BIOMETRICS 50, 632–639
September 1994

Regression Analysis of Multivariate Grouped Survival Data

S. W. Guo

Department of Biostatistics, University of Michigan, 109 S. Observatory, Ann Arbor, Michigan 48109, U.S.A.

and

D. Y. Lin

Department of Biostatistics, SC-32, University of Washington, Seattle, Washington 98195, U.S.A.

SUMMARY

Multivariate failure time data arise when each study subject may experience several types of event or when there are clusterings of observational units such that failure times within the same cluster are correlated. The failure times are often subject to interval grouping or have truly discrete measurements. In this paper, the marginal distribution for each discrete failure time variable is formulated by a grouped-data version of the proportional hazards model while the dependence structure is unspecified. Generalized estimating equations in the spirit of Liang and Zeger (1986, *Biometrika* **73,** 13–22) are proposed to estimate the regression parameters and survival probabilities. The resulting estimators are consistent and asymptotically normal. Robust estimators for the limiting covariance matrices are constructed. Simulation studies demonstrate that the asymptotic approximations are adequate for practical use and that ignoring the intracluster dependence in the variance–covariance estimation would lead to invalid statistical inference. A psychological experiment is provided for illustration.

1. Introduction

Multivariate failure time data arise frequently in biomedical sciences because each study subject may experience more than one type of event or because there exist natural or artificial clusterings of observational units such that failure times within the same cluster tend to be correlated. For example, AIDS researchers are often interested in time to the drop of the CD4-lymphocyte count below a threshold, in time to the first detectable level of HIV antigen, as well as in times to prominent changes of other biological markers for an HIV-infected person. As another example, in an ophthalmologic clinical trial to evaluate the efficacy of a new therapy on delaying the occurrence of severe visual loss, investigators may randomly assign the new therapy to one eye of each patient with the remaining eye serving as an untreated control.

The scientific interest of a multivariate survival study typically lies in the effects of covariates on the risks for failures. It is then natural to model the marginal distributions with proportional hazards models (Cox, 1972) while treating the dependence among multiple failure times as a nuisance. This marginal approach was taken by Wei, Lin, and Weissfeld (1989) and Lee, Wei, and Amato (1992); the former allowed the baseline hazard functions to be different among the marginal models, whereas the latter postulated a common baseline hazard function. Both papers dealt with only *continuous* failure times.

In many applications, failure times are rather broadly grouped. Most commonly, they arise when the (continuous) failure time is subject to interval grouping. For example, in HIV clinical trials, patients may be examined only at, say, 1-month intervals, and the decline of the CD4 count below a cutpoint or the development of a detectable HIV antigen level is known only up to the approximate 1-month gap between the last negative and the first positive follow-up. Similarly, the occurrence of

Key words: Correlated survival data; Cox model; Discrete failure time; Generalized estimating equation; Proportional hazards; Robust inference.

severe visual loss for patients in an ophthalmologic study may only be determined through periodic clinical visits. In other instances, the time measurement may truly be discrete, as, for example, when the time represents the number of attempts required to successfully perform a certain task. A real study of this kind will be seen in Section 4. Occasionally, even if the data are available as essentially exact response times, it may still be attractive to use grouping to simplify computations and data handling, especially when the sample size is exceedingly large.

Lawless (1982, pp. 372–390) gave a detailed account of existing regression methods for analyzing *univariate* grouped survival data. One approach that has met with great success in applications is due to Prentice and Gloeckler (1978). They proposed a grouped-data version of the proportional hazards model and provided the asymptotic results for both the regression coefficient and the survivor function. In the present paper, we extend the work of Prentice and Gloeckler to the multivariate setting. The resulting procedures can be viewed as discrete versions of Wei et al. (1989) and Lee et al. (1992).

The new methodology is described in the next section. Section 3 reports results from our simulation studies. A psychological experiment is provided in Section 4 for illustration. Several discussions follow in Section 5.

