unfortunately, in practice, some vectors  $\mathbf{X}$ 's and  $\mathbf{Y}$ 's may have

missing components due to, for example, censoring. Suppose that for patients with outcome vectors  $\mathbf{X}_i$  and  $\mathbf{Y}_j$ , the usual independent random censoring variables are denoted by  $C_{xi}$  and  $C_{yj}$ , respectively. Furthermore, let  $D_{xi}$  and  $D_{yj}$  be the corresponding censoring variables which may depend on  $\mathbf{X}_i$  and  $\mathbf{Y}_j$ , respectively. The  $\{D_{xi}\}$  and  $\{D_{yj}\}$  are assumed to be random samples from two, possibly distinct, populations whose cumulative hazard functions are  $\Lambda_x(\cdot)$  and  $\Lambda_y(\cdot)$ , respectively. Combining the above two types of censoring, the lengths of observation for  $\mathbf{X}_i$  and  $\mathbf{Y}_j$  are then  $T_{xi}(= \min(D_{xi}, C_{xi}))$  and  $T_{yj}(= \min(D_{yj}, C_{yj}))$ , respectively. Now, let  $\epsilon_{ki} = 1$ , if  $X_{ki}$  is observed, 0, otherwise; and  $\xi_{kj} = 1$ , if  $Y_{kj}$  is observed, 0, otherwise. If missing components of an outcome vector occurred before the censoring time  $T$ , they are assumed to be missing completely at random.

We model the relationship between the outcome vector and the censoring time semiparametrically. The joint distributions of  $(X_{1i}, \dots, X_{Ki}, D_{xi})'$  and  $(Y_{1j}, \dots, Y_{Kj}, D_{yj})'$  are totally unspecified. However, we assume that the joint distribution of  $(X_{1i}, \dots, X_{Ki}, D_{xi})'$  is identical to that of  $(Y_{1j} - \theta_{10}, \dots, Y_{Kj} - \theta_{K0}, g(D_{yj}))'$ , where  $g$  is an unknown function. Note that  $g(\cdot) = \Lambda_x^{-1}\{\Lambda_y(\cdot)\}$ . Although  $g$  is a monotone function, this does not imply that one of the censoring variables is stochastically larger than the other. Intuitively, this model assumes that while treatment may have an effect on the censoring time, the rank association between the outcome vector and the dropout time is not altered by the treatment. For example, in the aforementioned AIDS clinical trial, if a patient in the AZT group with a fast decline of CD4 cell counts was likely to withdraw from the study, the same phenomenon would be observed for a patient with a similar profile of repeated CD4 counts in the ddI group. On the other hand, the model assumption is violated when such dependence differs between the two groups. For example, to evaluate a new treatment in an open placebo-control trial, patients in the placebo arm with worsening clinical conditions withdrew in order to switch to the new treatment, while patients in the treated group, who did poorly, remained on the same arm since it was considered the only option.

A special case of this general model occurs if the logarithm of the dependent censoring time is linearly related to the entire set of repeated measurements of the outcome variable. That is,

$$\log D_{xi} = \gamma_0 + \sum_{k=1}^K \gamma_k X_{ki} + e_{xi}, i = 1, \dots, n_x,$$

and

$$\log D_{yj} = \gamma_0 + \sum_{k=1}^K \gamma_k Y_{kj} + e_{yj}, j = 1, \dots, n_y,$$

where  $\gamma$ 's are some constants and  $e$ 's are independent and identically distributed random variables. For this case,  $g(D_{yj}) = D_{yj} \exp(-\sum_k \gamma_k \theta_{k0})$ . Some versions of this model have been postulated for the joint distribution between repeated measurements and the censoring variable (DeGruttola and Tu 1994; Wulfsohn and Tsiatis 1997).

## 2.2 Point Estimation

If the censoring variables  $D_x$  and  $D_y$  are independent of the outcome variables, a rank estimator for  $\theta_{k0}$  based on the Wilcoxon test statistic can be easily obtained by solving the following estimating equations:

$$\sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \epsilon_{ki} \xi_{kj} \{I(Y_{kj} - \theta_k \geq X_{ki}) - 1/2\} = 0, \quad (1)$$

where  $k = 1, \dots, K$ , and  $I(\cdot)$  is the indicator function (Lehmann, 1975, pp. 91-94). If  $g(\cdot)$  is the identity function, that is, the distributions of the dependent censoring variables are identical, the solution to (1) is still consistent. This is no longer true, however, when the distributions of  $D_x$  and  $D_y$  are different. Here, we introduce an *artificial censoring* technique to make these two dependent censoring distributions equal. For illustration, let us consider a simple case in which  $D_y$  is stochastically larger than  $D_x$ . This implies that  $g(t) \leq t$ ,  $t \in R^+$ . We then replace the follow-up time  $T_{yj}$  with a shorter time  $g(T_{yj})$ ,  $j = 1, \dots, n_y$ . Now, if  $Y_{kj}$  is observed, but the new follow-up time  $g(T_{yj}) < t_k$ , then  $Y_{kj}$  is excluded from (1). In general, if  $g(T_y) < T_y$ , we replace  $T_y$  by

$g(T_y)$ . On the other hand, if  $g^{-1}(T_x) < T_x$ , we replace  $T_x$  with  $g^{-1}(T_x)$ . When  $g$  is known, this results in the following set of unbiased estimating functions for  $\{\theta_{k0}\}$ :

$$n^{-3/2} \sum_i \sum_j \epsilon_{ki} \xi_{kj} I(g^{-1}(T_{xi}) \geq t_k) I(g(T_{yj}) \geq t_k) \{I(Y_{kj} - \theta_k \geq X_{ki}) - 1/2\},$$

which is equivalent to  $\{S_{1k}(\theta_k, g_k, h_k)\}$ :

$$n^{-3/2} \sum_i \sum_j \epsilon_{ki} \xi_{kj} I(T_{xi} \geq g_k) I(T_{yj} \geq h_k) \{I(Y_{kj} - \theta_k \geq X_{ki}) - 1/2\}, \quad (2)$$

where  $g_k = g(t_k)$ ,  $h_k = g^{-1}(t_k)$ ,  $k = 1, \dots, K$ , and  $n = n_x + n_y$ .

