STATISTICS IN MEDICINE, VOL. 11, 1871–1879 (1992)

THE ACCELERATED FAILURE TIME MODEL: A USEFUL ALTERNATIVE TO THE COX REGRESSION MODEL IN SURVIVAL ANALYSIS

L. J. WEI

Department of Biostatistics, Harvard School of Public Health, 677 Huntington Ave., Boston, MA 02115, U.S.A.

SUMMARY

For the past two decades the Cox proportional hazards model has been used extensively to examine the covariate effects on the hazard function for the failure time variable. On the other hand, the accelerated failure time model, which simply regresses the logarithm of the survival time over the covariates, has seldom been utilized in the analysis of censored survival data. In this article, we review some newly developed linear regression methods for analysing failure time observations. These procedures have sound theoretical justification and can be implemented with an efficient numerical method. The accelerated failure time model has an intuitive physical interpretation and would be a useful alternative to the Cox model in survival analysis.

1. INTRODUCTION

Let T be the 'failure time', the response variable, and x be the corresponding covariate vector. Suppose that we are interested in making inferences about the effect from x on the response variable T. If there is no censored observation at the time of the analysis, most likely we would regress T or a transformation of it directly on the covariate x. This approach is quite appealing to practitioners owing to its ease of interpretation. However, if there are censored observations in the data, we usually do not take this conventional approach to examine the covariate effect. In fact, we almost exclusively use the proportional hazards (PH) model with the partial likelihood principle to draw inferences about the covariate effect.[1] The PH model is fairly flexible. It is semi-parametric. Its covariates can be time-dependent. Large-sample properties of its inference procedures have been beautifully justified with the martingale theory.[2] Moreover, those procedures can be easily obtained through commercial packages.

However, since the PH model specifies that the effect of the covariate x is to act multiplicatively on the hazard function, it is not easy to interpret, for example, the estimates of regression parameters. So, if the ordinary linear regression model can handle censored observations, it would be a useful alternative to the Cox model in the survival analysis. In fact, with the presence of censored observations, the linear regression analysis with log T as the response variable has been studied extensively.[3–6] This linear model with the log-transformation is called the accelerated failure time model (AFT) (see Reference 7, Chapter 6). Using Stanford heart transplant data, Miller and Halpern[8] conclude that the Buckley–James (BJ) procedure is preferred. Unfortunately, the original BJ method has no theoretical justification and does not provide a reliable numerical method for implementation. Therefore, it has seldom been used in practice.

0277–6715/92/141871–09$09.50

Recently quite a few interesting papers regarding the AFT model have been published. In this article, we give a brief review of some newly developed methods for the analysis of censored data with the AFT model. All these procedures have sound theoretical justification and can be implemented with an efficient numerical method. Other applications of the AFT model to the survival analysis are also discussed with real-life examples.

2. ACCELERATED FAILURE TIME MODEL

Let T_i be the failure time for the ith patient, $i = 1, \ldots, n$. For T_i, we can only observe a bivariate vector (Y_i, Δ_i), where $Y_i = \min(T_i, c_i)$ and $\Delta_i = 1$ if $T_i = Y_i$ and 0 otherwise. The c_is are censoring variables. Conditional on the covariates for the ith subject, c_i is assumed to be independent of the failure times T_i, $i = 1, \ldots, n$. Let x_i denote a $p \times 1$ vector of covariates for T_i. Suppose that the base 10 logarithm of T_i is linearly related to x_i, that is, there exists an unknown constant β such that

$$\log T_i = \beta' x_i + e_i, \tag{1}$$

where e_i, $i = 1, \ldots, n$, are independent and identically distributed random variables whose common distribution function F is *completely* unspecified. Note that the logarithm in (1) may be replaced by other strictly increasing functions. Also, in (1) the intercept parameter is not included in the vector β. Therefore, the mean of the error term may not be 0. Owing to the presence of censoring, usually the intercept parameter cannot be estimated well.

