Harvard University
Harvard University Biostatistics Working Paper Series

Year 2009 *Paper* 104

Marginalized Frailty Models for Multivariate Survival Data

Megan Othus* Yi Li†

*Harvard University and Dana Farber Cancer Institute, mothus@hsph.harvard.edu
†Harvard University and Dana Farber Cancer Institute, yili@jimmy.harvard.edu

Marginalized Frailty Models for Multivariate Survival Data

Megan Othus

Harvard University and Dana Farber Cancer Institute

Boston, MA, 02115

Yi Li

Harvard University and Dana Farber Cancer Institute

Boston, MA, 02115

Megan Othus is graduate student, Department of Biostatistics, Harvard University and Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, MA, 02115 (email: mothus@fas.harvard.edu); Yi Li is Associate Professor, Department of Biostatistics, Harvard University and Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, MA, 02115 (email: yili@jimmy.harvard.edu). This work was supported in part by National Cancer Institute grants R01 CA95747 and CA09337-25. The authors thank Jim Anderson for providing the Children's Oncology Group dataset and advice he gave for the data application.

Abstract

We consider a class of normal transformation models for clustered failure time data. The failure time outcomes are assumed to marginally follow a proportional hazards model, while the normally transformed variates allow a shared frailty. As a result, the model permits population-level interpretation of covariates in the proportional hazards model, but also directly models the correlation of the transformed failure times. The method allows for varying cluster sizes, and we are able to predict shared frailties for the transformed failure times. Predictions of the frailties allow us to evaluate the role of underlying cluster effects on subjects' survival. We propose a profile estimation procedure and derive asymptotic properties under this estimation scheme. Simulation studies verify finite sample utility. We apply the method to a Children's Oncology Group multi-center study of acute lymphoblastic leukemia. The analysis estimates marginal treatment effects and examines potential clustering within treatment institution.

Keywords: Semiparametric normal transformation, frailty model, marginal model, correlated survival data, proportional hazards model

1. Introduction

Two major branches of models for correlated survival data are frailty models (Oakes 1989; Murphy 1994, 1995; Parner 1998) and marginal models (Wei, Lin, and Weissfeld 1989; Prentice and Cai 1992; Cai and Prentice 1995). The parameters from these two approaches have different interpretations and hence are appropriate in different contexts. For example, frailty models are used when within cluster inferences are desired. Frailty models account for dependence by including a multiplicative term (called a frailty or a random-effect) in the model for the hazard. The multiplicative nature of the frailty term implies that parameters have interpretations conditional on the value of the frailty. In contrast, marginal models directly model the marginal failure time and within cluster correlations are treated as a nuisance. With marginal models, the correlated nature of the data is often accounted for by using a sandwich-type variance. The parameters of marginal models have population-average interpretations.

In some correlated survival data settings, practitioners have two primary interests: assessing the effect of treatment or exposure on the marginal survival distribution and determining the dependence between subjects. For example, in many multi-center clinical trials data are clustered within treatment center. Clinical researchers are interested in the unconditional treatment effect observed in the study, which can be found via marginal modeling. Although institutions participating in clinical trials follow trial-specific protocols, differences can still exist between outcomes (Fleiss 1986; Gray 1994; Jones, Teather, Wang, and Lewis 1998; Senn 1998; Anello, O'Neill, and Dubey 2005; Vierron and Giraudeau 2007; Logan, Nelson, and Klein 2008; Zheng and Zelen 2008). The correlation between patients treated at the same institution is an important component of a multi-center clinical trial analysis.

Our motivating data come from a large multi-center clinical trial for children with

"higher risk" acute lymphoblastic leukemia. The goal of the study was to evaluate the effect of different treatments on survival. We are interested in evaluating whether there existed substantial variation between institutions while concurrently assessing the efficacy of the new treatments.

To this end we propose a marginalized frailty model, which models the marginal failure times with the proportional hazards model (Cox 1972) and models the correlation by assuming that normally transformed survival times follow a shared frailty model. This project extends previous work on normal transformation models (e.g. Li and Lin (2006); Li, Prentice, and Lin (2008)) in four major ways. First, Li et al. (2008) use a likelihood method that does not allow for covariates and is restricted to bivariate data. Our model allows for covariates and varying cluster sizes. Second, Li and Lin (2006) assume a specific spatial correlation structure on the entire dataset. Our method explicitly allows for correlated survival times within independent clusters. Third, we establish a likelihood framework for inference for general regression models for the normally-transformed survival times. Finally, we provide a method for predicting cluster-level effects, providing information that can be used to evaluate individual clusters.

The rest of the paper is structured as follows: in Section 2 we define notation and describe the model; Section 3 provides methods to estimate the marginal survival parameters, the correlation parameter, and the frailties; we provide a summary of asymptotic results in Section 4; we outline a model-checking procedure in Section 5; simulations are presented in Section 6; Section 7 contains an analysis of Children's Oncology Group study 1961; and we finish with a brief discussion in Section 8. Regularity conditions and detailed proofs of theorems are contained in the Appendix.

