(although there is no simple empirical way of checking correctness). However, this is accompanied by at least two major inconveniences. Except in special cases, the GEE corresponds to no statistical model in the accepted sense of the term, i.e. no model that allows us to calculate the probability of the observed or any future data. Thus, no likelihood function is available, singularly complicating the tasks of obtaining useful measures of precision of the point estimates and of comparing "models". Generally only quasi-standard errors and a quasi-score function are available for making inferences and there is considerable debate about the choice of the former. Standard errors are well known to be unreliable in small samples of categorical data.

Although the examples given here have only involved simple binary responses, extensions to more complex **polytomous**, including ordinal, responses (*see* **Ordered Categorical Data**) are available. For more complex models, see, for example, [2], [3], [5], [6], and [9]. The reader may also wish to consult the review paper [10].

References

[1] Agresti, A. (1989). A survey of models for repeated ordered categorical response data, *Statistics in Medicine* **8**, 1209–1224.

[2] Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures, *Biometrika* **81**, 767–775.

[3] Balagtas, C.C., Becker, M.P. & Lang, J.B. (1995). Marginal modelling of categorical data from crossover experiments, *Applied Statistics* **44**, 63–77.

[4] Gilmour, A.R., Anderson, R.D. & Rae, A.L. (1985). The analysis of binomial data by a generalized linear mixed model, *Biometrika* **72**, 593–599.

[5] Glonek, G.F.V. (1996). A class of regression models for multivariate categorical responses, *Biometrika* **83**, 15–28.

[6] Lang, J.B. & Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses, *Journal of the American Statistical Association* **89**, 625–632.

[7] Liang, K.Y., Zeger, S.L. & Qaquish, B. (1992). Multivariate regression for categorical data, *Journal of the Royal Statistical Society, Series B* **54**, 3–40.

[8] Lindsey, J.K. (1993). *Models for Repeated Measurements*. Oxford University Press, Oxford.

[9] Molenberghs, G. & Lesaffre, E. (1994). Marginal modelling of correlated ordinal data using a multivariate Plackett distribution, *Journal of the American Statistical Association* **89**, 633–644.

[10] Pendergast, J.F., Gange, S.J., Newton, M.A., Lindstrom, M.J., Palta, M. & Fisher, M.R. (1996). A survey of methods for analyzing clustered binary response data, *International Statistical Review* **64**, 89–118.

[11] Sheiner, L.B., Beal, S.L. & Sambol, N.C. (1989). Study designs for dose-ranging, *Clinical Pharmacology and Therapeutics* **46**, 63–77.

[12] Zeger, S.L., Liang, K.Y. & Albert, P.S. (1988). Models for longitudinal data: a generalized estimating equation approach, *Biometrics* **44**, 1049–1060.

(*See also* **Binary Data; Correlated Binary Data; McNemar Test**)

J.K. LINDSEY

Marginal Models for Multivariate Survival Data

Multivariate survival or failure-time data arise when each study subject may experience several events or when there exists some natural or artificial grouping of subjects which induces dependence among failure times of the same group. Biomedical examples include the sequence of tumor recurrences or infection episodes, the development of physical symptoms or diseases in several organ systems, the occurrence of blindness in the left and right eyes, the onset of a disease among family members, the initiation of cigarette smoking by classmates, and the appearance of tumor in litter-mates exposed to a carcinogen.

Suppose that there are n independent units each of which can potentially experience K types of failures. Let T_{ik} be the time when the kth type of failure occurs on the ith unit, and let C_{ik} be the corresponding **censoring** time. Define $X_{ik} = \min(T_{ik}, C_{ik})$ and $\Delta_{ik} = I(T_{ik} \le C_{ik})$, where $I(\cdot)$ is the indicator function (*see* **Dummy Variables**). Also, let $\mathbf{Z}_{ik}(\cdot) = [Z_{1ik}(\cdot), \ldots, Z_{pik}(\cdot)]'$ denote a p-vector of possibly **time-dependent covariates** for the ith unit with respect to the kth type of failure. The failure time vector $\mathbf{T}_i = (T_{i1}, \ldots, T_{iK})$ and the censoring time vector $\mathbf{C}_i = (C_{i1}, \ldots, C_{iK})$ are assumed to be independent, conditional on the covariate vector $\mathbf{Z}_i = (\mathbf{Z}'_{i1}, \ldots, \mathbf{Z}'_{iK})$, $i = 1, \ldots, n$. The units are allowed to have unequal numbers of failures, which is achieved by setting C_{ik} to zero whenever T_{ik} is missing.