2. Multivariate Regression Methods

For definiteness, suppose that there are n units and that each unit can potentially experience M types of failure. (For clustered data, "units" and "types" correspond to clusters and memberships, respectively. Assuming a common M for all "units" entails no loss of generality, as will be discussed in Section 5.) Assume that the marginal distribution for each of the M types of failure satisfies a Cox proportional hazards model. Specifically, the mth hazard function associated with a p-vector of possibly time-dependent (type-specific) covariates Z_m takes the form

$$\lambda_m(t; Z_m) = \lambda_{m0}(t)\exp\{\beta' Z_m(t)\}, \quad m = 1, \dots, M, \tag{2.1}$$

where $\lambda_{m0}(\cdot)$ $(m = 1, \dots, M)$ are arbitrary baseline hazard functions, and β is a p-vector of unknown regression parameters. Note that we allow $\lambda_{m0}(\cdot)$ to be different but take β to be the same among the marginal models. The assumption of a common β vector, however, entails no loss of generality, since this can always be achieved by introducing extra type-specific covariates. The situation of a common baseline hazard function $\lambda_0(\cdot)$ will be discussed later.

Suppose that the time axis is partitioned into $(r + 1)$ intervals $I_j = [a_{j-1}, a_j), j = 1, \dots, r + 1$, with $a_0 = 0$ and $a_{r+1} = \infty$, and that failure times in I_j are recorded as t_j. The covariate vectors are allowed to be time-varying, but fixed within a specific time interval; that is, $Z_m(t) = Z_m(t_j)$ within interval I_j. For notational ease, the dependence of Z_m on t will be suppressed.

For $m = 1, \dots, M$ and $i = 1, \dots, n$, let $(Y_{mi}, \delta_{mi}, Z_{mi})$ represent the possibly censored failure time, failure indicator ($\delta_{mi} = 1$ if Y_{mi} is an uncensored observation and $\delta_{mi} = 0$ otherwise), and covariate vector. Assume that observations are independent between units and that the censoring and failure mechanisms are independent conditional on covariates. Also adopt the convention that a censored failure time t_k means that the unit is known only to have survived to the beginning of the time interval I_k. It is straightforward to modify subsequent results for other conventions.

In this section, we show how to make valid multivariate inference regarding regression parameters and survival probabilities under the working assumption that the M types of failures are independent of one another. Comments will be given in Section 5 on explicitly accounting for dependence. Under the independence working assumption, the contribution to the "likelihood" from the ith unit with respect to the mth type of failure is

$$\{1 - (\xi_{mk})^{\exp(\beta' Z_{mi})}\}^{\delta_{mi}} \prod_{j=1}^{k-1} (\xi_{mj})^{\exp(\beta' Z_{mi})} \tag{2.2}$$

given that $Y_{mi} = t_k$, where

$$\xi_{mj} = \exp\left\{-\int_{a_{j-1}}^{a_j} \lambda_{m0}(u)du\right\}.$$

To remove range restrictions on the parameters, we use the transformations $\gamma_{mj} = \log(-\log\xi_{mj})$ $(j = 1, \dots, r)$, which were introduced by Prentice and Gloeckler (1978). Let $\gamma_m = (\gamma_{m1}, \dots, \gamma_{mr})'$ and $\gamma = (\gamma_1', \dots, \gamma_M')'$. Then the logarithm of (2.2) can be written as

$$l_{mi}(\beta, \gamma) = \delta_{mi}\log[1 - \exp\{-\exp(\gamma_{mk} + \beta' Z_{mi})\}] - \sum_{j=1}^{k-1} \exp(\gamma_{mj} + \beta' Z_{mi}). \tag{2.3}$$

The overall "log-likelihood" is

$$l_{..}(\beta, \gamma) = \sum_{i=1}^{n} \sum_{m=1}^{M} l_{mi}(\beta, \gamma). \tag{2.4}$$

As in the ordinary likelihood analysis, we need to differentiate $l_{..}(\beta, \gamma)$ with respect to β and γ. To this end, the first-order and second-order derivatives of $l_{mi}(\beta, \gamma)$ are listed in the Appendix. Now, let

$$U_{mi}(\beta, \gamma) = \partial l_{mi}(\beta, \gamma)/\partial\beta, \quad W_{mi}(\beta, \gamma) = \partial l_{mi}(\beta, \gamma)/\partial\gamma_m,$$

$$U_{..}(\beta, \gamma) = \sum_{i=1}^{n} \sum_{m=1}^{M} U_{mi}(\beta, \gamma), \quad W_{m.}(\beta, \gamma) = \sum_{i=1}^{n} W_{mi}(\beta, \gamma).$$