2. Model Specification

Let T_{ij} and C_{ij} denote potentially unobserved failure and censoring times for subject j in cluster i, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$. The observed data are $X_{ij} = \min(T_{ij}, C_{ij})$ and $\Delta_{ij} = I(T_{ij} \leq C_{ij})$. Let $\mathbf{Z}_{ij}(t)$ denote an external time-dependent covariate vector (Kalbfleisch and Prentice 2003, page 197) of length p and write its covariate path up to time t as $\bar{\mathbf{Z}}_{ij}(t) = \{\mathbf{Z}_{ij}(s) \mid 0 \leq s \leq t\}$. Assume that T_{ij}, conditional on the covariate process $\bar{\mathbf{Z}}_{ij}(T_{ij})$, is independent of C_{ij}. Also, assume that, conditional on each individual's covariate path, the hazard of T_{ij} follows a proportional hazards model where $\lambda\{t \mid \bar{\mathbf{Z}}_{ij}(t)\}$ is equal to

$$\lim_{h \to 0} h^{-1} P\{t \leq T_{ij} < t + h \mid T_{ij} \geq t, \bar{\mathbf{Z}}_{ij}(t)\} = \lambda_0(t) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(t)\}. \tag{1}$$

Here $\boldsymbol{\beta}$ is a vector of regression coefficients and $\lambda_0(t)$ is an unspecified baseline hazard function with cumulative hazard function Λ. Equation (1) is a marginal model for each T_{ij}, hence $\boldsymbol{\beta}$ has a population-level interpretation not a cluster-specific interpretation.

To model the clustering of the T_{ij}, consider the semiparametric normal transformation:

$$\tilde{T}_{ij} = \Phi^{-1}[1 - S\{T_{ij} \mid \bar{\mathbf{Z}}_{ij}(T_{ij})\}], \tag{2}$$

where Φ is the standard normal distribution function, Φ^{-1} is the inverse of the standard normal distribution function, and S is the survival function associated with Equation (1). By the probability integral transform, $1 - S\{T_{ij} \mid \bar{\mathbf{Z}}_{ij}(T_{ij})\}$ has a Uniform$(0, 1)$ distribution. It necessarily follows that $\tilde{T}_{ij} \sim$ Normal$(0, 1)$. The transformation takes T_{ij} with support on $(0, \infty)$ and transforms it to a standard normal random variable.

Denote the correlation of $(\tilde{T}_{i1}, \ldots, \tilde{T}_{in_i})$ with $\mathbf{\Sigma}_i$. We propose to model the transformed survival times from Equation (2) with a shared frailty model:

$$\tilde{T}_{ij} = \sqrt{\sigma} b_i + \epsilon_{ij}, \tag{3}$$

where the cluster level frailty b_i has a standard normal distribution and the error terms, ϵ_{ij}, are independent and identically distributed $N(0, 1-\sigma)$ random variables that are also independent of b_i. We call the class of models described by Equations (1), (2), and (3) *marginalized frailty models*. The $\boldsymbol{\beta}$ parameters in Equation (1) have marginal interpretation, while σ and b_i from Equation (3) characterize the cluster effect. The term σ can have the interpretation as the proportion of explained variance shared by members of the same cluster. In the context of a multi-center clinical trial, the cluster level frailties characterize the center effect. Using this model, $\mathbf{\Sigma}_i$ will have an exchangeable structure with diagonal elements equal to 1 and off-diagonal elements equal to σ.

3. Inference

3.1. Likelihood Function

The survival function for $T_{ij} \mid b_i$ can be written $P(T_{ij} > t \mid b_i) = 1 - \Phi_{1-\sigma}(\tilde{t} - \sqrt{\sigma} b_i)$, with density function $f_{ij}(t) = \phi_{1-\sigma}(\tilde{t} - \sqrt{\sigma} b_i) f\{t \mid \bar{\mathbf{Z}}_{ij}(t)\} / \phi_{1-\sigma}(\tilde{t})$, where $\tilde{t} = \Phi^{-1}[1 - S\{t \mid \bar{\mathbf{Z}}_{ij}(t)\}]$, $\Phi_{1-\sigma}$ is the distribution function for ϵ_{ij}, $\phi_{1-\sigma}$ is the density of ϵ_{ij}, and $f\{t \mid \bar{\mathbf{Z}}_{ij}(t)\}$ is the density associated with Equation (1). Let $\tilde{\Phi}_{1-\sigma}(t) = 1 - \Phi_{1-\sigma}(t)$ and $\tilde{X}_{ij} = \Phi^{-1}[1 - S\{X_{ij} \mid \bar{\mathbf{Z}}_{ij}(X_{ij})\}]$. Here $\tilde{X}_{ij}$ is a potentially censored version of $\tilde{T}_{ij}$ (Equation (2)). The semiparametric normal transformation is monotone and thus preserves censoring patterns. The likelihood based on the observed data can be

written

$$\begin{aligned} L(\sigma,\boldsymbol{\beta},\Lambda) &= \prod_{i=1}^{m}\int\prod_{j=1}^{n_i}\tilde{\Phi}_{1-\sigma}(\tilde{X}_{ij}-\sqrt{\sigma}b_i)^{1-\Delta_{ij}}[\lambda_0(X_{ij})\exp\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(X_{ij})\}]^{\Delta_{ij}} \\ &\quad\times(\exp[\int_0^{X_{ij}}\exp\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(s)\}\,d\Lambda(s)]\phi_{1-\sigma}(\tilde{X}_{ij}-\sqrt{\sigma}b_i)/\phi_{1-\sigma}(\tilde{X}_{ij}))^{\Delta_{ij}}\,d\Phi(b_i) \\ &= \prod_{i=1}^{m}\int\prod_{j=1}^{n_i}[f\{X_{ij}\mid\bar{\mathbf{Z}}_{ij}(X_{ij})\}/\phi_{1-\sigma}(\tilde{X}_{ij})]^{\Delta_{ij}} \\ &\qquad\times\phi_{1-\sigma}(\tilde{X}_{ij}-\sqrt{\sigma}b_i)^{\Delta_{ij}}\tilde{\Phi}_{1-\sigma}(\tilde{X}_{ij}-\sqrt{\sigma}b_i)^{1-\Delta_{ij}}\,d\Phi(b_i). \end{aligned}$$