In this article, we present several valid methods of inferences about β_1, which is a subset of β, while regarding others as nuisance parameters. For instance, one may be interested in testing the null hypothesis H_0: $\beta_1 = \beta_{10}$, a given $q \times 1$ vector. Since the error distribution is completely arbitrary, test statistics for testing H_0 cannot be derived directly from usual likelihood functions.

Let us use the well-known Stanford heart transplant data to illustrate the above setup. These data contain the survival times of 184 heart-transplanted patients along with their ages at the time of the first transplant and T5 mismatch scores. Twenty-seven patients did not have T5 scores. Among the remaining 157 patients, 55 were censored as of February 1980.[8] To examine the age effect on T with the adjustment from the T5 score, an obvious model is

$$\log T_i = \beta_1 x_{1i} + \beta_2 x_{2i} + e_i, \tag{2}$$

where $i = 1, \ldots, 157$, x_{1i} is the age of the first transplant and x_{2i} is the corresponding T5 score for the ith patient. In the next section, we show how to make inferences about the age effect on the survival time T. Moreover, we will discuss if this simple additive model fits the data from the Stanford Study well.

3. PROCEDURES BASED ON RANK TEST STATISTICS

In this section, we present inference procedures studied in References 9–12. These methods are essentially derived from the ingenious Cox's PH model with the partial likelihood principle.

Consider the usual Cox's proportional hazards model with regression coefficient vecotr γ, say. The partial likelihood score function, which is the derivative of the logarithm of the partial likelihood function with respect to γ, is simply:

$$S(\gamma) = \sum_{i=1}^{n} \Delta_i \left(x_i - \frac{\sum_{k=1}^{n} x_k I(Y_k \geq Y_i) e^{\gamma' x_k}}{\sum_{k=1}^{n} I(Y_k \geq Y_i) e^{\gamma' x_k}} \right).$$

Now, if the parameter γ is 0, that is, the underlying failure times T_i are independent and identically

distributed, the above score function becomes

$$S(0) = \sum_{i=1}^{n} \Delta_i \left(x_i - \frac{\sum_{k=1}^{n} x_k I(Y_k \geq Y_i)}{\sum_{k=1}^{n} I(Y_k \geq Y_i)} \right). \tag{3}$$

It is important to note that as long as the underlying failure times T_i are independent and identically distributed, for large n, $S(0)$ is approximately centred around 0. This is true even if the censoring variable c_i depends on the covariate x_i. Furthermore, the large-sample distribution of $S(0)$ can be approximated by a normal random vector with mean 0 and variance–covariance matrix Λ, which is the inverse of the matrix $(\partial S(\gamma)/\partial\gamma|_{\gamma=0})$. These large-sample properties of $S(0)$ were given in Cox[1] and then justified elegantly by Andersen and Gill.[2] For the two-sample case, namely when the covariate x is simply the group indicator, $S(0)$ is the well-known logrank test statistic.

Now, let us go back to our log-linear model (1). Assume that β_0 is the true value of β. Consider $e_i(\beta_0) = \log Y_i - \beta_0' x_i$, which is the minimum of $\tilde{T}_i = (\log T_i - \beta_0' x_i)$ and $\tilde{c}_i = (\log c_i - \beta_0' x_i)$. Note that by definition, the $\tilde{T}_i$ are independent and identically distributed. Therefore, for large n, the score function $S(0)$ in (3) with $\{Y_i, \Delta_i, i = 1, \ldots, n\}$ being replaced by $\{e_i(\beta_0), \Delta_i, i = 1, \ldots, n\}$ is also centred about 0. Let the resulting function be denoted by $U(\beta_0)$. A good point estimate for β_0 is a quantity which is close to β_0. This implies that $U(\beta)$, evaluated at this estimate, would be around 0. Therefore, if the equation $U(\beta) = 0$ has a unique root, say $\hat{\beta}$, then it would be a reasonable estimate for β_0. This $\hat{\beta}$ is one of the rank estimates proposed by Tsiatis[9] and Lai and Ying.[11] For the two-sample case, various properties of $\hat{\beta}$ have been carefully studied by Louis[13] and Wei and Gail.[14] In general, we call $U(\beta)$ an estimating function of β.