Note that the estimating functions (2) depend on the nuisance function  $g(\cdot)$  only through a finite number of parameters  $\{g_k, h_k, k = 1, \dots, K\}$ . Estimators of these parameters can be easily obtained using the following estimating functions:

$$S_{2k}(g_k) = n^{1/2}(\hat{\Lambda}_y(t_k) - \hat{\Lambda}_x(g_k)), \quad (3)$$

and

$$S_{3k}(h_k) = n^{1/2}(\hat{\Lambda}_y(h_k) - \hat{\Lambda}_x(t_k)), \quad (4)$$

where  $k = 1, \dots, K$ , and  $\hat{\Lambda}_y$  and  $\hat{\Lambda}_x$  are Nelson estimators for  $\Lambda_y$  and  $\Lambda_x$ , respectively. Note that  $S_{2k}(\cdot)$  and  $S_{3k}(\cdot)$  are monotone functions and for fixed  $g_k$  and  $h_k$ ,  $S_{1k}$  is a monotone function in  $\theta_k$ . It follows that the roots  $\{\hat{\theta}_k, \hat{g}_k, \hat{h}_k; k = 1, \dots, K\}$  to the equations:  $S_{1k}(\theta_k, g_k, h_k) = 0, S_{2k}(g_k) = 0, S_{3k}(h_k) = 0, k = 1, \dots, K$ , are consistent and can be easily obtained numerically.

### 2.3 Interval Estimation

In the Appendix, we show that for large  $n_x$  and  $n_y$ , the joint distribution of  $\{S_{1k}(\theta_{k0}, g_{k0}, h_{k0}), S_{2k}(g_{k0}), S_{3k}(h_{k0}), k = 1, \dots, K\}$  can be approximated by a normal distribution with mean 0 and covariance matrix  $\hat{\Sigma}$ , where  $g_{k0}$  and  $h_{k0}$  are the true values of  $g_k$  and  $h_k$ , respectively. Using this large sample result and the arguments given in the appendices of Wei, Ying and Lin (1990), one can easily show that the above estimating functions have a linearity property in a neighborhood of  $(\theta_{k0}, g_{k0}, h_{k0}, k = 1, \dots, K)$

and that the joint distribution of  $\{\hat{\theta}_k, \hat{g}_k, \hat{h}_k, k = 1, \dots, K\}$  is approximately normal. The limiting covariance matrix, however, involves the unknown density functions of  $F_{xk}, F_{yk}, \Lambda_x$  and  $\Lambda_y$ , which are difficult to estimate directly with censored data. Here, we use a simple resampling method proposed by Parzen, Wei and Ying (1994) to make inferences about  $\theta_{k0}$ 's. Let  $\mathbf{Z} = (Z_{1k}, Z_{2k}, Z_{3k}; k = 1, \dots, K)$  be a normal vector with mean 0 and covariance matrix  $\hat{\Sigma}$  and let  $\{s_{lk}; l = 1, 2, 3; k = 1, \dots, K\}$  be the observed  $\{S_{lk}; l = 1, 2, 3; k = 1, \dots, K\}$ . Furthermore, let the random variables  $\{\theta_k^*, g_k^*, h_k^*, k = 1, \dots, K\}$  be the solutions to the equations:

$$s_{1k}(\theta_k, g_k, h_k) = Z_{1k}, \quad s_{2k}(g_k) = Z_{2k}, \quad s_{3k}(h_k) = Z_{3k}, \quad k = 1, \dots, K. \quad (5)$$

Then, the joint distribution of  $\{(\hat{\theta}_k - \theta_{k0}), (\hat{g}_k - g_{k0}), (\hat{h}_k - h_{k0}), k = 1, \dots, K\}$  can be approximated by that of  $\{(\theta_k^* - \tilde{\theta}_k), (g_k^* - \tilde{g}_k), (h_k^* - \tilde{h}_k), k = 1, \dots, K\}$ , where  $\tilde{v}$  is the observed value of the estimator  $\hat{v}$  (Parzen, Wei and Ying, 1994). In practice, to estimate the covariance matrix,  $\Gamma$ , for  $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_K)'$ , we generate a large random sample  $\{\mathbf{z}_m, m = 1, \dots, M\}$  from a normal population with mean 0 and covariance matrix  $\hat{\Sigma}$ . For a realization  $\mathbf{z}_m$ , we solve the equations (5) with  $\mathbf{Z}$  being  $\mathbf{z}_m$ . Let the resulting solutions be denoted by  $\{\theta_{km}^*, g_{km}^*, h_{km}^*, k = 1, \dots, K\}$ . Then,  $\Gamma$  can be approximated by  $\hat{\Gamma} = M^{-1} \sum_{m=1}^M (\boldsymbol{\theta}^*_m - \tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}^*_m - \tilde{\boldsymbol{\theta}})'$ , where  $\boldsymbol{\theta}^*_m = (\theta_{1m}^*, \dots, \theta_{km}^*)'$  and  $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \dots, \tilde{\theta}_K)'$ .