Conveniently, $L(\sigma,\boldsymbol{\beta},\Lambda)$ has a closed form expression proportional to a product of multivariate normal terms. To simplify the presentation, define $\Delta_i=\sum_{j=1}^{n_i}\Delta_{ij}$, order the observations such that $\Delta_{i1}=\ldots=\Delta_{i\Delta_i}=1$, and let $\tilde{\mathbf{X}}_i^{\Delta_i}=(\tilde{X}_{i1},\ldots,\tilde{X}_{i\Delta_i})$ and $\tilde{\mathbf{X}}_i^{n_i-\Delta_i}=(\tilde{X}_{i\Delta_{i+1}},\ldots,\tilde{X}_{in_i})$.

Then $L(\sigma,\boldsymbol{\beta},\Lambda)$ can be written

$$\prod_{i=1}^{m}\phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})\tilde{\Phi}_c^{n_i-\Delta_i}(\tilde{\mathbf{X}}_i^{n_i-\Delta_i}\mid\tilde{\mathbf{X}}_i^{\Delta_i})\prod_{j=1}^{n_i}[f\{X_{ij}\mid\bar{\mathbf{Z}}_{ij}(X_{ij})\}/\phi_{1-\sigma}(\tilde{X}_{ij})]^{\Delta_{ij}},$$

where $\phi_u^{\Delta_i}$ is the multivariate normal density corresponding to its argument and $\tilde{\Phi}_c^{n_i-\Delta_i}$ is the multivariate normal survival function corresponding to its argument. The distribution of each of the arguments will be outlined below. Let $\boldsymbol{\Sigma}_i$ be the covariance matrix for the transformed failure times. Write $\boldsymbol{\Sigma}_i$ as a partitioned matrix:

$$\boldsymbol{\Sigma}_i=\begin{pmatrix}\boldsymbol{\Sigma}_{i11} & \boldsymbol{\Sigma}_{i12}\\ \boldsymbol{\Sigma}_{i21} & \boldsymbol{\Sigma}_{i22}\end{pmatrix},$$

where $\boldsymbol{\Sigma}_{i11}$ has dimension $\Delta_i\times\Delta_i$. The vector $\tilde{\mathbf{X}}_i^{\Delta_i}$ follows a multivariate normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma}_{i11}$. For the second term of the

likelihood, $\tilde{\mathbf{X}}_i^{n_i-\Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i}$ is a censored observation from a normal distribution with mean $\boldsymbol{\Sigma}_{i21}\boldsymbol{\Sigma}_{i11}^{-1}\tilde{\mathbf{X}}_i^{i\Delta_i'}$ and covariance matrix $\boldsymbol{\Sigma}_{i22} - \boldsymbol{\Sigma}_{i21}\boldsymbol{\Sigma}_{i11}^{-1}\boldsymbol{\Sigma}_{i12}$.

To shed some light on the likelihood, $L(\sigma, \boldsymbol{\beta}, \Lambda)$, we consider an example likelihood contribution from a cluster of size two where one subject is observed to be censored at time C_{A1} and one subject is observed to fail at time T_{A2}. The covariate process for each subject is denoted $\bar{Z}_{A1}(C_{A1})$ and $\bar{Z}_{A2}(T_{A2})$. The normally transformed observed failure times are $\tilde{X}_{A1} = \Phi^{-1}[1 - S\{C_{A1} \mid \bar{Z}_{A1}(C_{A1})\}]$ and $\tilde{X}_{A2} = \Phi^{-1}[1 - S\{T_{A2} \mid \bar{Z}_{A2}(T_{A2})\}]$. In this case

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}.$$

The first term of $L(\sigma, \boldsymbol{\beta}, \Lambda)$ can be written

$$\phi_u^1(\tilde{X}_{A2}) = (2\pi)^{-1/2} \exp(-\tilde{X}_{A2}^2/2) \tag{4}$$

while the second term can be written

$$\Phi_c^1(\tilde{X}_{A1}) = \int_{\tilde{X}_{A1}}^{\infty} \{2\pi(1-\sigma^2)\}^{-1/2} \exp\{-(x - \sigma\tilde{X}_{A2})^2/2\sqrt{1-\sigma^2}\}\, dx. \tag{5}$$

Equation (4) is the density of a standard normal random variable, while Equation (5) corresponds to, conditional on $\tilde{X}_{A2}$, the probability that a Normal($\sigma\tilde{X}_{A2}$, $1-\sigma^2$) random variable is greater than $\tilde{X}_{A1}$.

3.2. Profile Likelihood Estimation

We propose a profile method to estimate $(\sigma, \boldsymbol{\beta}, \Lambda)$:

Step 1: Estimate $\boldsymbol{\beta}$ and Λ assuming working independence. Denote these estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}$.

Step 2: Estimate σ as the maximum of $L(\sigma, \hat{\boldsymbol{\beta}}, \hat{\Lambda})$. Write this maximum as $\hat{\sigma}$.

Formulas for the standard errors of $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}$ can be found using a sandwich formula (Spiekerman and Lin 1998). The standard error of $\hat{\sigma}$ needs to account for the variability from $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}$. This can be accomplished using a jackknife resampling scheme. To maintain the correlated structure of the failure times, the clusters should be the unit of removal for the jackknife calculations (Cai et al. 1997; Cai and Shen 2000).

