Semiparametric likelihood estimation in the Clayton-Oakes failure time model

BY D. V. GLIDDEN

University of California, San Francisco

AND

S. G. SELF

Fred Hutchinson Cancer Research Center

Running Title: Estimation in the Clayton-Oakes model

ABSTRACT. Multivariate failure time data arise when the sample consists of clusters and each cluster contains several possibly dependent failure times. The Clayton-Oakes model (Clayton, 1978; Oakes, 1982) for multivariate failure times characterizes the intracluster dependence parametrically but allows arbitrary specification of the the marginal distributions. In this paper, we discuss estimation in the Clayton-Oakes model when the marginal distributions are modeled to follow the Cox (1972) proportional hazards regression model. Parameter estimation is based on an approximate generalized maximum likelihood estimator. We illustrate the model's application with example datasets.

Some key words: frailty model, multivariate failure time, nonparametric maximum likelihood

1. Introduction

Multivariate failure time (MVFT) data arise when the sample consists of clusters and each cluster contains several possibly dependent failure times. Such data are routinely encountered in genetic epidemiology studies, ophthalmological studies and litter-matched carcinogenesis experiments; see Lin (1994) for various biomedical examples. The analysis of such data must account for the potential dependence of failure times within the same cluster.

One approach to modeling the intracluster dependence has been through specification of parametric copula models. These models express the joint distribution of failure times as a function of the marginal distributions and a dependence parameter. One such model is the Clayton-Oakes (CO) model (Clayton, 1978; Oakes, 1982). Suppose a cluster consists of K failure times $T_1, \ldots, T_K$ with marginal hazard functions

$$\lambda_k(t) = \lim_{h \downarrow 0} h^{-1} \mathrm{pr}(t \leq T_k < t + h | T_k \geq t)$$

and $\Lambda_k(t) = \int_0^t \lambda_k(u)du$. The CO model expresses the joint survival function of $(T_1, \ldots, T_K)$ as

$$\mathrm{pr}(T_1 > t_1, \ldots, T_K > t_K; \theta_0) = \left[\sum_{k=1}^{K} \exp\{\theta_0 \Lambda_k(t_k)\} - K + 1 \right]^{-\theta_0^{-1}}, \quad \theta_0 \in (0, \infty). \quad (1)$$

The model is parametric with respect to θ_0, which indexes intracluster dependence. When θ_0 equals 0, the joint survival function is the product of the marginal survival functions and failures are independent. The model is parametric with respect to θ_0 which describes the dependence between failures, but places no restriction of the marginal hazard functions.

In the bivariate case the parameter θ_0 can be expressed as the ratio of conditional hazard functions. In the bivariate case, i.e. $K = 2$, the hazard functions at time t_1 for the conditional distributions of T_1 given $T_2 = t_2$ and given $T_2 \geq t_2$ have ratio $1+\theta_0$, not depending on t_1 or t_2, and the same relation holds between the conditional distributions of T_2 given $T_1 = t_1$ and $T_1 \geq t_1$. The parameter $1 + \theta_0$ describes the modification in the hazard of one variable induced by failure of the other. When $\theta_0 = 0$, the failure of one variable does not affect the hazard of the other. Increasing values of θ_0 imply increasing positive association between failures.

This paper will consider estimation in a CO model when the marginal hazards obey a Cox (1972) regression model. Previous work in marginal regression models for MVFT data has emphasized a Cox regression specification without an explicit model for dependence. Other models, such as frailty models, have considered parametric models for dependence. These models have assumed that, conditional on a random effect, the regression effects are proportional. The marginal Cox specification within a CO model for dependence combines marginal hazard ratio parameters for covariate effects with a single, interpretable parameter for intracluster dependence.

A reparameterization of the CO model will permit us to approach the problem in the spirit of recent work in gamma frailty models (Nielsen *et al.*, 1992; Klein, 1992). In that work parameter estimation is based on constructing and maximizing a semi-parametric likelihood. This paper fits a marginal CO model into the framework of Nielsen *et al.*, (1992). An approximate fitting algorithm is developed in Section 3, and the properties of the estimators are explored in a simulation study in Section 4. Finally, the methods are applied to example datasets in Section 5.