Note that $\partial l_{..}(\beta, \gamma)/\partial\beta = U_{..}(\beta, \gamma)$ and $\partial l_{..}(\beta, \gamma)/\partial\gamma_m = W_{m.}(\beta, \gamma)$. Thus we estimate β and γ by solving the following system of "score" equations:

$$\begin{cases} U_{..}(\beta, \gamma) = 0, \\ W_{1.}(\beta, \gamma) = 0, \\ \vdots \\ W_{M.}(\beta, \gamma) = 0. \end{cases} \tag{2.5}$$

The resulting estimators are denoted by $\hat{\beta}$ and $\hat{\gamma}$.

The "observed Fisher information" matrix is

$$F(\beta, \gamma) = \begin{bmatrix} A_{..}(\beta, \gamma) & C_{.1}(\beta, \gamma) & \cdots & C_{.M}(\beta, \gamma) \\ C_{1.}(\beta, \gamma) & B_{11}(\beta, \gamma) & \cdots & B_{1M}(\beta, \gamma) \\ \vdots & \vdots & \vdots & \vdots \\ C_{M.}(\beta, \gamma) & B_{M1}(\beta, \gamma) & \cdots & B_{MM}(\beta, \gamma) \end{bmatrix}, \tag{2.6}$$

where

$$A_{..}(\beta, \gamma) = -\sum_{i=1}^{n} \sum_{m=1}^{M} \partial^2 l_{mi}(\beta, \gamma)/\partial\beta^2,$$

$$B_{mm}(\beta, \gamma) = -\sum_{i=1}^{n} \partial^2 l_{mi}(\beta, \gamma)/\partial\gamma_m^2, \quad B_{ml}(\beta, \gamma) = 0 \quad (m \neq l),$$

$$C_{.m}(\beta, \gamma) = -\sum_{i=1}^{n} \partial^2 l_{mi}(\beta, \gamma)/\partial\beta\partial\gamma_m', \quad C_{m.}(\beta, \gamma) = -\sum_{i=1}^{n} \partial^2 l_{mi}(\beta, \gamma)/\partial\gamma_m\partial\beta'.$$

The derivative matrix (2.6) plays a crucial role in solving estimating equations (2.5) with a Newton–Raphson algorithm.

Asymptotic arguments similar to those presented by White (1982) show that, for large n, $(\hat{\beta}', \hat{\gamma}')'$ is approximately $(p + Mr)$-variate normal with mean $(\beta', \gamma')'$ and covariance matrix $F^{-1}(\hat{\beta}, \hat{\gamma})D(\hat{\beta}, \hat{\gamma})F^{-1}(\hat{\beta}, \hat{\gamma})$, where, suppressing the arguments β and γ,

$$D = \begin{bmatrix} \Sigma_i\Sigma_m\Sigma_l U_{mi}U'_{li} & \Sigma_i\Sigma_m U_{mi}W'_{1i} & \cdots & \Sigma_i\Sigma_m U_{mi}W'_{Mi} \\ \Sigma_i\Sigma_m W_{1i}U'_{mi} & \Sigma_i W_{1i}W'_{1i} & \cdots & \Sigma_i W_{1i}W'_{Mi} \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_i\Sigma_m W_{Mi}U'_{mi} & \Sigma_i W_{Mi}W'_{1i} & \cdots & \Sigma_i W_{Mi}W'_{Mi} \end{bmatrix},$$

with the sums over $i = 1, \ldots, n$; $m = 1, \ldots, M$; and $l = 1, \ldots, M$. It is important to point out that, although the terms "likelihood," "Fisher information," etc. retain their original meanings only when the independence assumption holds, the preceding asymptotic results for $(\hat{\beta}', \hat{\gamma}')'$ are valid regardless of the true dependence structure. By contrast, the naive approach that uses $F^{-1}(\hat{\beta}, \hat{\gamma})$ as the covariance matrix estimator for $(\hat{\beta}', \hat{\gamma}')'$ could result in very misleading statistical inference, as will be demonstrated in Section 3.