This estimation procedure is computationally straightforward. Marginal estimates of the survival function assuming working independence are available in all standard computing programs. The likelihood for σ, $L(\sigma, \hat{\boldsymbol{\beta}}, \hat{\Lambda})$, is proportional to a product of multivariate normal terms. The multivariate normal terms can be evaluated quickly using existing software (e.g. `R` package `mvtnorm`).

3.3. Predictions of Shared Frailties

Our marginalized frailty model allows us to estimate the shared frailties, which is often of interest. In the context of institutional clustering, the b_i can provide information on the results of individual institutions participating in a clinical trial. Estimates of the b_i and their standard errors can be found using Laplace approximations to the b_i's first two moments. Denote the observed data for the i^{th} cluster, $(X_{i1}, \ldots, X_{in_i}, \Delta_{i1}, \ldots, \Delta_{in_i}, \bar{Z}_{i1}(X_{i1}), \ldots, \bar{Z}_{in_i}(X_{in_i}))$, with $\boldsymbol{\Psi}_i$. The conditional density of b_i given the observed data $\boldsymbol{\Psi}_i$, denoted $g(b_i \mid \boldsymbol{\Psi}_i; \sigma, \boldsymbol{\beta}, \Lambda)$, can be written

$$L_i^{-1}(2\pi)^{-1/2}\exp(-b_i^2/2)\prod_{j=1}^{n_i}[f\{X_{ij} \mid \bar{\mathbf{Z}}_{ij}(X_{ij})\}/\phi_{1-\sigma}(\tilde{X}_{ij})]^{\Delta_{ij}}$$
$$\times\phi_{1-\sigma}(\tilde{X}_{ij} - \sqrt{\sigma}b_i)^{\Delta_{ij}}\tilde{\Phi}_{1-\sigma}(\tilde{X}_{ij} - \sqrt{\sigma}b_i)^{1-\Delta_{ij}}$$

where L_i is the likelihood for $\boldsymbol{\Psi}_i \mid \sigma, \boldsymbol{\beta}, \Lambda$. Define k_i such that $g(b_i \mid \boldsymbol{\Psi}_i; \sigma, \boldsymbol{\beta}, \Lambda) = L_i^{-1} \exp\{k_i(b_i \mid \boldsymbol{\Psi}_i; \sigma, \boldsymbol{\beta}, \Lambda)\}$. Using the Laplace approximations to the first two moments of $g(b_i \mid \boldsymbol{\Psi}_i; \sigma, \boldsymbol{\beta}, \Lambda)$ (Booth and Hobert 1998), the predicted estimate and variance of b_i are taken to be:

$$\hat{b}_i = E(b_i \mid \boldsymbol{\Psi}_i) \approx \arg\max_{b_i} k_i(b_i \mid \boldsymbol{\Psi}_i; \hat{\sigma}, \hat{\boldsymbol{\beta}}, \hat{\Lambda}) \tag{6}$$

$$V(b_i \mid \boldsymbol{\Psi}_i) \approx -\ddot{k}_i(\hat{b}_i \mid \boldsymbol{\Psi}_i; \hat{\sigma}, \hat{\boldsymbol{\beta}}, \hat{\Lambda})^{-1}, \tag{7}$$

where double superscript dots denote second derivatives.

The prediction of the shared frailties is straightforward and computationally fast, particularly in contrast to standard gamma frailty algorithms, which can involve analytically complicated integrals. The expression for $k_i(b_i \mid \hat{\sigma}, \hat{\boldsymbol{\beta}}, \hat{\Lambda}, \boldsymbol{\Psi}_i)$ involves n_i normal terms and can be maximized using any standard optimization routine. The estimate of the variance of b_i has a closed form expression and can be found by plugging in relevant estimated quantities.

4. Theoretical Results

The following theorems establish the theoretical properties of $(\hat{\sigma}, \hat{\boldsymbol{\beta}}, \hat{\Lambda})$ where their true values are denoted with $(\sigma_0, \boldsymbol{\beta}_0, \Lambda_0)$.

Theorem 1. *Under Conditions C.1 – C.6 in the Appendix,* $(\hat{\sigma}, \hat{\boldsymbol{\beta}}, \hat{\Lambda})$ *converges in probability to* $(\sigma_0, \boldsymbol{\beta}_0, \Lambda_0)$ *as* $m \to \infty$.

Theorem 2. *Under Conditions C.1 – C.7 in the Appendix, as* $m \to \infty$, $\sqrt{m}(\hat{\sigma} - \sigma_0)$ *and* $\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *converge to zero-mean normal distributions and* $\sqrt{m}\{\hat{\Lambda}(t) - \Lambda_0(t)\}$ *converges to a zero-mean Gaussian process.*

Detailed proofs can be found in the Appendix. The proofs of both theorems for $\hat{\sigma}$ involve accounting for the proposed profile estimation method. These theorems verify

that $\hat{\sigma}$ is consistent and asymptotically normal when plug-in estimates of $\boldsymbol{\beta}$ and Λ are used in the likelihood function.

To prove that $\hat{\sigma}$ is consistent we first have to verify that the log-likelihood function is Hadamard differentiable with respect to Λ_0 (van der Vaart 1998). The next step is to show that the first order terms of a Taylor series expansion of the log-likelihood around $\boldsymbol{\beta}_0$ and Λ_0 are bounded. The final key step is to prove that the log-likelihood function with plug-in estimates for $\boldsymbol{\beta}$ and Λ converges uniformly to the expected value of the log-likelihood function evaluated at the true values for $\boldsymbol{\beta}$ and Λ.