2. The Model

2.1 Data and Notation

Suppose there are n independent clusters. Let T_{ik} and C_{ik} be the event and censoring times for the kth subject in the ith cluster, and let Z_{ik} be a p-vector of covariates. Let $T_i = (T_{ik}, k = 1, \ldots, K)$ with C_i and Z_i defined similarly. Assume that conditional on Z_i, T_i and C_i are independent and $(T_i, C_i), (i = 1, \ldots, n)$ are *i.i.d.* random vectors. Define $X_{ik} = \min(T_{ik}, C_{ik})$ and $\delta_{ik} = I(T_{ik} \le C_{ik})$ where $I(\cdot)$ is the indicator function. The maximum cluster size, K, is fixed but clusters may have different sizes by setting C_{ik} to 0 when T_{ik} is missing.

In the counting process notation, $Y_{ik}(t) = I(X_{ik} \ge t)$ and $N_{ik}(t) = \delta_{ik} I(X_{ik} \le t)$, $t \in [0, \tau]$. The variable τ denotes the maximum follow-up time. The marginal hazard function, $\lambda_{ik}(\cdot)$ for T_{ik} follows

$$\lambda_{ik}(t|Z) = \lim_{h \downarrow 0} h^{-1} \mathrm{pr}(t \le T_{ik} < t + h | T_{ik} \ge t, Z_{ik}) = \lambda_0(t) \exp(\beta_0' Z_{ik}). \tag{2}$$

The joint distributions of failures in a cluster follow the CO model with joint survival function given in (1).

2.2 The Clayton-Oakes Model

The Clayton-Oakes model (1) can also be derived as a "gamma frailty" model (Clayton, 1978; Oakes, 1982). Suppose for each cluster there is an associated frailty variable ξ shared by members of a cluster. A frailty model (Vaupel *et al.*, 1979) postulates that conditional on ξ, failures in a cluster are independent with hazard functions:

$$\lim_{h \downarrow 0} h^{-1} \mathrm{pr}(t \le T_{ik} < t + h | T_{ik} \ge t, \xi_i) = \xi_i \alpha_{ik}(t), \tag{3}$$

where $\alpha_{ik}(\cdot)$ and $A_{ik}(t) = \int_0^t \alpha_{ik}(s) ds$ are called the "basic" hazard functions. Under the gamma frailty model, ξ has a gamma distribution with mean one and variance θ_0.

Clayton (1978) showed that a gamma frailty model (3) has a joint survival function which can be written in the form of (1) with the marginal hazard function $\lambda_{ik}(\cdot)$ and basic hazard function $\alpha_{ik}(\cdot)$ satisfying:

$$\alpha_{ik}(t) = \exp\{\theta_0 \Lambda_{ik}(t-)\} \lambda_{ik}(t), \tag{4}$$

where $\Lambda_{ik}(t) = \int_0^t \lambda_{ik}(s)ds$. This implies a model with the form (1) also has a representation in the form (3). However, the relationship between basic and marginal hazards functions in (4) is awkward.

Nielsen *et al.* (1992) used generalized maximum likelihood estimation for models of the form (3) with gamma frailties for a wide class of specifications for $\alpha(\cdot)$. This paper will approach estimation by adapting the techniques of Nielsen to the CO model with marginal hazards which follow (2). The model is equivalent to a gamma frailty model with basic hazard function:

$$\alpha_{ik}(t) = \exp\{\beta_0' Z_{ik} + \theta_0 e^{\beta_0' Z_{ik}} \Lambda_0(t-)\}\lambda_0(t), \qquad (5)$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$.

3. Parameter Estimation

Estimation for gamma frailty models of the form (3) with basic hazards which follow a variety of non- and semi-parametric models have been considered by Nielsen *et al.* (1992) and Klein (1992). Their approach based estimation for θ_0 and the parameters which govern $\alpha(\cdot)$ on generalized maximum likelihood estimation. Generalized maximum likelihood (Gill, 1989) attempts to extend the principal of maximum likelihood to semi-parametric models. Many non-parametric estimators in survival models (cf. Kaplan-Meier and Nelson-Aalen estimators) developed by other principals can also be derived as generalized maximum likelihood estimators (GMLEs). The method frequently requires that the parameter space is 'extended', so the likelihood estimator $\hat{H}_n(\cdot)$ for the smooth function $H(\cdot)$ is often a step-function. The number of discontinuities is often on the order of the number of observations, so parameter estimation may require high-dimensional optimization. Nielsen *et al.* and Klein adapted GML estimation to produce semi-parametric estimates for the frailty models.