It is natural and convenient to formulate the marginal distribution for each type of failure with a **proportional hazards model**. Depending on whether the baseline hazard functions are different or identical among the K types of failures, the marginal hazard function for the kth type of failure on the ith unit is

$$\lambda_k(t;\mathbf{Z}_{ik}) = \lambda_{0k}(t)\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(t)], \qquad (1)$$

or

$$\lambda_k(t;\mathbf{Z}_{ik}) = \lambda_0(t)\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(t)], \qquad (2)$$

where $\lambda_{0k}(t)$, $k = 1, \ldots, K$, and $\lambda_0(t)$ are unspecified baseline hazard functions, and $\boldsymbol{\beta}$ is a p-vector of unknown regression parameters. In some applications it is necessary to allow $\lambda_{0k}(t)$, $k = 1, \ldots, K$, to be different, whereas in others it suffices to assume a common baseline hazard function. In both models (1) and (2), we set $\boldsymbol{\beta}$ to be the same among the K submodels, which entails no loss of generality since this structure can always be achieved by introducing appropriate type-specific covariates.

Inference Procedures

If all the failure times were independent, then the **partial likelihood** functions for $\boldsymbol{\beta}$ would be

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\frac{\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(X_{ik})]}{\sum_{j=1}^{n} Y_{jk}(X_{ik})\exp[\boldsymbol{\beta}'\mathbf{Z}_{jk}(X_{ik})]}\right\}^{\Delta_{ik}}$$

under model (1) and

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\frac{\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(X_{ik})]}{\sum_{j=1}^{n}\sum_{l=1}^{K} Y_{jl}(X_{ik})\exp[\boldsymbol{\beta}'\mathbf{Z}_{jl}(X_{ik})]}\right\}^{\Delta_{ik}}$$

under model (2), where $Y_{ik}(t) = I(X_{ik} \geq t)$. The corresponding score functions would be

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\left[\mathbf{Z}_{ik}(X_{ik}) - \frac{\mathbf{S}_k^{(1)}(\boldsymbol{\beta}, X_{ik})}{S_k^{(0)}(\boldsymbol{\beta}, X_{ik})}\right] \qquad (3)$$

and

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\left[\mathbf{Z}_{ik}(X_{ik}) - \frac{\overline{\mathbf{S}}^{(1)}(\boldsymbol{\beta}, X_{ik})}{\overline{S}^{(0)}(\boldsymbol{\beta}, X_{ik})}\right], \qquad (4)$$

where

$$\mathbf{S}_k^{(0)}(\boldsymbol{\beta}, t) = \sum_{j=1}^{n} Y_{jk}(t)\exp\left[\boldsymbol{\beta}'\mathbf{Z}_{jk}(t)\right],$$

$$\mathbf{S}_k^{(1)}(\boldsymbol{\beta}, t) = \sum_{j=1}^{n} Y_{jk}(t)\exp\left[\boldsymbol{\beta}'\mathbf{Z}_{jk}(t)\right]\mathbf{Z}_{jk}(t),$$

$$k = 1, \ldots, K,$$

and

$$\overline{\mathbf{S}}^{(r)}(\boldsymbol{\beta}, t) = \sum_{k=1}^{K}\mathbf{S}_k^{(r)}(\boldsymbol{\beta}, t),\ r = 0, 1.$$

In both cases, the solution to $[\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}]$ is denoted by $\hat{\boldsymbol{\beta}}$.