If there exists a common baseline hazard function for all the M marginal models, then $\gamma_1 = \gamma_2 = \cdots = \gamma_M$, denoted by γ. (The parameter vector γ is now r-dimensional.) In this setting, the estimators for β and γ, denoted by $\tilde{\beta}$ and $\tilde{\gamma}$, are the roots to the following system of estimating equations:

$$\begin{cases} U_{\cdot\cdot}(\beta, \gamma) = 0, \\ W_{\cdot\cdot}(\beta, \gamma) = 0, \end{cases} \tag{2.7}$$

where $W_{\cdot\cdot}(\beta, \gamma) = \Sigma_{m=1}^{M} W_{m\cdot}(\beta, \gamma)$. Redefine

$$F(\beta, \gamma) = \begin{bmatrix} A_{\cdot\cdot}(\beta, \gamma) & C_{\cdot\cdot}(\beta, \gamma) \\ C_{\cdot\cdot}'(\beta, \gamma) & B_{\cdot\cdot}(\beta, \gamma) \end{bmatrix},$$

where

$$B_{\cdot\cdot}(\beta, \gamma) = -\sum_{i=1}^{n}\sum_{m=1}^{M} \partial^2 l_{mi}(\beta, \gamma)/\partial\gamma^2, \quad C_{\cdot\cdot}(\beta, \gamma) = -\sum_{i=1}^{n}\sum_{m=1}^{M} \partial^2 l_{mi}(\beta, \gamma)/\partial\beta\partial\gamma'.$$

Then, for large n, $(\tilde{\beta}', \tilde{\gamma}')'$ is approximately $(p + r)$-variate normal with mean $(\beta', \gamma')'$ and covariance matrix $F^{-1}(\tilde{\beta}, \tilde{\gamma})D(\tilde{\beta}, \tilde{\gamma})F^{-1}(\tilde{\beta}, \tilde{\gamma})$, where

$$D = \begin{bmatrix} \Sigma_i\Sigma_m\Sigma_l U_{mi}U'_{li} & \Sigma_i\Sigma_m\Sigma_l U_{mi}W'_{li} \\ \Sigma_i\Sigma_m\Sigma_l W_{mi}U'_{li} & \Sigma_i\Sigma_m\Sigma_l W_{mi}W'_{li} \end{bmatrix}.$$

3. Simulation Studies

We have conducted Monte Carlo simulations to examine the finite-sample properties of the robust and naive procedures for making inference about the regression parameter β. Paired failure times were generated from the bivariate exponential family of Gumbel (1960). The distribution functions take the form $F(u, v) = F_1(u)F_2(v)[1 + \theta\{1 - F_1(u)\}\{1 - F_2(v)\}]$, where θ $(-1 \leqslant \theta \leqslant 1)$ measures the dependence between the two failure times in that the correlation, denoted by ρ, equals $\theta/4$. The algorithm of Lawrance and Lewis (1981) was used to generate bivariate exponentials with higher correlations. The two marginal distributions $F_1(\cdot)$ and $F_2(\cdot)$ were both taken to be exponentials with hazard function $e^{\beta Z}$, where Z is standard normal. Since the data were generated with a common baseline hazard function, it was interesting to compare the working model that allows different baseline hazard functions with the one that assumes a common baseline hazard function. We will refer to these two models as Models 1 and 2 in the sequel. We partitioned the time axis at $a_1 = .5$, $a_2 = 1.0$, and $a_3 = 2.0$, and censored the failure times at a_3. The following 18 combinations were considered: $n = 50, 100, 200$; $\rho = 0, .25, .5$; $\beta = 0, .25$. For each combination, 2,000 data sets were generated. All calculations were programmed in C and carried out on a SPARC 2 workstation. The random numbers were generated by the run-time library *drand48* in C.