3.1 Likelihood Construction

The approach of Nielsen *et al.* and Klein parallels maximum likelihood for a missing data problem. The "complete data" is observation of $\mathcal{G}_t = \{\xi_i, N_{ik}(s), Y_{ik}(s+), Z_{ik}, 0 \le s \le t,\ i = 1, \ldots, n,\ k = 1, \ldots, K\}$, a filtration which includes the latent frailties and the history of event to time t. The "incomplete data" is observation of $\mathcal{F}_t = \{N_{ik}(s), Y_{ik}(s+), Z_{ik}, 0 \le s \le t,\ i = 1, \ldots, n,\ k = 1, \ldots, K\}$, a filtration which includes the history of events to time t. The likelihood for the "incomplete data" is obtained from the "complete data" likelihood by margining over the distribution of the missing data. The (partial) likelihood for the "complete data" is proportional to

$$L_{\mathcal{G}} = \prod_{i=1}^{n}\prod_{k=1}^{K} f(\xi_i;\theta_0)\exp\{-\xi_i\int_0^\tau Y_{ik}dA_{ik}\}\prod_{t\in[0,\tau]}(\xi_i Y_{ik}(t)\alpha_{ik}(t))^{\Delta N_{ik}(t)}, \tag{6}$$

where f is the density for a gamma random variable with mean one and variance θ_0. Under regularity conditions given by Gill (1992), integrating over the frailties gives the (partial) likelihood for the "incomplete data"

$$L_{\mathcal{F}} = \prod_{i=1}^{n}\left\{\frac{\theta_0^{-\theta_0^{-1}}}{\Gamma(\theta_0^{-1})}\frac{\Gamma(\theta_0^{-1}+N_{i\cdot}(\tau))}{(\theta_0^{-1}+\sum_{k=1}^{K}\int_0^\tau Y_{ik}dA_{ik})^{\theta_0^{-1}+N_{i\cdot}(\tau)}}\prod_{k=1}^{K}\prod_{t\in[0,\tau]}(Y_{ik}(t)\alpha_{ik}(t))^{\Delta N_{ik}(t)}\right\}, \tag{7}$$

where $\cdot$ in the subscript indicates a sum over the corresponding index and $\Gamma(\cdot)$ denotes the gamma function. Nielsen *et al.* used (7) as the basis for GML estimation of θ_0 and the parameters which govern $\alpha(\cdot)$. If the model for the basic hazards is fully parametric, the likelihood may be maximized directly. If the model for $\alpha(\cdot)$ is semi- or non-parametric, maximization of (7) can be computationally prohibitive.

3.2 The algorithm of Nielsen, *et al.*

For many familiar non- and semi-parametric models, however, GMLEs for the parameters based on the "complete data" are computationally simple. This structure makes estimation approaches based on the EM algorithm appealing. This strategy was first applied to frailty models by Clayton & Cuzick (1985). Nielsen, *et al.*, using counting process notation, were able to develop a general expression for the expected value of the latent frailties given the observed data:

$$E(\xi_i|(\mathcal{F}_t)) = \frac{1+\theta_0 N_{i.}(t)}{1+\theta_0 \sum_{k=1}^{K} \int_0^t Y_{ik} dA_{ik}}. \tag{8}$$

In the EM-algorithm, this forms the basis for the E-step. The M-step involves repeated maximization of the "complete data" likelihood (6). The EM approach is very general and for many semi-parametric specifications of $\alpha(\cdot)$, maximization of (6) is considerably simpler than the direct maximization of (7). An example of the relative simplicity of the EM approach is a Cox model specification for the basic hazard:

$$\alpha_{ik}(t) = \exp\{\beta_c' Z_{ik}\}\lambda_c(t), \tag{9}$$

where $\Lambda_c(t) = \int_0^t \lambda_c(s)ds$. For this model, Nielsen, *et al.* used a double iterative algorithm. At the first level, θ is considered fixed at a value $\tilde{\theta}$ and an EM-algorithm is iterated to convergence to maximize (6) with $\alpha(t)$ given in (9). This finds the GMLEs $\hat{\beta}_c|\tilde{\theta}$ and $\hat{\Lambda}_c(\cdot)|\tilde{\theta}$ and gives the value of the profile likelihood for θ at $\tilde{\theta}$. At the second level, repeated evaluations of the profile likelihood are combined with the golden section search to maximize the profile likelihood of θ. The maximizer of the profile likelihood, $\hat{\theta}$, and the values $\hat{\Lambda}_c(\cdot)|\hat{\theta}$ and $\hat{\beta}_c|\hat{\theta}$ maximize (7). The method requires repeated evaluation of the profile likelihood for θ and each evaluation requires iteration of an EM-algorithm to convergence.