In fact, we can use a more general estimating function to estimate β. For instance, a weight function $\phi(\hat{F}_\beta(\cdot))$ could be placed in front of each term in $U(\beta)$, where ϕ is a smooth function whose values are between 0 and 1 and $\hat{F}_\beta(\cdot)$ is the usual Kaplan–Meier estimate for the distribution function of the error term e_i in (1) based on the incomplete data $\{e_i(\beta), \Delta_i\}$. Let the estimating function with weight ϕ be denoted by $U_\phi(\beta)$, where

$$U_\phi(\beta) = \sum_{i=1}^{n} \Delta_i \phi(\hat{F}_\beta(e_i(\beta))) \left(x_i - \frac{\sum_{k=1}^{n} x_k I(e_k(\beta) \geq e_i(\beta))}{\sum_{k=1}^{n} I(e_k(\beta) \geq e_i(\beta))} \right)$$

In this section, we will discuss inference procedures derived from two specific U_ϕ, where $\phi(u) = 1$ and $\phi(u) = (1 - u)$. For the two-sample case, the statistic $U_\phi(0)$ with $\phi(u) = (1 - u)$ is called the Peto–Prentice Wilcoxon test statistic which is sensitive to detect early differences between two survival functions.

For a smooth function ϕ, $U_\phi(\beta_0)$ is still approximately centred around 0. We let $\hat{\beta}_\phi$ be the estimate which minimizes $\| U_\phi(\beta) \|$. The estimate $\hat{\beta}_\phi$ is consistent; that is, for large n, $\hat{\beta}_\phi$ is fairly close to β_0. Moreover, the large-sample distribution of $\hat{\beta}_\phi$ can be approximated by a multivariate normal distribution with mean β_0 and variance–covariance matrix Γ.[9,11] In theory, we can use $\hat{\beta}_\phi$ with its large-sample distribution to make inferences about β_0. However, the matrix Γ involves the unknown hazard function of the error term e_i. This function may not be estimated well with censored data. At present, there does not exist a reliable estimate of Γ for practical usage.

To avoid estimating the underlying hazard function, we may use the method proposed by Wei *et al.*[12] for making inference about β_0. Like the $U(\beta_0)$, $U_\phi(\beta_0)$ is also approximately normal with mean 0 and covariance matrix Λ_ϕ, where

$$\Lambda_\phi = \sum_{i=1}^{n} \phi^2(\hat{F}_{\hat{\beta}_\phi}(e_i(\hat{\beta}_\phi))) \Delta_i \left[\frac{\sum_{j=1}^{n} I(e_j(\hat{\beta}_\phi) \geq e_i(\hat{\beta}_\phi)) x_j^{\otimes 2}}{\sum_{j=1}^{n} I(e_j(\hat{\beta}_\phi) \geq e_i(\hat{\beta}_\phi))} - \left\{ \frac{\sum_{j=1}^{n} I(e_j(\hat{\beta}_\phi) \geq e_i(\hat{\beta}_\phi)) x_j}{\sum_{j=1}^{n} I(e_j(\hat{\beta}_\phi) \geq e_i(\hat{\beta}_\phi))} \right\}^{\otimes 2} \right],$$

and $a^{\otimes 2}$ is aa' for a column vector a. Suppose that we are interested in the entire vector of parameters. For example, let us consider the problem of testing the hypothesis $\beta = \beta_0$. Then, an obvious test statistic is

$$U'_\phi(\beta_0)\Lambda_\phi^{-1}U_\phi(\beta_0).$$

Under H_0, for large n, the above statistic is approximately chi-square distributed with degrees of freedom p. A large observed value of this statistic suggests a rejection of the hypothesis $\beta = \beta_0$.