Inferences about  $\theta_{k0}$ 's can be obtained through the above large sample properties of  $\hat{\boldsymbol{\theta}}$ . For example, a  $(1 - \alpha)$  simultaneous confidence band  $B$  for  $\boldsymbol{\theta}_0$  may be constructed as follows: let  $u(\alpha)$  be a cut-off point such that  $\text{pr}(\max_{1 \leq k \leq K} \{\hat{\gamma}_{kk}^{-1/2} |\hat{\theta}_k - \theta_0|\} \leq u(\alpha)) = 1 - \alpha$ , where  $\hat{\gamma}_{kk}$  is the  $k$ th diagonal element of  $\hat{\Gamma}$  and  $0 < \alpha < 1$ . Then

$$B = \{ \boldsymbol{\theta} = (\theta_1, \dots, \theta_K)' : \hat{\theta}_k - u(\alpha)\hat{\gamma}_{kk}^{1/2} \leq \theta_k \leq \hat{\theta}_k + u(\alpha)\hat{\gamma}_{kk}^{1/2} \}. \quad (6)$$

The cut-off point  $u(\alpha)$  can be easily obtained by simulating a large number of normal vectors with mean 0 and covariance matrix  $\hat{\Gamma}$ .

If the treatment differences  $\theta_{k0}$  are roughly equal over the study period, that is,  $\theta_{k0} = \eta, k = 1, \dots, K$ , for some unknown constant  $\eta$ , one may use a linear combination

of  $\hat{\theta}_k$ 's:  $\hat{\eta} = \sum_k w_k \hat{\theta}_k$ , to estimate  $\eta$ , where  $\sum_k w_k = 1$ . An optimal choice for  $\mathbf{w} = (w_1, \dots, w_K)'$  is  $\hat{\Gamma}^{-1} \mathbf{e} / \{ \mathbf{e}' \hat{\Gamma}^{-1} \mathbf{e} \}$ , where  $\mathbf{e}$  is  $K \times 1$  vector of ones ( Wei and Johnson 1985). Confidence intervals of  $\eta$  can be obtained accordingly based on  $\hat{\eta}$  and its large sample distribution.

### 3. AN EXAMPLE

We use the data from the AIDS trial discussed in the Introduction to illustrate our proposal. Here,  $\mathbf{X}_i$  is the vector of CD4 count measurements for the  $i$ th patient from the AZT group taken at weeks 8, 16 and 24 after randomization,  $i = 1, \dots, 304$ , and  $\mathbf{Y}_j$  is the corresponding vector for the  $j$ th patient in the ddI group,  $j = 1, \dots, 298$ . Table 2 displays the patterns of missing observations for patients in the two groups. We assume that intermittent missing observations during the patient's follow-up are missing completely at random, and except for the administrative censoring, all other types of censoring are potentially dependent on the CD4 counts. Empirically, through the standard Cox regression analysis, we find that patients with a steep decline on the CD4 count tend to drop out of the study earlier than those with a gradual deterioration over time.

Using estimating functions (2)-(4), a point estimate  $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)'$  for  $\boldsymbol{\theta}_0$  is  $(18, 16, 10)'$  with an estimated covariance matrix

$$\hat{\Gamma} = \begin{pmatrix} 63 & 8 & 8 \\ 8 & 45 & 9 \\ 8 & 9 & 48 \end{pmatrix}$$

which is obtained based on the resampling method proposed by Parzen et al. (1994) with  $M = 500$ . The corresponding time-specific 0.95 confidence intervals are  $(6, 37)$ ,  $(3, 30)$  and  $(-1, 25)$ , respectively. A 0.95 simultaneous confidence band  $B$  in (6) for  $\boldsymbol{\theta}_0$  is  $\{(-1, 37) \times (0, 32) \times (-6, 27)\}$ . If we assume that  $\theta_{10} = \theta_{20} = \theta_{30} = \eta$ , with the optimal weights described in Section 2,  $\hat{\eta} = 14$  and the corresponding estimated standard error is 5.

Although patients in the ddI group tended to have longer follow-up times than those in the AZT group, we don't always artificially delete observations in the ddI group. In

fact, our method artificially censors 2 in the AZT group at week 8, but 17 patients at week 16 and 38 patients at week 24 in the ddI group. If the patient's follow-up time is independent of the observed and the missing CD4 cell counts, one may estimate  $\theta_0$  using (1) with all the observed data points in the analysis. The resulting point estimates for  $\theta_0$  are 18, 11 and 6. The corresponding estimated standard errors are 8, 6 and 6, respectively. Note that if patients with lower CD4 counts tend to withdraw from the study earlier than those with higher counts, the *naive* method using (1) may underestimate the effect of the ddI on the CD4 count. In the next section, we show through numerical studies that the above naive procedure may lead to seriously biased inferences.

Now, one may argue that censoring due to the major toxicity is independent of the underlying outcome variable. For this case, the point estimate for  $\theta_0$  is (18, 17, 12)' with estimated standard errors of 8, 7 and 7. These results are very similar to those for the case in which the only independent censoring is due to the patient's staggered entry.