For the model (9), the algorithm is relatively simple. Direct maximization of (7) would require maximization over $1+p+N_{..}(\tau)$ dimensions. The EM-algorithm, however, requires only the iterative maximization over p dimensions. The structure is apparent from the EM-algorithm of Nielsen *et al.* given below.

The algorithm fixes the value of θ at $\tilde\theta$ and requires initial values $\hat\beta_c^{(0)}$ and $\hat\Lambda_c^{(0)}(\cdot)$. The $l+1$th iteration of the algorithm takes the form:

C1: Calculation of Basic Intensity.

$$\hat{A}_{ik}^{(l)}(t) = \int_0^t \exp\{\hat\beta_c^{(l)\prime} Z_{ik}\} d\hat\Lambda_c^{(l)}(s),$$

E1: Posterior Expectation of ξ_i.

$$\hat{\xi}_i^{(l+1)} = (1+\tilde\theta N_{i.}(\tau))/(1+\tilde\theta \sum_{k=1}^K \int_0^\tau Y_{ik}(s) d\hat{A}_{ik}^{(l)}(s)),$$

M1: 'M'-step for β_c.

$$\hat\beta_c^{(l+1)} = \arg\max_{\beta_c} \prod_{i=1}^n \prod_{k=1}^K \prod_{t\in[0,\tau]} \left(\frac{\hat\xi_i^{(l+1)} \exp\{\beta_c' Z_{ik}\}}{\sum_{j=1}^n \sum_{m=1}^K \hat\xi_j^{(l+1)} \exp\{\beta_c' Z_{jm}\} Y_{jm}(t)} \right)^{\Delta N_{ik}(t)},$$

M2: 'M'-step for $\Lambda_c(\cdot)$ given $\hat\beta^{(l+1)}$ and $\hat\xi^{(l+1)}$. $\hat\Lambda_c^{(l+1)}(t) = F(t;\hat\beta_c^{(l+1)}, \hat\xi^{(l+1)})$ where

$$F(t;\beta_c,\xi) = \int_0^t \left[\sum_{i=1}^n \sum_{k=1}^K \xi_i \exp\{\beta_c' Z_{ik}\} Y_{ik}(s) \right]^{-1} dN_{..}(s).$$

Steps **C1,E1,M1** and **M2** are repeated until convergence ($l = 0, \cdots, L$). The values of $\hat\beta_c^{(l)}$, $\hat\Lambda_c^{(l)}(\cdot)$ at convergence are denoted as $\hat\beta_c|\tilde\theta$ and $\hat\Lambda_c(\cdot|\tilde\theta)$.

The M-step requires maximization in $p + N_{..}(\tau)$ dimensions. However, by maximizing in two steps **M1** and **M2**, this is reduced to iterative maximization in p dimensions. The GMLE of $\hat\Lambda_c(\cdot)$ is $F(t;\beta,\xi)$ which has a closed-form expression. Substituting $F(t;\beta,\xi)$ for $\Lambda_c(\cdot)$ in the likelihood (6) gives the profile likelihood for β which is maximized in **M1**. The closed-form expression for the GMLE of this high-dimensional parameter (with frailties known and θ and β fixed) allows the two-part M-step. This reduces high dimensional optimization to calculation of a weighted Cox partial likelihood estimate in **M1** and a weighted Breslow (1974) estimate in **M2**. This is a structure we will try to mimic for the CO model.

3.3 An Approximate EM-Algorithm

The Clayton-Oakes model with Cox regression margins may be expressed as a gamma frailty model with basic hazards given in (5). This allows the model to be placed in the likelihood-based framework of Nielsen. While some of Nielsen's results apply immediately, others do not. The likelihood in (6) and E-step in (8) translate exactly. However, the basic hazards follow a complicated form and the structure which allows the M-step to simplify computation does not apply. Examples considered by Nielsen *et al.* were models for the basic hazard which followed familiar models. In those situations GML estimates based on $(\mathcal{G}_t)$ were simple extensions of familiar estimators. For hazards which follow (5), parameter estimation is not simple even if the frailties are observed; hence, repeated maximization of (6) does not immediately simplify the problem. However, we develop an EM-algorithm which uses an approximate M-step and which mimics the structure of Nielsen *et al.*.