Now, let us assume that we are interested in a subset of β. For example, one would like to test $H_0: \beta_1 = \beta_{10}$, a given q-dimensional vector, against a general alternative. Let $\beta = (\beta_1', \beta_2')'$, where β_2 is the vector of nuisance parameters. Here, we present a simple way to get rid of the nuisance parameters β_2 in the analysis. Consider the logarithm of the likelihood ratio statistic based on a single observation $\hat{\beta}_\phi$ for testing H_0, assuming that $\hat{\beta}_\phi$ is exactly normally distributed with an unknown mean β and a 'known and fixed' covariance matrix $\hat{\Gamma}$, a consistent estimate of Γ. This 'log-likelihood ratio statistic' is proportional to

$$\min[(\hat{\beta}_{1\phi} - \beta_{10})', (\hat{\beta}_{2\phi} - \beta_2)']\hat{\Gamma}^{-1}[(\hat{\beta}_{1\phi} - \beta_{10})', (\hat{\beta}_{2\phi} - \beta_2)']', \tag{4}$$

where the minimization is taken with respect to β_2, $\hat{\beta}_\phi = (\hat{\beta}_{1\phi}\ \hat{\beta}_{2\phi})'$ and $\hat{\beta}_{1\phi}$ and $\hat{\beta}_{2\phi}$ are the corresponding estimates for β_1, and β_2, respectively.

It can be shown that, for large n, (4) is approximately chi-square distributed with q degrees of freedom (see Reference 15, Appendix B). Recall that Γ is rather difficult to estimate well with censored data. Therefore, (4) is not very useful in practice. Luckily, it is not difficult to show that (4) is asymptotically equivalent to

$$V_\phi(\beta_{10}) = \min\left\{U'_\phi\begin{pmatrix}\beta_{10}\\ \beta_2\end{pmatrix}\Lambda_\phi^{-1}U_\phi\begin{pmatrix}\beta_{10}\\ \beta_2\end{pmatrix}\right\}, \tag{5}$$

where the minimization is taken with respect to β_2. A large observed value of $V_\phi(\beta_{10})$ suggests that H_0 may not be correct. Confidence regions for β_1 can be obtained by inverting this test statistic. For example, a $1 - \alpha$ region for β_1 is

$$\{\beta_1 : V_\phi(\beta_1) < \chi_q^2(1-\alpha)\},$$

where $\chi_q^2(1-\alpha)$ is the upper α point of a chi-square distribution with degrees of freedom q. Note that in (5), the function to be minimized is discrete in β_2. Such minimization cannot be achieved with the conventional optimization algorithms for smooth functions. Fortunately, there is an efficient numerical method to solve such optimization problems (see Section 5).

Now, let us use the Stanford data to illustrate the above methods. For model (2), with $\phi = 1$, the estimated age effect $\hat{\beta}_1$ is $-0{\cdot}025$, and the estimated T5 effect $\hat{\beta}_2$ is $-0{\cdot}124$. Treating the T5 effect as a nuisance parameter in (5), a 0·95 confidence interval for the age effect β_1 is $(-0{\cdot}047, -0{\cdot}007)$. On the other hand, if the age effect is nuisance, a 0·95 interval for β_2 is $(-0{\cdot}395, 0{\cdot}197)$, indicating that the T5 score is not significant at all. Based on this simple additive model with age and T5 score, age has a negative effect on the survival. That is, the younger patients tend to survive longer than the older ones after the surgery. For example, a 40-year-old person could live about 1·75 ($= 10^{0{\cdot}25}$) times as long as a 50-year-old after the heart transplant.

If we use the weight function $\phi(u) = 1 - u$ in (5) for our analysis, then $\hat{\beta}_1$ and $\hat{\beta}_2$ are $-0{\cdot}021$ and $-0{\cdot}062$, respectively. Note that if model (2) fits the Stanford data well, we expect that $\hat{\beta}$ derived from different weight functions should not differ too much. However, the above two estimates for β_2 are significantly different.[12] This indicates that the simple model (2) may not be appropriate for our data. Indeed the residual plots by Miller and Halpern[8] also discredit this

model. In an attempt to get a better fit, a quadratic age model without T5 score was tried by Miller and Halpern.[8] For this quadratic model, using the method based on the estimating function $U_\phi(\beta)$ with ϕ being 1, the estimates of the age and age^2 effects are $+0{\cdot}099$ and $-0{\cdot}0016$, respectively. The corresponding 0·95 confidence intervals are $(-0{\cdot}007, 0{\cdot}189)$ and $(-0{\cdot}0028, -0{\cdot}0003)$. So, in fact, not only were the old patients doing poorly in the study, but the very young ones were having problems as well.