The proof of normality for $\hat{\sigma}$ accounts for the profile estimation scheme by inflating the variance term to account for using $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}$. The key step of the proof is to show that the score equation can asymptotically be written as the sum of independent and identically distributed terms, facilitating the use of the central limit theorem. To make this argument we need to prove that the score equation for σ is Hadamard differentiable with respect to Λ_0 and that the first order terms of a Taylor series expansion of the score function around $\boldsymbol{\beta}_0$ and Λ_0 are bounded.

The analytical formula for the variance of $\hat{\sigma}$ is so complicated that it is of little computational utility. To evaluate the finite sample variability of $\hat{\sigma}$ we propose to use a jackknife resampling scheme. We evaluate the performance of the jackknife by simulation studies, which are presented in Section 6. The jackknife estimates of the standard error match well with Monte Carlo estimates of the standard error with moderate numbers of clusters.

5. Model-checking

In practice, it is often of interest to verify that any assumptions made during modeling are plausible. We propose a two-step model-checking procedure. As the marginal survival function is modeled using proportional hazards regression, a natural first

model-checking step is to verify that the associated regression assumptions are not violated. In our application, we use the method of Grambsch and Therneau (1994), though there have been many other methods proposed, such as Lin, Wei, and Ying (1993). Next, provided that the proportional hazards model fit is adequate, the assumption that the transformed failure times follow a shared frailty model (Equation (3)) can be checked.

Recall that $S\{\cdot \mid \bar{\mathbf{Z}}_{ij}(T_{ij})\}$ is the survival function associated with the marginal survival time T_{ij}. If S is estimated well, with estimate denoted $\hat{S}$, the probability integral transform indicates that it should be the case that $\Phi^{-1}\{1-\hat{S}(T_{ij} \mid \bar{\mathbf{Z}}_{ij}(T_{ij}))\}$ approximately follows a standard normal distribution. In verifying the assumptions of Equation (3), we need to check that the estimates of the frailties are normally distributed. To this end, let $z_i = \hat{b}_i/\sqrt{V(\hat{b}_i)}$, where $\hat{b}_i$ and $V(\hat{b}_i)$ are defined in Equations (6) and (7). A simple graphical check of the assumption that the b_i are normally distributed is a quantile-quantile plot (Q-Q plot) of the cumulative distribution function of the z_i.

6. Simulation Results

We conducted a number of simulations to evaluate the efficacy of the proposed method. The presented simulations have marginal survival times from a proportional hazards model with a constant baseline hazard function equal to 1 and with two covariates: one Bernoulli(0.5) covariate with parameter equal to log(0.5) (denoted β_1) and one Uniform$(0,1)$ covariate with parameter equal to 0.75 (denoted β_2). Censoring times were taken from the Exponential(mean=3) distribution and produced about a 25% censoring rate. Clustering was induced through generating survival times using Equation (3). The precision of $\hat{\sigma}$ is determined by the number of clusters, while the precision of the predicted frailties is determined by the sample size

in individual clusters. Each simulation is based on 500 replications. We summarize results for simulated datasets with 30, 60, and 90 clusters. The individual clusters sizes either varied between 2 and 10 with a median size of 3, or varied between 10 and 50 with median cluster size 17.

Simulation results are presented in Table 1. Even with small cluster sizes the estimates of σ are relatively unbiased and the jackknife standard errors match well with the Monte Carlo standard errors. For example, in a simulation of 30 clusters with median cluster size 3 the average estimate for σ is 0.531 (truth = 0.5), while in a simulation of 90 clusters with median cluster size 3 the average estimate for σ is 0.512.

The jackknife estimates of the standard errors for β_1 and β_2 are closer to the Monte Carlo standard errors than the sandwich-based standard errors for almost all the simulations. For example, in a simulation of 30 clusters with median size 3, the jackknife standard error of β_2 is 0.400, which is very close to the Monte Carlo standard error of 0.403, while the sandwich-based standard error is 0.366. In a simulation of 60 clusters with median size 17, the Monte Carlo and jackknife standard error estimates of β_1 are 0.079 and the sandwich-based standard error estimate is 0.077.

With respect to the frailty estimates, the median relative bias decreases and the likelihood-based standard error and the Monte Carlo standard error approach one as the individual cluster sizes increase. For example, in a simulation of 90 clusters with median cluster size 3 the median relative bias was -0.032 and in a simulation of 30 clusters with median size 3 the likelihood-based standard errors have an average value of 0.911.

7. Data Application: Children's Oncology Group Study 1961

We apply our method to a Children's Oncology Group study (protocol number 1961) (Seibel et al. 2008). The study population included 460 children with enlarged livers. The goal of this study was to determine whether increasing the intensity of therapy by increasing duration or strength for "higher risk" acute lymphoblastic leukemia patients would improve survival. To evaluate the relative benefit or harm for each type of intensification, a 2x2 factorial design with block randomization of patients was used. The distribution of subjects with enlarged livers among the four arms is presented in Table 2. We focus on the overall survival endpoint.

We present regression results for an analysis with a marginal survival model containing covariates for strength and duration of treatment in Table 3. We note that there was no evidence of a duration by strength interaction in the initial analysis of the entire dataset (Seibel et al. 2008) or in this enlarged liver subset. The p-value for increased strength of treatment is marginally significant (p-value = 0.056, using sandwich standard error) and indicates that increased strength of treatment may be associated with improved survival. The p-value for increased duration of treatment is marginally significant (p-value = 0.063, using sandwich standard error) and indicates that duration strength of treatment may be associated with worsened survival. For both survival parameters, the sandwich-based standard error and jackknife standard error estimates that account for clustering within institutions provide smaller standard error estimates than the naïve method assuming independence. The estimate for σ is 0.171 with a standard error of 0.152.