The M-step for an EM-algorithm should maximize (6) when the basic hazards follow (5). To mimic the prior EM-algorithm, the structure of an **M2** step must permit a closed form for $\Lambda_0(\cdot)$ with fixed β and θ and ξ. When basic hazards follow (5), the "complete data" GMLE for $\hat{\Lambda}_0(\cdot)$ can only be expressed implicitly as the solution to the score equation $U(t, \hat{\Lambda}_0|\beta, \theta) = 0$ for all $t \in [0, \tau]$ where

$$U(t) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t dN_{ik}(s) - \xi_i \exp\{\beta' Z_{ik} + \theta e^{\beta' Z_{ik}}\Lambda_0(s-)\}Y_{ik}(s)d\Lambda_0(s)+$$

$$\theta\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t \left[\int_{s+}^{\tau} dN_{ik}(u) - \xi_i \exp\{\beta' Z_{ik} + \theta e^{\beta' Z_{ik}}\Lambda_0(u-)\}d\Lambda_0(u)\right] Y_{ik}(s)e^{\beta' Z_{ik}}d\Lambda_0(s) \tag{10}$$

The structure of (10) does not allow a closed-form expression for $\hat{\Lambda}_0(\cdot)$. However, there is a structure to (10) which suggests an approximate GMLE. The first term forms a estimating equation which is easily solved and which yields an unbiased estimating equation since,

$$\int_0^t dN_{ik}(s) - \xi_i \exp\{\beta_0' Z_{ik} + \theta_0 e^{\beta_0' Z_{ik}} \Lambda_0(s-)\} Y_{ik}(s) d\Lambda_0(s)$$

is a $(\mathcal{G}_t)$ martingale. The solution to the first term is :

$$\hat{\Lambda}_0(t) = \int_0^t \left[\sum_{i=1}^n \sum_{k=1}^K \xi_i \exp\{\beta' Z_{ik} + \theta e^{\beta' Z_{ik}} \hat{\Lambda}_0(s-) Y_{ik}(s)\} \right]^{-1} dN_{..}(s) \qquad (11)$$

When $\beta_0 = 0$ or $\theta_0 = 0$, this approximate GMLE is exact.

Using this approximation as the basis for an **M2** step, it is possible to develop an algorithm for calculating profile GMLEs of $\hat{\Lambda}_0(\cdot)$ and $\hat{\beta}$ for a fixed $\theta = \tilde{\theta}$. Given initial values $\hat{\beta}^{(0)}$ and $\hat{\Lambda}^{(0)}(\cdot)$, the $l+1$th step takes the form:

C1: Calculation of Basic Intensity.

$$\hat{A}_{ik}^{(l)}(t) = \int_0^t \exp\{\hat{\beta}^{(l)\prime} Z_{ik} + \tilde{\theta} e^{\hat{\beta}^{(l)\prime} Z_{ik}} \hat{\Lambda}_0^{(l)}(s-)\} d\hat{\Lambda}_0^{(l)}(s),$$

E1: Posterior Expectation of ξ_i.

$$\hat{\xi}_i^{(l+1)} = (1 + \tilde{\theta} N_{i.}(\tau))/(1 + \tilde{\theta} \sum_{k=1}^K \int_0^\tau Y_{ik}(s) d\hat{A}_{ik}^{(l)}(s)),$$

M1: 'M'-step for β.

$$\hat{\beta}^{(l)} = \arg\max_\beta \prod_{i=1}^n \prod_{k=1}^K \prod_{t \in [0,\tau]} \left(\frac{\hat{\xi}_i^{(l+1)} \exp\{\beta' Z_{ik} + \tilde{\theta} e^{\beta' Z_{ik}} R(t-; \beta, \hat{\xi}^{(l+1)})\}}{\sum_{j=1}^n \sum_{m=1}^K \hat{\xi}_j^{(l+1)} \exp\{\beta' Z_{jm} + \tilde{\theta} e^{\beta' Z_{jm}} R(t-; \beta, \hat{\xi}^{(l+1)})\} Y_{jm}(t)} \right)^{\Delta N_{ik}(t)}$$

where $R(t, \beta, \xi)$ is given in **M2**.