One of the main goals for fitting failure time observations with a regression model is to predict, for example, the t-year survival probability for future patients with some specific covariate vector x_0. With an accelerated failure time model, a consistent estimate for such conditional probability $\tau = Pr(T > t | X = x_0)$ is $\hat{\tau} = \hat{S}(t; \hat{\beta}_\phi)$, where $\hat{S}(t; \beta) = 1 - \hat{F}_\beta(\log t - \beta' x_0)$. The large-sample distribution of $\hat{\tau}$ can be obtained through large-sample properties of $\hat{\beta}_\phi$ and the so-called delta method. However, the large-sample variance of $\hat{\tau}$ is again a function of the unknown hazard function of F.

Now, let us consider interval estimation procedures for τ without using non-parametric functional estimates. It is easy to show that, for large n, $U_\phi(\beta_0)$ and $(\hat{S}(t; \beta_0) - \tau)$ are approximately independent and are normally distributed.[16] Using the Greenwood formula for the KM estimate, a large-sample variance $\sigma^2(\beta_0)$ for $\hat{S}(t; \beta_0)$ can be easily obtained. For testing the hypothesis that the t-year survival probability is, say, τ_0, the logarithm of the 'likelihood ratio statistic', assuming that $\hat{\beta}_\phi$ and $\hat{\tau}_0$ are exactly normally distributed with known covariance matrices, is asymptotically equivalent to

$$Q(\tau_0) = \min[(\hat{S}(t; \beta) - \tau_0)^2 (\sigma(\beta))^{-2} + U_\phi(\beta)' \Lambda_\phi^{-1} U_\phi(\beta)],$$

where the minimization is taken with respect to β. If $\tau = \tau_0$, $Q(\tau_0)$ is approximately chi-square distributed with one degree of freedom. Confidence intervals for τ can then be obtained by inverting the above test statistic; for example, a $1 - \alpha$ interval is

$$\{\tau : Q(\tau) < \chi_1^2(1 - \alpha)\}.$$

For the Stanford study, suppose that we are interested in the estimation of 1·5-year survival probability with patient's entry age being 45. Using weight $\phi = 1$ in $U_\phi(\beta)$ and a log-linear model with linear and quadratic age effects, the point and 0·95 confidence intervals for τ are 0·55 and (0·45, 0·63), respectively. If $\phi(u) = 1 - u$, the estimates are 0·55 and (0·47, 0·63), respectively. More details on the prediction based on the accelerated failure time model are given in Ying *et al.*[16]

An interesting question is how to choose the weight function ϕ for inferences about β_0. Presumably one can use the so-called adaptive procedure in the analysis. That is, we can estimate the hazard function of F non-parametrically and then choose proper weight ϕ in U_ϕ based on this estimate. This kind of approach is interesting in theory. However, with the presence of censored data, it is not clear that we can gain much from such complicated procedures for practical sample sizes.

Since the logrank and Peto–Prentice Wilcoxon tests are the most popular two-sample tests in the survival analysis, one may ask which weight function, either $\phi = 1$ or $\phi(u) = 1 - u$, would be a better choice in general. Based on extensive numerical studies, Lin and Wei[15] conclude that the procedure with the Wilcoxon score generally performs better than that with the logrank score.

4. BUCKLEY–JAMES PROCEDURE

In this section, we show how to use the least squares principle to make inferences about β_0 in the accelerated failure time model (1). One such approach was introduced by Buckley and James.[5] Unfortunately, their method did not have theoretical justification.