The results of our simulations are summarized in Table 1. The main conclusions are as follows:

Table 1
Summary statistics for the simulation studies

			Model 1						Model 2					
					Naive		Robust				Naive		Robust	
n	β	ρ	Bias	SSE	SE	CP	SE	CP	Bias	SSE	SE	CP	SE	CP
50	0	.00	.000	.118	.115	.946	.110	.930	.000	.115	.114	.948	.110	.933
		.25	.000	.137	.115	.910	.124	.926	.000	.134	.114	.915	.124	.934
		.50	.001	.149	.115	.884	.136	.931	.001	.146	.115	.889	.136	.932
	.25	.00	.013	.123	.119	.945	.113	.922	.009	.121	.118	.952	.112	.929
		.25	.007	.133	.119	.927	.128	.939	.004	.131	.118	.928	.127	.939
		.50	.013	.151	.119	.890	.140	.929	.009	.148	.118	.889	.139	.932
100	0	.00	.000	.081	.079	.944	.078	.937	.000	.080	.079	.946	.077	.938
		.25	−.002	.090	.079	.920	.087	.938	−.002	.089	.079	.924	.087	.938
		.50	−.005	.100	.079	.885	.095	.934	−.005	.100	.079	.885	.095	.936
	.25	.00	.007	.083	.082	.957	.080	.940	.005	.082	.082	.957	.080	.943
		.25	.008	.093	.082	.919	.090	.935	.007	.092	.082	.920	.090	.936
		.50	.005	.103	.082	.889	.098	.941	.003	.102	.082	.892	.098	.940
200	0	.00	.001	.056	.055	.950	.055	.954	.001	.056	.055	.951	.055	.954
		.25	.002	.062	.055	.926	.062	.947	.002	.062	.055	.924	.061	.947
		.50	.002	.069	.055	.884	.067	.944	.002	.069	.055	.887	.067	.944
	.25	.00	.004	.058	.057	.953	.057	.950	.004	.058	.057	.951	.056	.949
		.25	.006	.064	.057	.921	.063	.944	.005	.064	.057	.921	.063	.944
		.50	.004	.070	.057	.891	.069	.942	.003	.070	.057	.889	.069	.941

Note: Bias and SSE are, respectively, the sampling bias and sampling standard error for the estimator of β. SE and CP are the sampling averages for the standard error estimator and for the coverage probability of the 95% confidence interval, respectively.

(1) The biases of the regression estimators are negligible. The bias and standard error of the regression estimator under Model 2 tend to be smaller than the corresponding quantities under Model 1. (2) The bias of the robust variance estimator is fairly small, and the robust confidence interval has reasonable coverage probability, especially in large samples. (3) For correlated failure times, the naive variance estimator underestimates the true variance and the actual coverage probability of the naive confidence interval falls below the nominal level. The discrepancies are alarming even for moderate correlations.

4. A Real Example

In this section, we illustrate the proposed methods with a psychological experiment that studied children's abilities to locate hidden objects. The study was conducted by Dr David H. Uttal of Northwestern University and was funded by the Spencer Foundation. Each of the 83 children was asked to search for objects hidden in 10 different locations. For each location, the child was given three chances to find the object. The experiments might differ in terms of whether a map was taken while searching for the object and whether the map was rotated. These two factors were expected to affect how quickly the child could find the object. The child's age (which was categorized into two groups, 4–5 years vs. 6–7 years old) was also thought to be predictive.

For this study, $n = 83$, $M = 10$, and $r = 3$. If the ith child was unable to find the mth object in the third trial, then $Y_{mi} = t_4$ and $\delta_{mi} = 0$. The three covariates are coded as follows: $Z_1 = 1$ if the child was 6–7 years old and $Z_1 = 0$ otherwise; $Z_2 = 1$ if the search was done with a map and $Z_2 = 0$ otherwise; $Z_3 = 1$ if the map was rotated and $Z_3 = 0$ otherwise. Because of learning effects, it seems sensible to assume that the success probabilities are different among the 10 types of experiment. For simplicity of illustration, we postulate that the regression effects of age, map use, and map rotation are constant across the experiments.