Although the failure times within the same unit tend to be **correlated**, the estimator $\hat{\boldsymbol{\beta}}$ can be shown to be consistent for $\boldsymbol{\beta}$ and asymptotically p-variate normal provided that the marginal models are correctly specified. However, the conventional **covariance matrix** estimator $\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})$, where $\mathcal{I}(\boldsymbol{\beta}) = -\partial^2 \log L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^2$, is no longer valid, the reason being that $\mathcal{I}(\boldsymbol{\beta})$ is not the covariance matrix of $\mathbf{U}(\boldsymbol{\beta})$ in the presence of intraclass dependence. By approximating $\mathbf{U}(\boldsymbol{\beta})$ with a sum of independent and identically distributed zero-mean random vectors, one can show that, for large n and relatively small K, the random vector $\mathbf{U}(\boldsymbol{\beta})$ is approximately zero-mean normal with covariance matrix estimator

$$\mathbf{V}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{l=1}^{K}\mathbf{W}_{ik}(\hat{\boldsymbol{\beta}})\mathbf{W}_{il}(\hat{\boldsymbol{\beta}})',$$

where

$$\mathbf{W}_{ik}(\boldsymbol{\beta}) = \Delta_{ik}\left[\mathbf{Z}_{ik}(X_{ik}) - \frac{\mathbf{S}_k^{(1)}(\boldsymbol{\beta}, X_{ik})}{S_k^{(0)}(\boldsymbol{\beta}, X_{ik})}\right] - \sum_{j=1}^{n}\frac{\Delta_{jk}Y_{ik}(X_{jk})\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(X_{jk})]}{S_k^{(0)}(\boldsymbol{\beta}, X_{jk})} \times \left[\mathbf{Z}_{ik}(X_{jk}) - \frac{\mathbf{S}_k^{(1)}(\boldsymbol{\beta}, X_{jk})}{S_k^{(0)}(\boldsymbol{\beta}, X_{jk})}\right]$$

and

$$\mathbf{W}_{ik}(\boldsymbol{\beta}) = \Delta_{ik}\left[\mathbf{Z}_{ik}(X_{ik}) - \frac{\bar{\mathbf{S}}^{(1)}(\boldsymbol{\beta}, X_{ik})}{\bar{S}^{(0)}(\boldsymbol{\beta}, X_{ik})}\right] - \sum_{j=1}^{n}\sum_{l=1}^{K}\frac{\Delta_{jl}Y_{ik}(X_{jl})\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(X_{jl})]}{\bar{S}^{(0)}(\boldsymbol{\beta}, X_{jl})} \times \left[\mathbf{Z}_{ik}(X_{jl}) - \frac{\bar{\mathbf{S}}^{(1)}(\boldsymbol{\beta}, X_{jl})}{\bar{S}^{(0)}(\boldsymbol{\beta}, X_{jl})}\right]$$

under models (1) and (2), respectively. Consequently, $\hat{\boldsymbol{\beta}}$ is approximately normal with covariance matrix estimator $\mathbf{D}(\hat{\boldsymbol{\beta}}) = \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{V}(\hat{\boldsymbol{\beta}})\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})$. We call $\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})$ and $\mathbf{D}(\hat{\boldsymbol{\beta}})$ the naive and robust estimators, respectively. In the case of $K = 1$, the matrix $\mathbf{D}(\hat{\boldsymbol{\beta}})$ reduces to the Lin–Wei [14] robust covariance matrix estimator for the maximum partial likelihood estimator under **misspecified** proportional hazards models. To test the global hypothesis that $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, one may use the chi-square statistic $\mathbf{U}'(\boldsymbol{\beta}_0)\mathbf{V}^{-1}(\boldsymbol{\beta}_0)\mathbf{U}(\boldsymbol{\beta}_0)$ or $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{D}^{-1}(\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$; to test the general linear hypothesis H_0: $\mathbf{L}\boldsymbol{\beta} = \mathbf{d}$, where $\mathbf{L}$ is an $r \times p$ matrix of constants and $\mathbf{d}$ is an $r \times 1$ vector of constants, one refers $(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{d})'\{\mathbf{L}\mathbf{D}(\hat{\boldsymbol{\beta}})\mathbf{L}'\}^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{d})$ to the **chi-square distribution** with r **degrees of freedom**.