We also include results for a larger survival model that includes gender and potential prognostic factors age and platelet count at diagnosis. Regression results are summarized in Table 3. The p-values for increased strength (p-value = 0.067, using

sandwich standard error) and increased duration (p-value = 0.071, using sandwich standard error) are slightly larger in this larger model, but the direction of the effect remains the same. None of the p-values for gender, age, or platelets at diagnosis is significant. The estimate of σ increases slightly to 0.175 with standard error of 0.144.

We predicted the frailties for each of the institutions in the sample using the treatment only survival model. A summary of the distributions of the predictions can be found in Figure 1. The mean of the frailties is close to zero (0.024). A Q-Q plot with standardized predicted frailties compared to theoretical normal quantiles and the smoothed density of the predicted frailties indicate that the normal assumption in approximately satisfied.

We also provide a scatter plot of the predicted frailties and the number of patients enrolled at each institution in Figure 2. In general, it appears that the predicted frailties are larger with larger sample sizes. It may be the case that institutions with more patients have more practice with a trial protocol, which could lead to better performance.

The two most extreme standardized frailties in the Q-Q plot are marked at the bottom of Figure 1. The same institutions are marked in Figure 2. These institutions have more negative predicted frailty values than might be expected given the trends in the rest of the data. This might be due to the worse than normal outcomes in these two institutions. Specifically, among the 460 patients in this subset, 14.8% have died. However, in the institution with 22 patients, the failure rate was 27.3% and, in the institution with 10 patients, the failure rate was 60%.

8. Discussion

There is a need for flexible survival regression models that allow for marginal interpretations of treatment or exposure, while concurrently evaluating potential clustering.

The method proposed here establishes a general likelihood framework for this type of analysis. Marginal treatment or exposure effects are modeled with a proportional hazards model, while correlated transformed survival times are described by a shared frailty model.

In future work, we hope to extend the transformation in Equation (3) to normal distributions with variances that can depend on covariates. This type of model can allow for more explicit analyses into what is driving differences between clusters.

References

Anello, C., O'Neill, R., and Dubey, S. (2005), "Multicentre Trials: A US Regulatory Perspective," *Statistical Methods in Medical Research*, 14, 303–318.

Booth, J. and Hobert, J. (1998), "Standard Errors of Prediction in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 93, 262–272.

Cai, J. and Prentice, R. (1995), "Estimating Equations for Hazard Ratio Parameters Based on Correlated Failure Time Data," *Biometrika*, 82, 151–164.

Cai, J. and Shen, Y. (2000), "Permutation Tests for Comparing Marginal Survival Functions with Clustered Failure Time Data," *Statistics in Medicine*, 19, 2963–2973.

Cai, J., Zhou, H., and Davis, C. (1997), "Estimating the Mean Hazard Ratio Parameters for Clustered Survival Data with Random Clusters," *Statistics in Medicine*, 16, 2009–2020.

Cox, D. (1972), "Regression Models and Life-tables (with Discussion)," *Journal of the Royal Statistical Society, Series B (Methodological)*, 34, 187–220.

Fleiss, J. (1986), "Analysis of Data from Multiclinic Trials." *Controlled Clinical Trials*, 7, 267–275.

Grambsch, P. and Therneau, T. (1994), "Proportional Hazards Tests and Diagnostics Based on Weighted Residuals," *Biometrika*, 81, 515–526.

Gray, R. (1994), "A Bayesian Analysis of Institutional Effects in a Multicenter Cancer Clinical Trial," *Biometrics*, 244–253.

Jones, B., Teather, D., Wang, J., and Lewis, J. (1998), "A Comparison of Various Estimators of a Treatment Difference for a Multi-centre Clinical Trial," *Statistics in Medicine*, 17, 1767–1777.

Kalbfleisch, J. and Prentice, R. (2003), *The Statistical Analysis of Failure Time Data*, Wiley, 2nd ed.

Li, Y. and Lin, X. (2006), "Semiparametric Normal Transformation Models for Spatially Correlated Survival Data," *Journal of the American Statistical Association*, 101, 591–603.

Li, Y., Prentice, R., and Lin, X. (2008), "Semiparametric Maximum Likelihood Estimation in Normal Transformation Models for Bivariate Survival Data," *Biometrika*, 95, 947–960.

Lin, D., Wei, L., and Ying, Z. (1993), "Checking the Cox Model with Cumulative Sums of Martingale-based Residuals," *Biometrika*, 80, 557–572.

Logan, B., Nelson, G., and Klein, J. (2008), "Analyzing Center Specific Outcomes in Hematopoietic Cell Transplantation." *Lifetime Data Analysis*, 14, 389404.

Murphy, S. (1994), "Consistency in a Proportional Hazards Model Incorporating a Random Effect," *Annals of Statistics*, 22, 712–731.

— (1995), "Asymptotic Theory for the Frailty Model," *Annals of Statistics*, 23, 182–198.

Oakes, D. (1989), "Bivariate Survival Models Induced by Frailties," *Journal of the American Statistical Association*, 84, 487–493.

Parner, E. (1998), "Asymptotic Theory for the Correlated Gamma-frailty Model," *Annals of Statistics*, 26, 183–214.

Prentice, R. and Cai, J. (1992), "Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data," *Biometrika*, 79, 495–512.

Seibel, N., Steinherz, P., Sather, H., Nachman, J., Delaat, C., Ettinger, L., Freyer, D., Mattano Jr., L., Hastings, C., Rubin, C., Bertolone, K., Franklin, J., Heerema, N., Mitchell, T., Pysemany, A., La, M., Edens, C., and Gaynon, P. (2008), "Early Post-induction Intensification Therapy Improves Survival for Children and Adolescents with High-risk Acute Lymphoblastic Leukemia: A Report from the Children's Oncology Group," *Blood*, 11, 2548–2555.