Wu and Carroll (1988) and Wu and Bailey (1989) considered a growth curve model for the repeated measurements and parametric models for the probabilities of the patient's dropout at the observation times. For example, at a specific observation time point, one may assume a logistic regression model for the dropout with the slope of the repeated measurements as the covariate. In AIDS clinical trials, most marker measurements, for example CD4 counts, tend to go up first and then decline. Such a phenomenon may not be well described by a parametric growth curve model. For our AIDS example, using the linear minimum variance unbiased estimation procedure developed by Wu and Bailey (1989), the estimated average difference of the slopes for the repeated CD4 counts between the AZT and ddI groups is 0.86 with an estimated standard error of 0.41.

#### 4. NUMERICAL STUDIES

Simulation studies were conducted to evaluate the performance of the proposed

procedure. In one of the numerical studies, we considered a trial with 150 patients randomly allocated to two treatments with equal probability. For each patient, four repeated measurements at 0, 20, 30, and 40 weeks and a dependent censoring time were generated. The outcome vectors followed a growth curve model:

$$X_{ki} = \alpha_i + \beta_{ix}t_k + e_{ik}, \quad Y_{kj} = \alpha_j + \beta_{jy}t_k + e_{jk} \quad k = 1, \dots, K,$$

where  $(\alpha_i, \beta_{ix}, e_{ik})'$  and  $(\alpha_j, \beta_{jy}, e_{jk})'$ ,  $i = 1, \dots, n_x$  and  $j = 1, \dots, n_y$ , were independent normal random variables with mean  $(400, 0, 0)'$  and  $(400, 1.5, 0)'$  and variance  $(81, 33, 100)'$ . The dependent censoring times  $D_{xi}$  and  $D_{yj}$  were exponential random variables with means  $100e^{0.5\beta_i}$  and  $120e^{0.5\beta_j}$ , respectively. In this study, there were no intermittent missing observations and the independent censoring variable  $C$  was set to  $\infty$ . The group differences were  $(\theta_{10}, \theta_{20}, \theta_{30}) = (30, 45, 60)$  and patients in the first group were more likely to be censored earlier than patients in the second group. For example, approximately 35%, 40% and 45% of patients in the first group had missing values at Weeks 20, 30, and 40. The corresponding missing proportions for patients in the second group were only 23%, 27%, and 30%, respectively.

This empirical study was designed to compare the procedure for the conditional linear model investigated by Wu and Bailey (1989) with the new method when the mean structure of the response variable follows a growth curve model, but a significant number of subjects do not have post-baseline measurements. For each of the 500 simulated samples, point estimates and 0.95 confidence intervals for  $\theta_{10}$ ,  $\theta_{20}$  and  $\theta_{30}$  were obtained based on the 'naive' method (treating censoring as non-informative), the conditional linear model approach, and the new procedure with  $M = 250$ . The results are summarized in Table 3. The naive method and Wu and Bailey's procedure give extremely biased estimators and the empirical coverage probability of the confidence intervals for both methods are smaller than the corresponding nominal confidence coefficient. On the other hand, our method performs quite well.

The second simulation was conducted to compare the performance of the above methods when the mean structure of the data does not closely follow a growth curve

model. We again considered a trial with 150 patients randomly allocated to two treatments with equal probability. For each patient, a baseline measurement, two repeated measurements and a dependent censoring time were generated: one at Week 16 and the other at Week 48. The vectors:  $(X_{0i}, X_{1i}, X_{2i}, \log D_{xi})$  and  $(Y_{0j}, Y_{1j}, Y_{2j}, \log D_{yj})$ , were simulated from normals with mean vectors  $(380, 400, 390, 4.4)'$  and  $(380, 425, 440, 7.4)'$ , respectively, and with covariance matrices given in Table 4. There were no intermittent missing observations and the independent censoring variable  $C$  is set to  $\infty$ . The group differences were  $(\theta_{10}, \theta_{20}) = (25, 50)$  and the proportions of missing observations in the two groups were 13% and 35% ( $x$  group) and 11% and 17% ( $y$  group), respectively. Table 5 gives average point estimates and the empirical coverage probabilities of 0.95 confidence intervals for  $\theta_{10}$  and  $\theta_{20}$  based on 500 simulated samples. Both the naive method and Wu and Bailey's estimation procedure show strong evidence of bias while the proposed method performs well.

We find that the proposed interval procedure has the correct coverage probability even for the case when  $K$ , the number of observation times, is large. For example, with 150 patients randomly allocated to the two groups and observations taken at Week 16, 48, 60, 70 and 80, we let  $(\mathbf{X}, \log D_x)$  and  $(\mathbf{Y}, \log D_y)$  be generated from normal random vectors with mean  $(400, 390, 380, 370, 360, 2.77)'$  and  $(425, 440, 440, 410, 390, 4.8)'$ , respectively, and with covariance matrices given in Table 6. In this setting, on average, the proportions of missing observations for  $X$  and  $Y$  at these five time points are  $(0.55, 0.6, 0.62, 0.63, 0.65)'$  and  $(0.2, 0.35, 0.38, 0.41, 0.43)'$ , respectively. Based on 500 simulated samples, the empirical coverage probabilities of 0.95 intervals for  $\theta_{k0}$  range from 0.95 to 0.96,  $k = 1, \dots, 5$ . Other simulation studies have demonstrated that the new method has negligible bias and provides reliable confidence intervals for sample sizes as low as 15 to 20 subjects per group.