The above results are analogous to those of the **generalized estimation equations** (GEE) for the analysis of **marginal models** for **longitudinal data** with an independence working assumption. A similar idea can be used to estimate the cumulative baseline hazard functions $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(u)\mathrm{d}u$, $k = 1, \ldots, K$, and $\Lambda_0(t) = \int_0^t \lambda_0(u)\mathrm{d}u$ for models (1) and (2). Specifically, under the independence working assumption, the Aalen–Breslow type estimators for $\Lambda_{0k}(t)$ and $\Lambda_0(t)$ are

$$\hat{\Lambda}_{0k}(t) = \sum_{i=1}^{n}\frac{I(X_{ik} \le t)\Delta_{ik}}{S_k^{(0)}(\hat{\boldsymbol{\beta}}, X_{ik})}, \quad k = 1, \ldots, K, \qquad (5)$$

and

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n}\sum_{k=1}^{K}\frac{I(X_{ik} \le t)\Delta_{ik}}{\bar{S}^{(0)}(\hat{\boldsymbol{\beta}}, X_{ik})}. \qquad (6)$$

These estimators are consistent and asymptotically normal. In fact, the p-vector of random processes,

$$n^{1/2}[\hat{\Lambda}_{01}(t) - \Lambda_{01}(t), \ldots, \hat{\Lambda}_{0K}(t) - \Lambda_{0K}(t)]',$$

converges weakly to a p-dimensional zero-mean Gaussian random field, and the covariance between $\hat{\Lambda}_{0k}(t)$ and $\hat{\Lambda}_{0l}(s)$ can be estimated by $\sum_{i=1}^n \xi_{ik}(t;\hat{\boldsymbol{\beta}})\xi_{il}(s;\hat{\boldsymbol{\beta}})$, where

$$\xi_{ik}(t;\boldsymbol{\beta}) = \frac{I(X_{ik} \le t)\Delta_{ik}}{S_k^{(0)}(\boldsymbol{\beta}, X_{ik})} - \sum_{j=1}^{n}\frac{I(X_{jk} \le t)\Delta_{jk}Y_{ik}(X_{jk})\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(X_{jk})]}{S_k^{(0)}(\boldsymbol{\beta}, X_{jk})^2} - \left[\sum_{j=1}^{n}\frac{I(X_{jk} \le t)\Delta_{jk}S_k^{(1)}(\boldsymbol{\beta}, X_{jk})}{S_k^{(0)}(\boldsymbol{\beta}, X_{jk})^2}\right]' \times \mathcal{I}^{-1}(\boldsymbol{\beta})\sum_{l=1}^{K}\mathbf{W}_{il}(\boldsymbol{\beta}).$$

In addition, $n^{1/2}[\hat{\Lambda}_0(t) - \Lambda_0(t)]$ converges weakly to a zero-mean Gaussian process, and the covariance between $\hat{\Lambda}_0(t)$ and $\hat{\Lambda}_0(s)$ can be estimated by $\sum_{i=1}^n\sum_{k=1}^K\sum_{l=1}^K \xi_{ik}(t;\hat{\boldsymbol{\beta}})\xi_{il}(s;\hat{\boldsymbol{\beta}})$, where

$$\xi_{ik}(t;\boldsymbol{\beta}) = \frac{I(X_{ik} \le t)\Delta_{ik}}{\bar{S}^{(0)}(\boldsymbol{\beta}, X_{ik})} - \sum_{j=1}^{n}\sum_{l=1}^{K}\frac{I(X_{jl} \le t)\Delta_{jl}Y_{ik}(X_{jl})\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(X_{jl})]}{\bar{S}^{(0)}(\boldsymbol{\beta}, X_{jl})^2} - \left[\sum_{j=1}^{n}\sum_{l=1}^{K}\frac{I(X_{jl} \le t)\Delta_{jl}\bar{\mathbf{S}}^{(1)}(\boldsymbol{\beta}, X_{jl})}{\bar{S}^{(0)}(\boldsymbol{\beta}, X_{jl})^2}\right]' \times \mathcal{I}^{-1}(\boldsymbol{\beta})\mathbf{W}_{ik}(\boldsymbol{\beta}).$$

The large-sample properties for the corresponding baseline survival function estimators $\exp[-\hat{\Lambda}_{0k}(t)]$, $k = 1, \ldots, K$, and $\exp[-\hat{\Lambda}_0(t)]$ follow from the **delta method**. Furthermore, simple modifications can be made to estimate the survival functions associated with specific covariate values.

Software Availability

The estimators $\hat{\boldsymbol{\beta}}$, $\hat{\Lambda}_{0k}$, $k = 1, \ldots, K$, and $\hat{\Lambda}_0$ are constructed under the independence working assumption, and therefore can be obtained from any existing software for the **Cox regression**. The robust covariance matrix estimator for $\hat{\boldsymbol{\beta}}$ is available in **S-PLUS**, SAS, and STATA packages, as well as in a special FORTRAN program [12]. The robust variance–covariance estimators for $\hat{\Lambda}_{0k}, k = 1, \ldots, K$, and $\hat{\Lambda}_0$ have not been implemented in commercially available software packages.