Consider the case that there are no censored observations. The normal equation for β is simply:

$$\sum_{i=1}^{n} (x_i - \bar{x}) (\log Y_i - \beta' x_i) = 0, \tag{6}$$

where $\bar{x} = n^{-1} \sum x_i$. However, if the log Y_i are censored, we replace $e_i(\beta) = (\log Y_i - \beta' x_i)$ with

$$e_i(\beta) + \int_{e_i(\beta)}^{\infty} (1 - \hat{F}_\beta(s))\mathrm{d}s / \{1 - \hat{F}_\beta(e_i(\beta))\} \tag{7}$$

in (6). Let the resulting equation be denoted by $W(\beta)$. If Δ_i is 0, that is, log Y_i is censored, the second part of (7) is an estimate of the 'expected residual life time' for $e_i(\beta)$.

For large n, $W(\beta_0)$ is approximately centred around 0. The estimate $\tilde{\beta}$ which minimizes $\| W(\beta) \|$ is consistent.[17] The large-sample distribution of $\tilde{\beta}$ can be approximated by a normal distribution with mean β_0. However, the corresponding covariance matrix again involves the unknown hazard function. On the other hand, $W(\beta_0)$ is approximately normal with mean 0 and covariance Λ_{BJ}, which does not involve any non-parametric functional estimate. Therefore, for testing $H_0 : \beta_1 = \beta_{10}$, one may use the following 'likelihood ratio'-like test statistic:

$$\tilde{Q}(\beta_{10}) = \min \left\{ W' \begin{pmatrix} \beta_{10} \\ \beta_2 \end{pmatrix} \Lambda_{BJ}^{-1} W \begin{pmatrix} \beta_{10} \\ \beta_2 \end{pmatrix} \right\},$$

where the minimization is taken with respect to β_2. For large n, under H_0, $\tilde{Q}$ is approximately chi-square distributed with q degrees of freedom. Confidence regions for β_{10} can be obtained based on such $\tilde{Q}$ accordingly.[18]

For the Stanford study, with the age and age^2 in model (1) and using base 10 logarithm for the transformation of T_i, the BJ estimate for the age effect is 0·109, and a 0·95 confidence interval is (0·022, 0·192). A 0·95 interval for the age effect based on the original BJ procedure[5] is (0·034, 0·18), which is shorter than the previous one. However, we find that the original BJ method tends to give liberal confidence intervals, that is, the coverage probabilities for the original BJ interval procedures tend to be smaller than their nominal levels.

From extensive numerical studies conducted by Lin and Wei,[15] as expected, the BJ method based on $W(\beta)$ is not robust, that is, it is sensitive to the outliers. In an excellent paper by Ritov,[10] the Buckley–James estimating function $W(\beta)$ has been extended to a more general setting to deal with the robustness issue.

5. COMPUTATIONAL METHOD

Recall that all the procedures discussed here are obtained through minimization of discrete functions. For example, to get a point estimate $\hat{\beta}_\phi$ we have to minimize $\| U_\phi(\beta) \|$, which is discontinuous and may have multiple local minima. Such minimization problems cannot be solved with a conventional optimization algorithm designed for smooth functions. Naturally, the reliable, but extremely time consuming, grids search method can be used to minimize discrete functions. However, for high-dimensional cases, that is when p is large, this method is not practical at all.

Recently, Lin and Geyer[19] have successfully applied the so-called simulated annealing method[20] to the present problems. Here, we simply give a brief description of their method. To avoid extremely uneven contributions from individual components from the covariate vector x to the value of the objective function, we divide all covariates by their sample standard deviations. For illustration, let us consider the problem of minimizing $\| U_\phi(\beta) \|$ with respect to β, a $p \times 1$

vector. Let the current value of β be denoted by β_c. Then, we generate a multivariate normal vector β^* from a distribution $MN(\beta_c, \sigma^2 I_p)$, where σ is some given constant and I_p is the $p \times p$ identity matrix. Let $D = \| U_\phi(\beta^*) \| - \| U_\phi(\beta_c) \|$. If $D \leqslant 0$, set $\beta_c = \beta^*$. If $D \geqslant 0$, to avoid being trapped in a local 'valley', we set $\beta_c = \beta^*$ with probability $\exp(-D/w)$, where w is some preassigned constant. We then decrease w and σ slightly, generate another normal random vector β^*, and check if β_c should be replaced by this new β^*. Repeat this process, say, m times, and declare that the last β_c is the estimate $\hat{\beta}_\phi$ for β_0. Based on their extensive numerical studies, Lin and Geyer[19] provide some practical guidelines for choosing w, σ, and m. They also conclude that the simulated annealing method is very efficient and reliable in dealing with the regression problem presented here.