Our procedure also performs well when the proportions of missing observations are extremely different between the two treatment groups. For example, under a similar simulation setup, we generated data  $(\mathbf{X}, \log D_x)$  and  $(\mathbf{Y}, \log D_y)$  from normals with

means  $(400, 390, 8.5)'$  and  $(425, 440, 0.96)'$  and with covariance matrices:

$$\begin{pmatrix} 25600 & 15300 & 255 \\ 15300 & 25600 & 255 \\ 255 & 255 & 4.5 \end{pmatrix}$$

and

$$\begin{pmatrix} 15600 & 15300 & 223 \\ 15300 & 25600 & 223 \\ 223 & 223 & 3.5 \end{pmatrix}.$$

On average, the proportions of missing observations for  $X$  and  $Y$  are  $(0.1, 0.15)$  and  $(0.7, 0.8)$ , respectively. Based on 500 simulated samples, the empirical coverage probabilities for our interval procedure are nearly identical to the nominal levels.

Since the new method artificially deletes some observations in the analysis, it is interesting to know the “cost” of applying it when censoring is, in fact, non-informative. To this end, we simulated data from the distributions of  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $D_x$  and  $D_y$  used in the first simulation, but with the outcome variables being independent of the censoring time. For the naive method, the average of the 1,000 estimates  $\hat{\boldsymbol{\theta}}$  is  $(30, 45, 61)'$  with empirical root mean squared errors 28, 44, and 60. The mean of the estimates from the conditional linear model is  $(30, 45, 60)'$  with empirical root mean squared errors 28, 42, and 57. For our method, the average of the 1,000 estimates  $\hat{\boldsymbol{\theta}}$  is  $(31, 46, 61)'$  with empirical root mean squared errors 30, 46, and 64.

The model is flexible with respect to the joint distribution of  $(X_1, \dots, X_K, D_x)$ . However, the model makes the fundamental assumption that the rank correlation between repeated measurements is unaltered by treatment. We examined the robustness of the new proposal to this assumption by generating data with a similar design to the second simulation:  $K = 2$ ,  $\boldsymbol{\theta}_0 = (25, 50)$ . Measurements for the  $x$ -group were generated as normal with the covariance matrix given in Table 4. Measurements for the  $y$ -group were also normal with a covariance matrix which was similar to the one in Table 4 but the element which governed the correlation  $\rho$  between  $(Y_1, Y_2)$  and  $D_y$  was allowed to vary from 0.75 to -0.5. The results based on the naive method and the new proposal are displayed in Table 7. When the correlations in the two treatment groups are not

too drastically different (for example, 0.75 vs. 0.5), our method performs reasonably well. Otherwise, the new estimator is biased as is its naive counterpart.

## 5. REMARKS AND CONCLUSIONS

In this article, we propose a simple method to analyze incomplete repeated measures in the presence of dependent censoring. Although there is no specific parametric modeling for the dependence between the outcome and censoring variables, we assume that for the  $i$ th patient in the AZT group, the joint distribution of  $(X_{1i}, \dots, X_{Ki}, D_{xi})'$  is identical to that of  $((Y_{1j} - \theta_{10}), \dots, (Y_{Kj} - \theta_{K0}), g(D_{yj}))'$  from the  $j$ th patient in the ddI group, where  $g(t) = \Lambda_x^{-1}\{\Lambda_y(t)\}$ . If the missing observations are not missing at random, this model assumption is not testable.

Based on the results from the empirical study summarized in Table 7, we find that if the true model moderately deviates from the assumed one, our method performs reasonably well. However, our proposal may perform poorly when the treatment considerably alters the correlation between the outcome and censoring variables. A referee kindly provided the following interesting example to illustrate such a scenario. Suppose that a diuretic is evaluated for treating liver disease in a two-arm controlled clinical study. The diuretic ameliorates edema and ascites, both of which are strong predictors of disease severity. It does not, however, affect the underlying liver disease. If a patient in the treated group is doing well with respect to edema, it is likely that he/she would stop taking the diuretic. However, a patient in the untreated group with worsening clinical condition would be likely to drop out of the study in order to receive the other treatment. Thus, the correlation between edema induced dropout and the outcome measure can be quite different in the two groups. For this case, our method may not be better than the naive procedure.

It is important to note that if the mechanism of generating missing observations is not ignorable, all the existing methods in the literature for analyzing such incomplete data are valid only under a set of non-testable assumptions. Therefore, it is essential that statisticians make every effort to understand the underlying process causing the

missing or censored observations and then choose appropriate methods for the analysis. If the cause of missing data is unknown, one may try several different methods for handling missing data and perform a sensitivity analysis.

If the treatment effect on the repeated measurements is of primary importance, a clinical trial should attempt to obtain measurements even after subjects have discontinued the study treatment. In this situation, continuing to obtain measurements allows for analysis by the intent-to-treat principal. Comparing the results from the intent-to-treat analysis with those obtained from the ‘explanatory’ analysis presented in this article can be quite informative (Fisher, et al., 1990; Oakes, et al., 1993).

It would be interesting and important to extend the proposed technique to the general regression problem. When the covariate vector is discrete, this extension is straightforward. For the case with continuous covariates, the generalization does not seem obvious without further modeling assumptions on the informative censoring variable and the covariates.

Recently, Lin, Robins and Wei (1996) have taken a similar approach to tackle the two-sample problem in survival analysis with dependent censoring. In an unpublished report, Robins (1994) has discussed some general issues of the artificial censoring techniques.