The key regression results are displayed in Table 2. The robust tests indicate that age and map rotation are highly significant, whereas map use is marginally significant. The naive approach yields much smaller standard error estimates and consequently far more significant results. The observed value of the robust Wald statistic for testing the equality of the 10 baseline hazard functions turns out to be 157.33 on 27 degrees of freedom, which is highly significant (P-value $< .00001$). In response to one referee's suggestion, we also ran an analysis using a common baseline hazard function with 9 indicator covariates to contrast the 10 types of experiment. From this analysis, the parameter estimates for age, map use, and map rotation are .538, .266, and −.874, respectively, the corresponding naive standard error estimates being .094, .088, and .091, and the corresponding robust standard error estimates being .155, .163, and .163.

Table 2
Regression analysis of the psychological experiment data

Variable	Estimate	Naive method SE	Naive method Est./SE	Robust method SE	Robust method Est./SE
Age	.537	.095	5.66	.153	3.51
Map use	.251	.089	2.82	.161	1.56
Map rotation	−.890	.092	−9.69	.161	−5.51

5. Discussion

We have assumed that each study unit can potentially experience M types of failure. If the mth type of failure is missing on unit i, we simply let $Y_{mi} = t_1$ and $\delta_{mi} = 0$. In fact, this method was used implicitly in Section 4 to handle a child whose two test results were unknown. Clearly, such records do not contribute to the estimating equations under our censoring convention. Note that the validity of our analysis requires missing data to be missing completely at random in the sense of Rubin (1976).

As in Liang and Zeger (1986), we may use estimating equations that take the intracluster dependence into account to increase efficiency. This amounts to weighting the contributions from the M types of failure. The resulting estimators remain consistent and asymptotically normal with an estimable covariance matrix under mild regularity conditions on the weight function. We have focused our attention on the independence working assumption for several reasons. First, for generalized linear models without censoring, experience has indicated that the efficiency gains from explicitly accounting for the dependence are small unless the correlations are high and the working correlation structure is close to the true (e.g., Liang and Zeger, 1986, §5). Second, it is not clear what kinds of weight function might lead to efficiency improvements in our case because of the nonlinear nature of model (2.1), although investigations into this issue would be valuable. Lastly, an elaborate working model that contains extra nuisance parameters will complicate the analysis considerably.

As a referee pointed out, it is important to assess the adequacy of the assumed mean function with the two proposed models. One simple method for examining the key proportional hazards assumption is to test for the significance of interaction terms between covariates and t or $\log t$, as was originally suggested by Cox (1972). Various authors, most recently Lin, Wei, and Ying (1993), have exploited the use of martingale-based residuals for checking survival models. It would be useful to adapt their techniques to the current setting.

Acknowledgements

The authors are grateful to the referees for their reviews, to Dr David H. Uttal for the use of his psychological experiment data, and to Dr Thomas Ten Have for helpful discussions. The first author would also like to thank Dr Michael Boehnke for his encouragement throughout the course of this research. This research was supported in part by the National Institutes of Health Grants HG-00209 (for Guo), and R29 GM-47845 and R01 AI-291968 (for Lin).

Résumé

Les données de survie multivariées se rencontrent quand plusieurs types d'événements peuvent advenir à chaque sujet ou quand les observations individuelles sont associées de sorte que les temps de survie dans chaque groupe sont corrélés. Les temps de survie sont souvent sujets à des groupements par intervalles ou prennent des valeurs réellement discrètes. Dans ce papier, la distribution marginale de chaque temps de survie est formulée selon une version du modèle des risques instantanés proportionnels pour données groupées tandis que la structure de dépendance n'est pas spécifiée. Des équations d'estimation généralisées dans l'esprit de Liang et Zeger (1986, *Biometrika* **73,** 13–22) sont proposées pour estimer les paramètres de régression et les probabilités de survie. Les estimateurs qui en résultent sont convergents et asymptotiquement normaux. Des estimateurs robustes des limites des matrices de covariance sont construits. Des études de simulation montrent que les approximations asymptotiques conviennent à une utilisation pratique et que ne pas tenir compte de la dépendance intra-association conduirait à une inférence statistique erronée. Une expérience psychologique est montrée pour illustration.