Senn, S. (1998), "Some Controversies in Planning and Analysing Multi-centre Trials." *Statistics in Medicine*, 17, 1753–1765.

Spiekerman, C. and Lin, D. (1998), "Marginal Regression Models for Multivariate Failure Time Data," *Journal of the American Statistical Association*, 93, 1164–1175.

van der Vaart, A. (1998), *Asymptotic Statistics*, Cambridge University Press.

Vierron, E. and Giraudeau, B. (2007), "Sample Size Calculation for Multicenter Randomized Trial: Taking the Center Effect into Account," *Contemporary Clinical Trials*, 28, 451–458.

Wei, L., Lin, D., 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.

Zheng, L. and Zelen, M. (2008), "Multi-center Clinical Trials: Randomization and Ancillary Statistics," *Annals of Applied Statistics*, 2, 582–600.

Table 1: Summary of Simulation Results

Number of clusters = 30

		Median cluster size = 3				**Median cluster size = 17**			
	Truth	Estimate	SE_J	SE_S	SE_{MC}	Estimate	SE_J	SE_S	SE_{MC}
σ	0.5	0.531	0.117	–	0.107	0.559	0.070	–	0.067
$\boldsymbol{\beta}_1$	-0.693	-0.718	0.244	0.217	0.236	-0.709	0.112	0.106	0.109
$\boldsymbol{\beta}_2$	0.75	0.824	0.400	0.366	0.403	0.770	0.177	0.168	0.179
			R. Bias	SE_L	SE_{MC}		R. Bias	SE_L	SE_{MC}
Frailties			-0.099	0.911	0.759		-0.076	0.797	0.838

Number of clusters = 60

		Median cluster size = 3				**Median cluster size = 17**			
	Truth	Estimate	SE_J	SE_S	SE_{MC}	Estimate	SE_J	SE_S	SE_{MC}
σ	0.5	0.520	0.081	–	0.076	0.535	0.051	–	0.050
$\boldsymbol{\beta}_1$	-0.693	-0.710	0.163	0.157	0.157	-0.704	0.079	0.077	0.079
$\boldsymbol{\beta}_2$	0.75	0.749	0.275	0.260	0.282	0.762	0.125	0.120	0.124
			R. Bias	SE_L	SE_{MC}		R. Bias	SE_L	SE_{MC}
Frailties			-0.040	0.900	0.798		-0.067	0.773	0.886

Number of clusters = 90

		Median cluster size = 3				**Median cluster size = 17**			
	Truth	Estimate	SE_J	SE_S	SE_{MC}	Estimate	SE_J	SE_S	SE_{MC}
σ	0.5	0.512	0.066	–	0.063	0.528	0.042	–	0.042
$\boldsymbol{\beta}_1$	-0.693	-0.704	0.133	0.128	0.131	-0.700	0.064	0.063	0.065
$\boldsymbol{\beta}_2$	0.75	0.773	0.221	0.212	0.229	0.748	0.102	0.100	0.104
			R. Bias	SE_L	SE_{MC}		R. Bias	SE_L	SE_{MC}
Frailties			-0.032	0.896	0.811		-0.055	0.767	0.906

Estimate = average of estimates
SE_J = average of jackknifed based standard errors
SE_S = average of sandwich-formula based standard errors (only for $\boldsymbol{\beta}$)
SE_L = average of likelihood based standard errors
SE_{MC} = standard deviation of estimates
R. Bias = median relative bias, where relative bias = bias/parameter estimate

Table 2: Treatment arms for subjects of CCG-1961 with enlarged livers

Arm	Strength	Duration	N
A	Standard	Standard	119
B	Standard	Double	104
C	Increased	Standard	117
D	Increased	Double	120

Table 3: Data Analysis Results

Parameter	Estimate	SE_J	SE_S	Naïve SE
Treatment Only Model				
Covariates				
Increased Strength	-0.430	0.231	0.225	0.247
Increased Duration	0.419	0.231	0.226	0.245
Frailty Variance				
σ	0.171	0.152	–	–
Larger Model				
Covariates				
Increased Strength	-0.418	0.236	0.229	0.247
Increased Duration	0.426	0.247	0.236	0.248
Gender (ref=males)	-0.101	0.254	0.243	0.251
Age				
1–9	(ref)			
10–15	0.415	0.253	0.244	0.262
16+	0.394	0.491	0.430	0.384
Platelets ($\times 10^3/\text{mm}^3$)				
1–49	(ref)			
50–150	0.386	0.303	0.288	0.258
150+	-0.118	0.643	0.568	0.529
Frailty Variance				
σ	0.175	0.144	–	–

Estimate = log hazard ratios and estimates of σ
SE_J = average of jackknife based standard errors
SE_S = average of sandwich-formula based standard errors (only for $\boldsymbol{\beta}$)
Naïve SE = SE assuming independence