## APPENDIX: ASYMPTOTIC DISTRIBUTION OF $(S_{1k}, S_{2k}, S_{3k}, k = 1, \dots, K)$

For fixed time points  $s$  and  $t$ , consider the random quantity:

$$\rho(s, t) = n^{1/2}(\hat{\Lambda}_y(s) - \Lambda_y(s)) - n^{1/2}(\hat{\Lambda}_x(t) - \Lambda_x(t)). \quad (A.1)$$

Note that  $S_{2k}(g_{k0}) = \rho(t_k, g_{k0})$  and  $S_{3k}(h_{k0}) = \rho(h_{k0}, t_k)$ . Now, let  $\delta_{yj} = 1$ , if  $D_{yj}$  is observed, 0, otherwise. Also, let  $M_{yj}(t) = I(T_{yj} \leq t, \delta_{yj} = 1) - \int_0^t I(T_{yj} \geq s) d\Lambda_y(s)$ . Then,  $n^{1/2}(\hat{\Lambda}_y(s) - \Lambda_y(s)) \approx$

$$n^{-1/2} \sum_{j=1}^{n_y} \int_0^s n \left\{ \sum_{l=1}^{n_y} I(T_{yl} \geq u) \right\}^{-1} dM_{yj}(u), \quad (A.2)$$

which is asymptotically equivalent to a sum of independent identically distributed random variables:

$$n^{-1/2} \sum_{j=1}^{n_y} \int_0^s \{\tau_y(u)\}^{-1} dM_{yj}(u), \quad (\text{A.3})$$

where  $\tau_y(u)$  is the limit of  $\sum_{l=1}^{n_y} I(T_{yl} \geq u)/n$  (Fleming and Harrington 1991, Chapter 5). Similarly,  $n^{1/2}(\hat{\Lambda}_x(t) - \Lambda_x(t)) \approx$

$$n^{-1/2} \sum_{i=1}^{n_x} \int_0^t \{\tau_x(u)\}^{-1} dM_{xi}(u), \quad (\text{A.4})$$

which is a sum of independent identically distributed random variables, where  $M_{xi}(u) = I(T_{xi} \leq u, \delta_{xi} = 1) - \int_0^u I(T_{xi} \geq v) d\Lambda_x(v)$ ,  $\delta_{xi} = 1$  if  $D_{xi}$  is observed, 0, otherwise, and  $\tau_x$  is obtained by replacing  $y$  in  $\tau_y$  with  $x$ . It follows that  $\rho(s, t)$  in (A.1) is asymptotically equivalent to the quantity: (A.3) – (A.4).

Now, let  $S_{1k}(\theta_{k0}, g_{k0}, h_{k0})$  be denoted by  $n^{-3/2} \sum_i \sum_j \phi_{kij}(\theta_{k0}, g_{k0}, h_{k0})$ , which is a U-statistic. It follows from the standard large sample theory for the U-statistics, (A.3) and (A.4), that the limiting joint distribution of  $\{S_{1k}(\theta_{k0}, g_{k0}, h_{k0}), S_{2k}(g_{k0}), S_{3k}(h_{k0}), k = 1, \dots, K\}$  is normal with mean 0 and covariance matrix  $\Sigma$ .

A consistent estimator  $\hat{\Sigma}$  for  $\Sigma$  can be easily obtained. Firstly, the covariance between  $S_{1k}(\theta_{k0}, g_{k0}, h_{k0})$  and  $S_{1l}(\theta_{l0}, g_{l0}, h_{l0})$ ,  $k, l = 1, \dots, K$ , can be estimated by

$$n^{-3} \left\{ \sum_{i=1}^{n_x} \sum_{i'=1}^{n_x} \sum_{j=1}^{n_y} \hat{\phi}_{kij} \hat{\phi}_{li'j} + \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \sum_{j'=1}^{n_y} \hat{\phi}_{kij} \hat{\phi}_{li'j'} \right\},$$

where  $\hat{\phi}_{kij} = \phi_{kij}(\hat{\theta}_k, \hat{g}_k, \hat{h}_k)$  (Wei and Johnson 1985). Secondly, using the standard martingale theory (Fleming and Harrington 1991, Chapter 2), a consistent estimator  $\psi(s, t, \tilde{s}, \tilde{t})$  for the covariance between  $\rho(s, t)$  and  $\rho(\tilde{s}, \tilde{t})$  is

$$n \left[ \sum_{j=1}^{n_y} \delta_{yj} I(T_{yj} \leq \min(s, \tilde{s})) \left\{ \sum_{j'=1}^{n_y} I(T_{yj'} \geq T_{yj}) \right\}^{-2} + \sum_{i=1}^{n_x} \delta_{xi} I(T_{xi} \leq \min(t, \tilde{t})) \left\{ \sum_{i'=1}^{n_x} I(T_{xi'} \geq T_{xi}) \right\}^{-2} \right].$$

It follows that the covariance for  $S_{2k}(g_{k0})$  and  $S_{2l}(g_{l0})$  can be approximated by  $\psi(t_k, \hat{g}_k, t_l, \hat{g}_l)$ ,  $k, l = 1, \dots, K$ . Similarly,  $\text{cov}(S_{3k}(h_{k0}), S_{3l}(h_{l0}))$  and