References

Cox, D. R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association* **55,** 698–707.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: Wiley.

Lawrance, A. J. and Lewis, P. A. W. (1981). A new autoregressive time series model in exponential variables. *Advances in Applied Probability* **13,** 826–845.

Lee, E. W., Wei, L. J., and Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art*, J. P. Klein and P. K. Goel (eds), 237–247. Dordrecht: Kluwer.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80,** 557–572.

Prentice, R. L. and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with applications to breast cancer data. *Biometrics* **34,** 57–67.

Rubin, D. B. (1976). Inference and missing values. *Biometrika* **63,** 81–92.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50,** 1–25.

Received May 1992; *revised February* 1993; *accepted March* 1993.

Appendix

Derivatives of $l_{mi}(\beta, \gamma)$

It is convenient to introduce the notation:

$$h_{mij} = \exp\{\gamma_{mj} + \beta' Z_{mi}(t_j)\}, \quad b_{mij} = h_{mij} e^{-h_{mij}}/(1 - e^{-h_{mij}}),$$

$$d_{mij} = b_{mij}(e^{-h_{mij}} + h_{mij} - 1)/(1 - e^{-h_{mij}}), \quad j = 1, \ldots, r.$$

Then the first-order and second-order derivatives of $l_{mi}(\beta, \gamma)$ are as follows:

$$\partial l_{mi}(\beta, \gamma)/\partial \beta = \left(\delta_{mi} b_{mik} - \sum_{j=1}^{k-1} h_{mij}\right) Z_{mi},$$

$$\partial l_{mi}(\beta, \gamma)/\partial \gamma_{m'j} = \begin{cases} 0, & m \neq m' \\ -h_{mij}, & j < k,\ m = m' \\ \delta_{mi} b_{mij}, & j = k,\ m = m' \\ 0, & j > k,\ m = m' \end{cases},$$

$$\partial l_{mi}(\beta, \gamma)/\partial \gamma_m = \{\partial l_{mi}(\beta, \gamma)/\partial \gamma_{m1}, \ldots, \partial l_{mi}(\beta, \gamma)/\partial \gamma_{mr}\}',$$

$$-\partial^2 l_{mi}(\beta, \gamma)/\partial \beta^2 = \left(\delta_{mi} d_{mik} + \sum_{j=1}^{k-1} h_{mij}\right) Z_{mi} Z'_{mi},$$

$$-\partial^2 l_{mi}(\beta, \gamma)/\partial \gamma_{mj}^2 = \begin{cases} h_{mij}, & j < k \\ \delta_{mi} d_{mij}, & j = k \\ 0, & j > k \end{cases},$$

$$-\partial^2 l_{mi}(\beta, \gamma)/\partial \gamma_{mj} \partial \gamma_{ml} = 0, \quad j \neq l,$$

$$-\partial^2 l_{mi}(\beta,\ \gamma)/\partial\gamma_m^2 = \{-\partial^2 l_{mi}(\beta,\ \gamma)/\partial\gamma_{mj}\partial\gamma_{ml};\ j,\ l = 1,\ \ldots\ ,\ r\}_{r\times r}\ ,$$

$$-\partial^2 l_{mi}(\beta,\ \gamma)/\partial\beta\partial\gamma_{mj} = \{-\partial^2 l_{mi}(\beta,\ \gamma)/\partial\gamma_{mj}^2\}Z_{mi},$$

$$-\partial^2 l_{mi}(\beta,\ \gamma)/\partial\beta\partial\gamma_m' = \{-\partial^2 l_{mi}(\beta,\ \gamma)/\partial\beta\partial\gamma_{m1},\ \ldots\ ,\ -\partial^2 l_{mi}(\beta,\ \gamma)/\partial\beta\partial\gamma_{mr}\}_{p\times r}\ ,$$

$$-\partial^2 l_{mi}(\beta,\ \gamma)/\partial\gamma_m\partial\beta' = -\{\partial^2 l_{mi}(\beta,\ \gamma)/\partial\beta\partial\gamma_m'\}'.$$