An Example

We now provide an illustration with the well-known Diabetic Retinopathy Study [4], which was conducted by the National Eye Institute to evaluate the effectiveness of laser photocoagulation in delaying the onset of blindness in patients with diabetic retinopathy. The study enrolled 1742 patients. One eye of each patient was randomly selected for photocoagulation and the other eye was observed without treatment. The patients were followed over several years for the occurrence of blindness in their left and right eyes.

We confine our attention to a subset of the data with 197 high-risk patients previously analyzed by Huster et al. [7] and Lin [13]. By the end of the study, 54 treated eyes and 101 control eyes in this subsample had developed blindness. In this example, each patient could potentially experience blindness in both eyes; therefore, there are two failure types with $k = 1$ and 2, denoting the left and right eyes, respectively. Since there are no biological differences between the left and right eyes, it is natural to assume a common baseline hazard function for the two failure types.

As mentioned above, the primary objective of this study was to assess whether laser photocoagulation delays the occurrence of blindness. Because juvenile and adult diabetes have very different courses, it is desirable to examine how the age at onset of diabetes may affect the time to blindness. Thus, we consider model (2) with $\mathbf{Z}_{ik} = (Z_{1ik}, Z_{2ik}, Z_{3ik})', i = 1, \ldots, 197; k = 1, 2$, where

$$Z_{1ik} = \begin{cases} 1, & \text{if the } k\text{th eye of the } i\text{th patient was on treatment,} \\ 0, & \text{otherwise;} \end{cases}$$

$$Z_{2ik} = \begin{cases} 1, & \text{if the } i\text{th patient had adult onset diabetes,} \\ 0, & \text{if the } i\text{th patient had juvenile onset diabetes;} \end{cases}$$

Table 1

Variable	Parameter estimate	Stand. error estimate	
		Naive	Robust
Treatment (Z_1)	−0.425	0.218	0.185
Diabetic type (Z_2)	0.341	0.199	0.196
Interaction ($Z_1 \times Z_2$)	−0.846	0.351	0.304

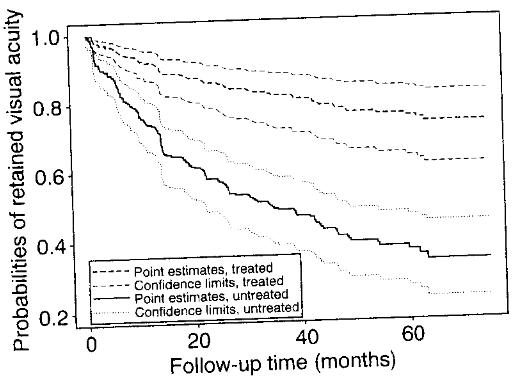

Figure 1 Estimates and pointwise 95% confidence intervals for survival functions

and $Z_{3ik} = Z_{1ik} \times Z_{2ik}$. The results for the estimation of the regression parameters are shown in Table 1. The robust standard error estimates are appreciably smaller than the naive estimates, the latter ignoring the dependence between the left and right eyes. The treatment appears to be effective, and this effect is much stronger for adult-onset diabetes than for juvenile-onset diabetes.

Figure 1 displays the estimates and pointwise 95% confidence intervals for the survival functions, namely, the probabilities of retained visual acuity, for adult-onset diabetes, separated by treatment groups. As expected, these probabilities are much higher for the treated eyes than for the untreated ones.

Further Results

The estimation of $\boldsymbol{\beta}$ under models (1) and (2) was first studied by Wei et al. [23] and Lee et al. [9], respectively, and further developed by Lin [13], while the estimation of $\Lambda_{0k}, k = 1, \ldots, K$, and Λ_0 was investigated by Spiekerman & Lin [22]. The latter authors established a rigorous asymptotic theory for the estimation of both the regression parameters and baseline hazard functions under a general marginal model which allows M, $1 \leq M \leq K$, different baseline hazard functions among the K types of failures. In a separate paper, they [21] developed a class of graphical and numerical techniques for checking the adequacy of models (1) and (2). The readers are referred to the aforementioned papers for further theoretical details as well as additional numerical examples. Incidentally, Huster et al. [7] studied model (2) with a parametric baseline hazard function, while Guo & Lin [6] deal with discrete-time versions of models (1) and (2).