$\text{cov}(S_{2k}(g_{k0}), S_{3l}(h_{l0}))$  can be estimated by  $\psi(\hat{h}_k, t_k, \hat{h}_l, t_l)$  and  $\psi(t_k, \hat{g}_k, \hat{h}_l, t_l)$ , respectively. Finally, the covariance between  $S_{1k}(\theta_{k0}, g_{k0}, h_{k0})$  and  $\rho(s, t)$  can be estimated by

$$\zeta_k(s, t) = n^{-1} \left[ \sum_i \sum_j \hat{\phi}_{kij} \int_0^s \left\{ \sum_{l=1}^{n_y} I(T_{yl} \geq u) \right\}^{-1} d\hat{M}_{yj}(u) \right. \\ \left. - \sum_i \sum_j \hat{\phi}_{kij} \int_0^t \left\{ \sum_{l=1}^{n_x} I(T_{xl} \geq u) \right\}^{-1} d\hat{M}_{xi}(u) \right],$$

where  $\hat{M}_{yj}(u) = I(T_{yj} \leq u, \delta_{yj} = 1) - \int_0^u I(T_{yj} \geq v) d\hat{\Lambda}_y(v)$  and  $\hat{M}_{xi}(\cdot)$  is obtained by replacing  $y$  and  $j$  in  $\hat{M}_{yj}$  with  $x$  and  $i$ , respectively. It follows that the covariance between  $S_{1k}(\theta_{k0}, g_{k0}, h_{k0})$  and  $S_{2k}(g_{k0})$  can be estimated by  $\zeta_k(t_k, \hat{g}_k)$ ; and the corresponding estimator for the covariance between  $S_{1k}(\theta_{k0}, g_{k0}, h_{k0})$  and  $S_{3k}(h_{k0})$  is  $\zeta_k(\hat{h}_k, t_k)$ .

Figure 1. Kaplan-Meier estimates of dependent censoring distributions for AIDS clinical trial

Table 1. Patterns of patient's censoring for the AZT and ddI groups

Reason for censoring	AZT	ddI
Patient's request	100	62
Physician's decision	24	24
Major toxicity	16	19
Death	9	10
Ad hoc reasons*	25	32
Administrative censoring	130	151

\* Loss to follow-up, protocol violation etc.

Table 2. Numbers of missing observations for the AIDS trial

	AZT Group $n_x=304$			ddI Group $n_y = 298$		
	Week			Week		
	8	16	24	8	16	24
Intermittent missing <sup>†</sup>	32	31	31	43	31	21
Administrative censoring	2	2	2	1	12	1
Dependent censoring	13	57	86	14	31	55

<sup>†</sup> *Mainly owing to missed visits*

Table 3. Simulation results – Growth Curve Model

Method	Week	$\theta_{k0}$	Average of $\hat{\theta}_k$	Empirical level of 0.95 confidence interval
New	20	30	29	0.93
	30	45	43	0.92
	40	60	57	0.93
Wu & Bailey	20	30	8	0.88
	30	45	12	0.88
	40	60	16	0.88
Naive	20	30	6	0.83
	30	45	7	0.83
	40	60	6	0.83

Table 4. Covariance matrices for the simulation study with  $K = 2$

Group $x$				Group $y$			
$X_{0i}$	$X_{1i}$	$X_{2i}$	$\log(D_{xi})$	$Y_{0j}$	$Y_{1j}$	$Y_{2j}$	$\log(D_{yj})$
$\left( \begin{array}{cccc} 25600 & 15300 & 15300 & 175 \\ 15300 & 25600 & 15300 & 175 \\ 15300 & 15300 & 25600 & 175 \\ 175 & 175 & 175 & 2 \end{array} \right)$	$\left( \begin{array}{cccc} 25600 & 15300 & 15300 & 450 \\ 15300 & 25600 & 15300 & 450 \\ 15300 & 15300 & 25600 & 450 \\ 450 & 450 & 450 & 14 \end{array} \right)$						

Table 5. Simulation results with  $K = 2$

Method	Week	$\theta_{k0}$	Average of $\hat{\theta}_k$	Empirical level of 0.95 confidence interval
New	16	25	26	0.95
	48	50	50	0.95
Naive	16	25	22	0.94
	48	50	18	0.75
Wu & Bailey	16	25	9	0.59
	48	50	27	0.87

Table 6. Covariance matrices for the simulation study with  $K = 5$

Group x

$X_{1i}$   $X_{2i}$   $X_{3i}$   $X_{4i}$   $X_{5i}$   $\log(D_{xi})$

---



---


$$\begin{pmatrix} 25600 & 15300 & 15300 & 15300 & 15300 & 693 \\ 15300 & 25600 & 15300 & 15300 & 15300 & 693 \\ 15300 & 15300 & 25600 & 15300 & 15300 & 693 \\ 15300 & 15300 & 15300 & 25600 & 15300 & 693 \\ 15300 & 15300 & 15300 & 15300 & 25600 & 693 \\ 693 & 693 & 693 & 693 & 693 & 19 \end{pmatrix}$$

Group y

$Y_{1j}$   $Y_{2j}$   $Y_{3j}$   $Y_{4j}$   $Y_{5j}$   $\log(D_{yj})$

---



---


$$\begin{pmatrix} 25600 & 15300 & 15300 & 15300 & 15300 & 289 \\ 15300 & 25600 & 15300 & 15300 & 15300 & 289 \\ 15300 & 15300 & 25600 & 15300 & 15300 & 289 \\ 15300 & 15300 & 15300 & 25600 & 15300 & 289 \\ 15300 & 15300 & 15300 & 15300 & 25600 & 289 \\ 289 & 289 & 289 & 289 & 289 & 6 \end{pmatrix}$$