M2: Approximate 'M'-step for $\Lambda_0(\cdot)$ given $\hat{\beta}^{(l+1)}$, $\hat{\xi}^{(l+1)}$. $\hat{\Lambda}_0^{(l+1)}(t) = R(t; \hat{\beta}^{(l+1)}, \hat{\xi}^{(l+1)})$

where

$$R(t; \beta, \xi) = \int_0^t \left[\sum_{i=1}^n \sum_{k=1}^K \xi_i \exp\{\beta' Z_{ik} + \tilde{\theta} e^{\beta' Z_{ik}} R(s-; \beta, \xi)\} Y_{ik}(s) \right]^{-1} dN_{..}(s),$$

where $R(0, \beta, \xi) = 0$ for β and $\xi_i (i = 1, \ldots, n)$. The step **M2** is based on (11) and produces a step function with jumps at the observed failure times. The step function is solved inductively. The step at the first failure time is the jump in the Breslow estimator. The jump at t incorporates the cumulative hazard prior to time t. The step **M1** maximizes the approximate profile likelihood obtained when $R(\cdot; \beta, \xi)$ replaces $\hat{\Lambda}_0(\cdot)$ in (6). The steps **C1,E1,M1** and **M2** are repeated until convergence $(l = 0, \cdots, L)$. The values of $\hat{\beta}_0^{(l)}$ and $\hat{\Lambda}_0^{(l)}(\cdot)$ at convergence are denoted as $\hat{\beta}|\tilde{\theta}$ and $\hat{\Lambda}_0(\cdot|\tilde{\theta})$. One dimensional methods are used to maximize the (approximate) profile likelihood in θ.

The algorithm works well for producing approximate GMLEs. When θ is fixed at 0 (i.e., if failures are considered independent), the joint estimators for β and $\Lambda_0(\cdot)$ are the Cox's maximum partial likelihood estimator and the Breslow estimator respectively. When either β or θ are near zero, the algorithms give an excellent, rapid approximation to the true GMLEs. The estimates can be used as starting values in a more intensive optimization, but the simulation studies suggest they are good estimates in their own right. The algorithm is easily adapted to calculate profile likelihoods in β and θ which can form the basis for approximate likelihood-based intervals and tests.

4. Simulation Studies

In our numerical studies, we assessed the approximate GMLE algorithm by simulating data from a CO model, and studying the behavior of the estimators and the likelihood ratio tests. The simulations were designed to assess the performance of the point estimates $(\hat{\beta}, \hat{\theta})$ and to estimate the empirical size of the 0.05 level tests of $\beta = \beta_0$ and $\theta = \theta_0$. The simulations also compare the efficiency of the estimate of the marginal hazard ratio calculated by our methods to the naive Cox estimate (the Cox maximum partial likelihood estimator when clustering is ignored), and to the approximate GMLE if the value of θ_0 is known.

For each set of parameter values we simulated $M = 300$ datasets. Each dataset included $n = 150$ independent sets of paired failure times: $(T_{i1}, T_{i2}), i = 1, \ldots, n$. The data followed a CO model such that T_{i1} had a marginal exponential distribution with mean R, and T_{i2} had a marginal unit exponential distribution. Hence, the marginal log-hazard ratio of T_{i1} to T_{i2} was β_0 ($\beta_0 = \log R$). The joint survival function for $(T_{i1}, T_{i2}$ $(i = 1, \ldots, n)$ followed:

$$S(t_1, t_2) = (\exp(\theta_0 R t_1) + \exp(\theta_0 t_2) - 1)^{-\theta_0^{-1}}$$

for all $t_1, t_2 \in [0, \tau] \times [0, \tau]$.

The data (t_1, t_2) were simulated as:

$$t_{i1} = -R^{-1}\theta^{-1} \log(-\theta_0 \log u_{i1}/\xi_i + 1) \qquad t_{i2} = -\theta_0^{-1} \log(-\theta_0 \log u_{i2}/\xi_i + 1)$$

where u_{ik} $(k = 1, 2)$,$(i = 1, \ldots, 150)$ are mutually independent random variables with unit exponential distributions, and $\xi_i (i = 1, \ldots, 150)$ are mutually independent gamma random variables with mean one and variance θ_0. The failure times are censored at $t = 1.75$ to give approximately 30% censoring.

Each dataset was analyzed using the approximate EM-algorithm described in Section 3.3. Four sets of values of (θ_0, β_0) were generated and 300 datasets were analyzed for each scenario. The point estimates and size of the likelihood ratio statistics are summarized in Table 1. Each dataset also had the log-hazard ratio estimated by a naive Cox model (i.e., $\hat{\beta}|\theta = 0$) and by an approximate GMLE of β from a CO model with θ fixed at the true value θ_0 (i.e., $\hat{\beta}|\theta_0$). Those results are summarized by Table 2.

A small negative bias is apparent in the estimation of θ_0. This has been reported by other authors (Nielsen *et al.*, 1992; Hsu & Prentice, 1996) who found the bias is mitigated by censoring large failure times and by increasing n. Likelihood ratio tests appear to have satisfactory size.