Liang et al. [11] proposed a different procedure for analyzing model (2). Their estimating function is similar to (4), but they replaced $\overline{\mathbf{S}}^{(1)}/\overline{S}^{(0)}$ by an analog which exploits pairwise comparisons of independent observations. The actual form of their **estimating function** is

$$\sum_{i=1}^{n}\sum_{k=1}^{K} I[n_i(X_{ik}) > 0]\Delta_{ik}$$

$$\times \left[\mathbf{Z}_{ik}(X_{ik}) - n_i^{-1}(X_{ik})\sum_{j\neq i}\sum_{l}\mathbf{e}_{ik,jl}(\boldsymbol{\beta}, X_{ik})\right],$$

where $n_i(t) = \sum_{j\neq i}\sum_l Y_{jl}(t)$ and

$$\mathbf{e}_{ik,jl}(\boldsymbol{\beta}, t) = \frac{Y_{ik}(t)\mathbf{Z}_{ik}(t)\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(t)] + Y_{jl}(t)\mathbf{Z}_{jl}(t)\exp[\boldsymbol{\beta}'\mathbf{Z}_{jl}(t)]}{Y_{ik}(t)\exp[\boldsymbol{\beta}'\mathbf{Z}_{ik}(t)] + Y_{jl}(t)\exp[\boldsymbol{\beta}'\mathbf{Z}_{jl}(t)]}.$$

The resultant estimator is **consistent** and asymptotically normal. The relative efficiency of $\hat{\boldsymbol{\beta}}$ vs. the Liang et al. estimator has not been investigated.

Estimating functions (3) and (4) were derived under the independence working assumption. As in the case of **longitudinal data**, it may be more efficient to use estimating functions that take into account the nature of dependence explicitly. This amounts to incorporating certain weight functions into estimating functions (3) and (4). The resultant estimators remain consistent and asymptotically normal with a sandwich-type variance estimator under mild regularity conditions on the weight function. Due to censoring and the nonlinear nature of the proportional hazards model, it is difficult to construct optimal weight functions. Cai & Prentice [2] investigated a weight function that is the inverse of the covariance matrix of the marginal martingales associated with the T_{ik}s. Their theoretical calculations and simulation studies indicated that the efficiency gains in using such weighted estimating functions over estimating functions (3) and (4) are small unless the correlations of failure times are unusually high.

There has been considerable research on semiparametric multivariate failure time distributions which characterize the strength of association among failure time components by a limited number of parameters while leaving the forms of the marginal distributions unspecified (e.g. [3], [18], [1]). One may extend these multivariate distributions by formulating their marginal distributions with model (1) or (2). One may then estimate the marginal regression parameters and baseline hazard functions by (3) and (5) or (4) and (6) and proceed to estimate the association parameters by the pseudo-**maximum likelihood** method [5]. This approach was mentioned by Bandeen-Roche & Liang [1], but its inferential properties have yet to be investigated.

Prentice & Hsu [19] studied simultaneous regression on the marginal hazard ratios and pairwise dependencies, which is analogous to the regression on the means and covariances of noncensored multivariate responses [20]. They used the estimating function of Cai & Prentice [2] for the marginal hazard ratio parameters and developed a similar *ad hoc* estimating function for the dependence parameters. They showed that the solutions to this pair of estimating functions are consistent and asymptotically normal, with a sandwich-type covariance matrix estimator.

The **accelerated failure-time** and **additive hazards models** are two important alternatives to the proportional hazards model. The former relates the logarithm of the failure time linearly to the covariates [8], while the latter relates the conditional hazard function linearly to the covariates [16]. One may formulate the marginal distributions of multivariate failure time data with accelerated failure time models or additive hazards models rather than proportional hazards models. The corresponding inference procedures were studied, respectively, by Lin & Wei [15] and Lee et al. [10], and by Lin & Ying [17].

References

[1] Bandeen-Roche, K.J. & Liang, K.-Y. (1996). Modelling failure time associations in data with multiple levels of clustering, *Biometrika* **83**, 29–39.