Table 7. Simulation results when correlation between the outcome and censoring variables depends on the treatment\*

$$(\theta_{01} = 25, \theta_{02} = 50)$$

Correlation $\rho$ in group $y$	New method		Naive		
	Average of $\hat{\theta}$	Level of 0.95 interval	Average of $\hat{\theta}$	Level of 0.95 interval	
0.75					
	$\theta_1$	24	0.96	22	0.95
	$\theta_2$	50	0.95	18	0.78
0.50					
	$\theta_1$	17	0.91	16	0.92
	$\theta_2$	28	0.87	9	0.67
0.20					
	$\theta_1$	2	0.85	6	0.89
	$\theta_2$	1	0.65	-7	0.51
0.00					
	$\theta_1$	-4	0.80	-3	0.85
	$\theta_2$	-17	0.42	-17	0.38
-0.50					
	$\theta_1$	-26	0.53	-18	0.64
	$\theta_2$	-61	0.04	-39	0.10

\*The correlation for x-group is 0.75.

## REFERENCES

- Dawson, J.D., and Lagakos, S.W. (1993), "Size and Power of Two-Sample Tests of Repeated Measures Data," *Biometrics*, 49, 1022-32.
- DeGruttola, V.G., and Tu, X.M. (1994), "Modeling Progression of CD4-Lymphocyte Count and its Relationship to Survival Time," *Biometrics*, 50, 1003-1014.
- Fisher, L.D., Dixon, D.O., Herson, J., Frankowski, R.K., Hearron, M.S., and Peace, K.E. (1990), Intention to treat in clinical trials. In *Statistical Issues in Drug Research and Development*, in K.E. Peace (ed), 331-350. New York: Marcel Dekker.
- Fleming, T.R., and Harrington, D.P. (1991), *Counting Processes and Survival Analysis*, New York: John Wiley.
- Kahn, J.O., Lagakos, S.W., Richman, D.D., Cross, A., Pettinelli, C., Liou, S.H., Brown, M., Volberding, P.A., Crumpacker, C.S., Beall, G., Sacks, H.S., Merigan, T.C., Beltangady, M., Smaldone, L., Dolin, R., and the AIDS Clinical Trials Group. (1992), "A Controlled Trial Comparing Continued Zidovudine with Didanosine in Human Immunodeficiency Virus Infection," *The New England Journal of Medicine*, 327, 581-587.
- Laird, N.M., and Ware, J.H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963-974.
- Lehmann, E.L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.
- Liang, K.Y., and Zeger, S.L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13-22.
- Lin, D.Y., Robins, J.M., and Wei, L.J. (1996), "Comparing Two Failure Time Distributions in the Presence of Informative Censoring," *Biometrika*, 83, 381-393.

- Little, R.J.A., and Rubin, D.B. (1987), *Statistical Analysis with Missing Data*, New York: John Wiley.
- Mori, M., Woodworth, G.G., and Woolson, R.F. (1992), "Application of Empirical Bayes Inference to Estimation of Rate Change in the Presence of Informative Right Censoring," *Statistics in Medicine*, 11, 621-631.
- Oakes, D., Moss, A.J., Fleiss, J.L., Bigger, J.T., Jr., Therneau, T., Eberly, S.W., McDermott, M.P., Manatunga, A., Carleen, E., Benhorin, J., and Multicenter Diltiazem on mortality and reinfarction after myocardial infarction. *Journal of the American Statistical Association*, 88, 44-49.
- Parzen, M.I., Wei, L.J., and Ying, Z. (1994), "A Resampling Method Based on Pivotal Estimating Functions," *Biometrika*, 81, 341-50.
- Robins, J.M. (1994), "An Analytic Method for Randomized Trials with Informative Censoring," Technical Report, Department of Epidemiology, Harvard School of Public Health.
- Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1995), "Analysis of Semi-parametric Regression Models Under the Presence of Missing Data," *Journal of the American Statistical Association*, 90, 106-121.
- Rubin, D.B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581-590.
- Wei, L.J., and Johnson, W.E. (1985), "Combining Dependent Tests with Incomplete Repeated Measurements," *Biometrika*, 72, 359-364.
- Wei, L.J., and Lachin, J.M. (1984), "Two-Sample Asymptotically Distribution-Free Tests for Incomplete Multivariate Observations," *Journal of the American Statistical Association*, 79, 653-661.
- Wei, L.J., Ying, Z., and Lin, D.Y. (1990), "Linear Regression Analysis of Censored Survival Data Based on Rank Tests," *Biometrika*, 77, 845-51.

- Wu, M.C., and Bailey, K.R. (1988), "Analyzing Changes in the Presence of Informative Right Censoring Caused by Death and Withdrawal," *Statistics in Medicine*, 7, 337-346.
- Wu, M.C., and Bailey, K.R. (1989), "Estimation and Comparison of Changes in the Presence of Informative Right Censoring: Conditional Linear Model," *Biometrics*, 45, 939-955.
- Wu, M.C., and Carroll, R.J. (1988), "Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process," *Biometrics*, 44, 175-188.
- Wulfsohn, M.S., and Tsiatis, A.A. (1997), "A Joint Model for Survival and Longitudinal Data Measured with Error," *Biometrics*, 53, 330-339.