6. ANALYSIS OF MULTIVARIATE FAILURE TIME DATA

Suppose that during his or her follow-up period, each patient may experience a number of failures which correspond to repeated occurrences of the same type of event or to the occurrence of events of entirely different natures. The corresponding event times, which are possibly censored, are recorded for each patient. For example, in a randomized trial to evaluate the efficacy of AZT for treating patients with AIDS or AIDS-related complex,[21] 281 patients were enrolled in the study, among whom 144 were assigned to AZT and 137 to placebo. At the end of the study, 44 patients had first opportunistic infection episodes and 19 patients died from AIDS.[22] Death is a natural endpoint for the study. However, opportunistic infections are life-threatening infections and are also regarded as major endpoints for typical AIDS trials. It is important to know how to use those event times from various types of failures to make inferences regarding the efficacy of AZT.

Recently, Lin and Wei[15] have generalized the univariate linear regression method *described in* Sections 3 and 4 to the multivariate case by regressing the logarithm of each marginal failure time linearly on its covariates. However, no specific structure of dependence among the distinct failure times on each subject has to be imposed. Their methods provide an overall evaluation for the treatment difference in a typical two-arm trial with multiple endpoints.

It is not clear that the procedures proposed by Lin and Wei[15] are useful when the related event times are from the same type of failure. For example, in the Diabetes Control and Complications Trial[23] patients are randomly assigned to receive either experimental or standard therapy. The purpose of the study is to assess the relationship between glycaemic control and the development or progression of early vascular complications in persons with insulin-dependent diabetes mellitus. Experimental therapy involves the use of an intensive insulin regimen designed to maintain near normal glycaemic levels in the absence of severe hypoglycaemia. Standard treatment is designed to maintain subjects free of clinical symptoms related to hyper- or hypoglycaemia while receiving up to two insulin injections daily. One of the principal outcomes in this trial is the initial appearance of background retinopathy for individual eyes of the study patients. Therefore, for the DCCT, it is natural to treat individual eyes as sample units. Since the treatment affects the whole body of the patient, there are two failure time observations from each study patient. Lee *et al.*[24] use a population-averaged linear model approach (see Reference 25 for an excellent review on this topic) to analyse such highly stratified failure time data. Their idea is similar to that proposed by Liang and Zeger[26] for non-censored correlated observations.

7. OTHER DEVELOPMENTS

Danyu Lin at the University of Washington has proposed an interesting and useful data monitoring scheme for two-group comparisons with adjustment from other covariates based on

accelerated failure time model (1) using the estimating function $U_\phi(\beta)$. However, it seems rather difficult to invert his repeated test procedure to obtain the repeated confidence intervals for the treatment difference.

Recently Robins and Tsiatis[27] have extended (1) to the case with time-dependent covariates.

8. REMARKS

The purpose of this article is to show that the accelerated failure time model is a useful alternative to the Cox model in the analysis of survival observations. Naturally there are still many interesting and important research problems for such log-linear models. For example, analytical and graphical methods are needed for model checking and outlier detection. It is highly desirable to have robust inference procedures which are valid under weaker conditions than the 'independent and identically distributed' assumption for the error terms in the model. In fact, almost all the design and analysis issues for the ordinary linear model with non-censored observations need to be addressed in the presence of censoring.

ACKNOWLEDGEMENT

The author is grateful to a reviewer for helpful comments on the original version of the paper. This work was supported by NIH grants AI29131 and CA56844, and contract AI95030.