[2] Cai, J. & Prentice, R.L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data, *Biometrika* **82**, 151-164.
[3] Clayton, D.G. (1978). A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika* **65**, 141-151.
[4] Diabetic Retinopathy Study Research Group (1981). Diabetic retinopathy study, *Investigative Ophthalmology and Visual Science* **21**, 149-226.
[5] Gong, G. & Samaniego, F.J. (1981). Pseudo maximum likelihood, *Annals of Statistics* **9**, 861-869.
[6] Guo, S.W. & Lin, D.Y. (1994). Regression analysis of multivariate grouped survival data, *Biometrics* **50**, 632-639.
[7] Huster, W.J., Brookmeyer, R. & Self, S.G. (1989). Modelling paired survival data with covariates, *Biometrics* **45**, 145-156.
[8] Kalbfleisch, J.D. & Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.
[9] Lee, E.W., Wei, L.J. & 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 & P.K. Goel, eds. Kluwer, Dordrecht, pp. 237-247.
[10] Lee, E.W., Wei, L.J. & Ying, Z. (1993). Linear regression analysis for highly stratified failure time data, *Journal of the American Statistical Association* **88**, 557-565.
[11] Liang, K.-Y., Self, S.G. & Chang, Y.-C. (1993). Modeling marginal hazards in multivariate failure-time data, *Journal of the Royal Statistical Society, Series B* **55**, 441-453.
[12] Lin, D.Y. (1993). MULCOX2: a general computer program for the Cox regression analysis of multivariate failure time data, *Computer Methods and Programs in Biomedicine* **40**, 279-293.
[13] Lin, D.Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach, *Statistics in Medicine* **13**, 2233-2247.
[14] Lin, D.Y. & Wei, L.J. (1989). The robust inference for the Cox proportional hazards model, *Journal of the American Statistical Association* **84**, 1074-1078.
[15] Lin, J.S. & Wei, L.J. (1992). Linear regression analysis for multivariate failure time observations, *Journal of the American Statistical Association* **87**, 1071-1097.
[16] Lin, D.Y. & Ying, Z. (1994). Semiparametric analysis of the additive risk model, *Biometrika* **81**, 61-71.
[17] Lin, D.Y. & Ying, Z. (1997). Additive hazards regression models for survival data, in *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, D.Y. Lin & T.R. Fleming, eds. Springer-Verlag, New York, pp. 185-198.
[18] Oakes, D. (1989). Bivariate survival models induced by frailties, *Journal of the American Statistical Association* **84**, 487-493.
[19] Prentice, R.L. & Hsu, L. (1997). Regression on hazard ratios and cross-ratios in multivariate failure time analysis, *Biometrika*, **84**, 349-363.
[20] Prentice, R.L. & Zhao, L.P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses, *Biometrics* **47**, 825-839.
[21] Spiekerman, C.F. & Lin, D.Y. (1996). Checking the marginal Cox model for correlated failure time data, *Biometrika* **83**, 143-156.
[22] Spiekerman, C.F. & Lin, D.Y. (1996). Marginal Regression Models for Multivariate Failure Time Data, *Technical Report 144*. Department of Biostatistics, University of Washington.
[23] Wei, L.J., Lin, D.Y. & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, *Journal of the American Statistical Association* **84**, 1065-1073.

D.Y. LIN

Marginal Probability

In many situations, interest focuses on probability distributions for multiple random variables. For example, one may be studying height, weight, blood pressure, and cholesterol levels in a population, variables likely to be **correlated** with one another. Knowledge of the joint distribution of these variables allows one to calculate the probabilities associated with any particular outcome of interest. Marginal probabilities relate to the univariate distribution, or marginal distribution, associated with any of the variables under consideration.

To fix notation, first consider a **bivariate** model for two random variables X and Y. Let $f_{X,Y}(x, y)$ denote the joint probability mass function if X and Y are discrete or the joint probability density function if X and Y are continuous. If X and Y are discrete, the marginal probability mass functions of X and Y are given by

$$f_X(x) = \sum_y f_{X,Y}(x, y)$$

and

$$f_Y(y) = \sum_x f_{X,Y}(x, y),$$

where the summations are taken over all of the values of Y or X. In this case, the joint probability mass function can be written in tabular form, with the columns corresponding to the possible values