In Table 2, the empirical variance of the estimators of the log-hazard ratio are very similar, suggesting that knowledge of a CO structure for paired data does not improve estimation of β. In fact, even when the data follows a CO structure with known $\theta = \theta_0$, the naive Cox estimate is nearly fully efficient.

5. Examples

We consider a subset of 197 patients selected and analyzed by Huster *et al.* (1989) from the Diabetic Retinopathy Study. The study was designed to evaluate the effectiveness of laser photocoagulation in delaying the onset of blindness in patients with diabetic retinopathy. Secondary questions included whether treatment effect varied by type of diabetes (juvenile or adult-onset) and whether there was significant intra-individual correlation in disease onset times. Each patient had a single eye randomized to laser treatment and the other eye was observed without treatment.

Table 3 displays the fit of CO models to the data. The GMLE for $\hat{\beta}$ and $\hat{\theta}$ are (-0.78,0.92). The likelihood ratio statistic for $H_0 : \theta_0 = 0$ is 12.8, ($p < 0.001$) and the likelihood ratio statistic for $H_0 : \beta_0 = 0$ is 6.7 ($p < 0.01$). The results show significant effects of treatment, intra-cluster correlation, and greater effectiveness of treatment among patients with adult-onset diabetes. The results of the final model are compared with estimates from other methods in Table 4. The results are similar to previous results from a CO model with Weibull regression margins by Huster *et al.* (1989), a pseudolikelihood method proposed by Liang *et al.* (1993), and the method of Wei *et al.* fit by Lin (1994).

Batchelor & Hackett (1970) studied typing of allogenic skin grafts for the human leukocyte antigen (HLA) to determine the degree of antigenic incompatibility between host and donor. Sixteen patients with burns on more than 18% of the body surface were typed for HLA. For each patient, two skin grafts were selected: one graft with a slight HLA incompatibility and a second graft with a serious HLA incompatibility. The survival of poorly matching grafts was compared to the survival of closely matching grafts. The results appear in Table 5. After controlling for the percentage of the body burned, antigenic compatibility is significant ($\chi_1^2 = 6.1, p = 0.013$), but there is no evidence of clustering of graft survival times within a recipient ($\chi_1^2 = 0.10, p > 0.50$).

6. Discussion

We have described an approach to estimation in a model which naturally incorporates simultaneous inference about regression effects and dependence between failures. The GML approach provides joint estimates of all the model parameters: regression coefficients, dependence parameters and a marginal baseline hazard function. The approach also accommodates censoring, the presence of ties and varying cluster sizes. Asymptotic theory for our estimators is largely unavailable. Recent work by Murphy (1995; 1996), however, may provide a template for the development of a large sample theory.

The estimation of β and θ has been proposed by two-stage methods. Liang (1991) and Genest *et al.* (1995) used pseudolikelihood methods to estimate β and $\Lambda_0(\cdot)$. They then estimated θ by treating the first stage estimates as fixed and maximized the one dimensional likelihood. These approaches appear to be computationally simpler than our methods, but their robustness and relative efficiency is unknown.

Analysis of MVFT data with a CO model is a promising possibility. The practical analysis of data from a CO model would benefit from methods which can diagnose the appropriateness of the model for a given dataset, which can efficiently estimate marginal hazard ratios and dependence parameters, and which can distinguish the relative fit of a marginal Cox model and a gamma frailty Cox model.

References

Batchelor, J.R. & Hackett, M. (1970). HLA matching in treatment of burned patients with skin allografts. *Lancet* **2**, 581-583.

Breslow, N.E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89-99.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141-151.

Clayton, D.G. & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *J. Roy. Statist. Soc. Ser. A* **148**, 82-108.

Cox, D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Genest, C., Ghoudi, K. & Rivest, L.P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543-552.

Gill, R.D. (1989). Non- and semi-parametric maximum likelihood and the von Mises method (Part 1). *Scand. J. Statist.* **16**, 97-128.

Gill, R.D. (1992). Marginal partial likelihood. *Scand. J. Statist.* **19**, 133-137.

Hsu, L. & Prentice, R.L. (1996). On assessing the strength of dependency between failure time variates. *Biometrika* **83**, 491-506.

Huster, W.J., Brookmeyer, R. & Self, S.G. (1989). Modeling paired survival data with covariates. *Biometrics* **45**, 145-156.

Liang, K.Y. (1991). Estimating the effect of probands' characteristics on familial risk: I. adjustment for censoring and correlated ages at onset. *Genet. Epidem.* **8**, 329-338.