REFERENCES

1. Cox, D. R. 'Regression models and life-tables' (with discussion), *Journal of the Royal Statistical Society, Series B*, **34**, 187–220 (1972).
2. Andersen, P. K. and Gill, R. D. 'Cox's regression model for counting processes: a large sample study', *The Annals of Statistics*, **10**, 1100–1120 (1982).
3. Miller, R. 'Least squares regression with censored data', *Biometrika*, **63**, 449–464 (1976).
4. Prentice, R. L. 'Linear rank tests with right censored data', *Biometrika*, **65**, 167–179 (1978).
5. Buckley, I. V. and James, I. 'Linear regression with censored data', *Biometrika*, **66**, 429–436 (1979).
6. Koul, H., Susarla, V. and Van Ryzin, J. 'Regression analysis with randomly right censored data', *The Annals of Statistics*, **9**, 1276–1288 (1981).
7. Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1980.
8. Miller, R. and Halpern, J. 'Regression with censored data', *Biometrika*, **69**, 521–531 (1982).
9. Tsiatis, A. A. 'Estimating regression parameters using linear rank tests for censored data', *The Annals of Statistics*, **18**, 354–372 (1990).
10. Ritov, Y. 'Estimation in a linear regression model with censored data', *The Annals of Statistics*, **18**, 303–328 (1990).
11. Lai, T. L. and Ying, Z. 'Linear rank statistics in regression analysis with censored or truncated data', *Journal of Multivariate Analysis*, **40**, 13–45 (1992).
12. Wei, L. J., Ying, Z. and Lin, D. Y. 'Linear regression analysis of censored survival data based on rank tests', *Biometrika*, **77**, 845–851 (1990).
13. Louis, T. 'Nonparametric analysis of an accelerated failure time model', *Biometrika*, **68**, 381–390 (1981).
14. Wei, L. J. and Gail, M. H. 'Nonparametric estimation for a scale-change with censored observations', *Journal of American Statistical Association*, **78**, 382–388 (1983).
15. Lin, J. S. and Wei, L. J. 'Linear regression analysis for multivariate failure time observations', *Journal of American Statistical Association*, to appear (1992).
16. Ying, Z., Wei, L. J. and Lin, J. S. 'Predication of survival probability based on linear regression model', *Biometrika*, **79**, 205–209 (1992).
17. Lai, T. L. and Ying, Z. 'Large-sample theory of a modified Buckley–James estimator in regression analysis with censored data', *The Annals of Statistics*, **19**, 1370–1402 (1991).
18. Lin, J. S. and Wei, L. J. 'Linear regression analysis based on Buckley–James estimating equations', *Biometrics*, September (1992).

19. Lin, D. Y. and Geyer, C. J. *Computational Methods for Semiparametric Linear Regression with Censored Data*, Technical Report, Department of Biostatistics, University of Washington, 1991.
20. Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. 'Optimization by simulated annealing', *Science*, **220**, 671–680 (1983).
21. Fischl, M. A. *et al.* 'The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex', *New England Journal of Medicine*, **317**, 185–191 (1987).
22. Lin, D. Y. 'Nonparametric sequential testing in clinical trials with incomplete multivariate observations', *Biometrika*, **78**, 123–132 (1991).
23. The DCCT Group. 'The Diabetes Control and Complications Trial: the design and methodological considerations for feasibility phase', *Diabetes*, **35**, 530–545 (1986).
24. Lee, E., Wei, L. J. and Ying, Z. 'Linear regression analysis for highly stratified failure time data', *Journal of American Statistical Association*, to appear (1993).
25. Zeger, S. L., Liang, K. Y. and Albert, P. S. 'Models for longitudinal data: a generalized estimating equation approach', *Biometrics*, **44**, 1049–1060 (1988).
26. Liang, K. Y. and Zeger, S. L. 'Longitudinal data analysis using generalized linear models', *Biometrika*, **73**, 13–22 (1986).
27. Robins, J. M. and Tsiatis, A. A. 'Semiparametric estimation of an accelerated failure time model with time-dependent covariates', *Biometrika*, **79**, 311–320 (1992).