Liang, K.Y., Self, S.G. & Chang, Y.C. (1993). Modeling marginal hazards in multivariate failure time data. *J. Roy. Statist. Soc. Ser. B* **55**, 441-453.

Lin, D.Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statist. Med.* **13,** 2233-2247.

Klein, J.P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48**, 795-806.

Murphy, S.A. (1995). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712-731.

Murphy, S.A. (1996). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182-198.

Nielsen, G.G., Gill, R.D., Andersen, P.K. & Sørensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25-43.

Oakes, D. (1982). A model for association in bivariate survival data. *J. Roy. Statist. Soc. Ser. B* **44**, 414-422.

Vaupel, J.W., Manton, K.G. & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439-454.

Wei, L.J., Lin, D.Y. & Weissfeld, L. (1989). Regression analysis of incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065-1073.

David V. Glidden, Division of Biostatistics, Box 0560, University of California, San Francisco, CA 94143-0560, U.S.A.

Table 1

Behaviour of Point Estimates of β and θ and Size of a 0.05 Level Likelihood Ratio Test

		$\beta_0 = 0$		$\beta_0 = -\log(2)$	
		$\theta_0 = 1$	$\theta_0 = 3$	$\theta_0 = 1$	$\theta_0 = 3$
β:					
	mean $\hat{\beta}$	0.01	0.00	-0.69	-0.69
	standard deviation $\hat{\beta}$	0.10	0.05	0.12	0.09
θ:					
	mean $\hat{\theta}$	0.97	2.94	1.00	2.92
	standard deviation $\hat{\theta}$	0.25	0.51	0.29	0.52
LR Tests:					
	P(Reject $H_0 : \theta = \theta_0$)	0.06	0.05	0.05	0.05
	P(Reject $H_0 : \beta = \beta_0$)	0.05	0.04	0.05	0.07

Table 2: Estimates for β

Moments of Estimates of β with Varying Assumptions for θ

		$\beta_0 = 0$		$\beta_0 = -\log(2)$	
		$\theta_0 = 1$	$\theta_0 = 3$	$\theta_0 = 1$	$\theta_0 = 3$
Naive: $\hat{\beta}\|\theta = 0$					
	mean $\hat{\beta}$	0.01	0.00	-0.69	-0.69
	standard deviation $\hat{\beta}$	0.10	0.06	0.12	0.10
GMLE : $\hat{\beta}\|\theta = \hat{\theta}$					
	mean $\hat{\beta}$	0.01	0.00	-0.69	-0.69
	standard deviation $\hat{\beta}$	0.10	0.05	0.12	0.09
Profile: $\hat{\beta}\|\theta = \theta_0$					
	mean $\hat{\beta}$	0.01	0.00	-0.69	-0.69
	standard deviation $\hat{\beta}$	0.10	0.05	0.12	0.09

Table 3:

Fit to the Diabetic Retinopathy Data

Model	Log-Likelihood	$\hat{\theta}$	Treatment ($\hat{\beta}_1$)	Type of Diabetes ($\hat{\beta}_2$)	Type By Treat ($\hat{\beta}_3$)
θ only	-994	0.56	(-)	(-)	(-)
Treatment	-979	0.92	-0.76	(-)	(-)
Main Effects	-979	0.94	-0.77	0.12	(-)
Interaction	-975	1.02	-0.42	0.35	-0.82

Table 4: *Estimates of regression parameters for the Diabetic Retinopathy Study by various methods*

	Methods			
Covariate	SGML[1]	Wei[2]	Liang[3]	Huster[4]
Treatment	-0.42	-0.43	-0.42	-0.43
		(0.19)	(0.19)	(0.22)
Diabetic type	0.35	0.34	0.34	0.37
		(0.20)	(0.20)	(0.20)
Interaction	-0.82	-0.85	-0.84	-0.84
		(0.30)	(0.30)	(0.35)

Standard errors are given in parentheses.

1, Semiparametric approximate GMLEs

2. Naive estimator with robust standard error due to Wei *et al.*, (1989)

3. Pseudolikelihood method due to Liang *et al.*, (1993)

4. Parametric model fit by Huster *et al.*, (1989)

Table 5:

Model results for skin graft data

Model	Log-Likelihood	$\hat{\theta}$	Close Match ($\hat{\beta}_1$)	Percent Burned ($\hat{\beta}_2$)
θ only	-80	0.302	(-)	(-)
Close Match	-77	0.575	-.943	(-)
% Burned and Match	-73	0.110	-1.252	-0.078