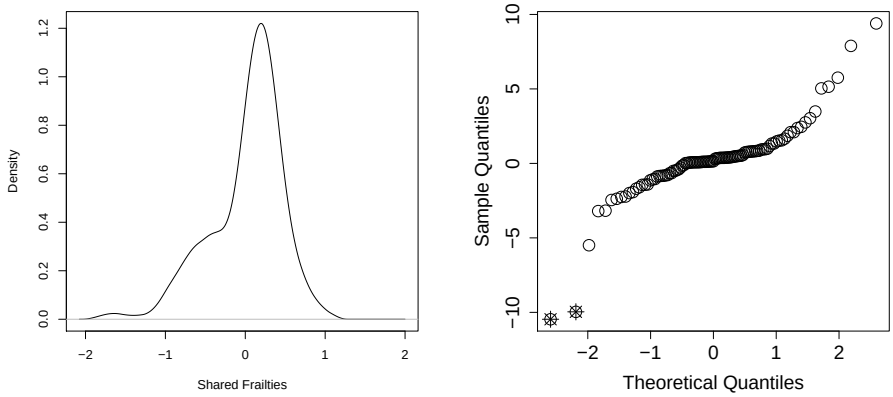

Figure 1: Left figure: smoothed density of predicted frailties. Right figure: Q-Q plot of predicted frailties and standard normal distribution quantiles

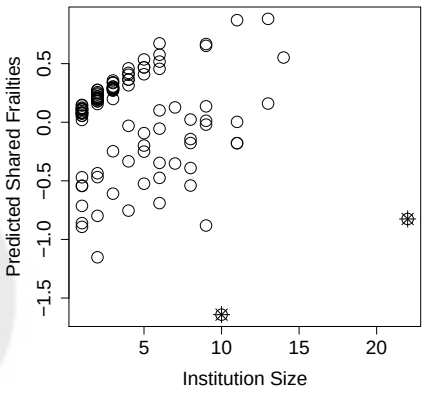

Figure 2: A scatterplot of predicted frailties and institution size

Appendix A. Regularity Conditions and Notation

Assume the following regularity conditions where $\tau > 0$ is a constant (for example, study duration):

C.1: $\boldsymbol{\beta}$ is in a compact subset of $\mathbb{R}^p$

C.2: $\Lambda(\tau) < \infty$

C.3: $\sigma \in \nu$, where ν is a compact subset of $(0, 1)$

C.4: $P(C_{ij} \geq t \, \forall t \in [0, \tau] \mid \mathbf{Z}_{ij}) > \delta_c > 0$ for $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$

C.5: Write $\mathbf{Z}_{ij}(t) = \{Z_{ij1}(t), \ldots, Z_{ijp}(t)\}$. $|Z_{ijk}(0)| + \int_0^\tau |dZ_{ijk}(t)| \leq B_Z < \infty$ almost surely for some constant B_Z and $i = 1, \ldots, m$, $j = 1, \ldots, n_i$, $k = 1, \ldots, p$

C.6: $E \log\{L(\sigma_1; \boldsymbol{\beta}, \Lambda)/L(\sigma_2; \boldsymbol{\beta}, \Lambda)\}$ exists for all σ_1, $\sigma_2 \in (0, 1)$

C.7: Let $Y_{ij}(t) = I(X_{ij} \geq t)$, $K = \max_i n_i$, $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}'\mathbf{a}$,

$$Q_j^{(\kappa)}(\boldsymbol{\beta}, t) = m^{-1} \sum_{i=1}^m Y_{ij}(t) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(t)\} \mathbf{Z}_{ij}(t)^{\otimes \kappa}, \;\; q_j^{(\kappa)}(\boldsymbol{\beta}, t) = E Q_j^{(\kappa)}(\boldsymbol{\beta}, t),$$
$$\eta_j(\boldsymbol{\beta}, t) = \frac{q_j^{(1)}(\boldsymbol{\beta}, t)}{q_j^{(0)}(\boldsymbol{\beta}, t)}, \;\; \varrho_j(\boldsymbol{\beta}, t) = \frac{q_j^{(2)}(\boldsymbol{\beta}, t)}{q_j^{(0)}(\boldsymbol{\beta}, t)} - \eta_k(\boldsymbol{\beta}, t)^{\otimes 2} \text{ for } j = 1, \ldots, K.$$

Assume $\sum_{j=1}^K \int_0^\tau \varrho_j(\boldsymbol{\beta}_0, t) q_j^{(0)}(\boldsymbol{\beta}_0, t) \lambda_0(t) \, dt$ is positive definite

Condition C.3 allows us to avoid boundary issues. Condition C.5 assumes that all the covariates are of bounded variation, which is necessary to ensure the Hadamard differentiability of the likelihood and score function. Condition C.6 is useful to help prove that the expected likelihood is maximized at σ_0. Condition C.7 is a technical condition from Spiekerman and Lin (1998) that is needed for the results for $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}$.

To simplify the presentation of the proofs we define several terms. Define

$$L^*(\sigma, \boldsymbol{\beta}, \Lambda) = \prod_{i=1}^{m} \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i}) \tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{\mathbf{X}}_i^{n_i - \Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i}),$$

where $L^*(\sigma, \boldsymbol{\beta}, \Lambda) = c^* L(\sigma, \boldsymbol{\beta}, \Lambda)$ and c^* does not depend on σ. Let

$$l_{m0}(\sigma) = m^{-1} \log L^*(\sigma, \boldsymbol{\beta}_0, \Lambda_0),\ l_m(\sigma) = m^{-1} \log L^*(\sigma, \boldsymbol{\beta}, \Lambda),$$
$$\hat{l}_m(\sigma) = m^{-1} \log L^*(\sigma, \hat{\boldsymbol{\beta}}, \hat{\Lambda}),\ U_{m0}(\sigma) = \partial l_{m0}(\sigma)/\partial\sigma,\ U_{m0}(\sigma) = \partial l_m(\sigma)/\partial\sigma,$$
$$\text{and } \hat{U}_m(\sigma) = \partial \hat{l}_m(\sigma)/\partial\sigma$$

Expectations are with respect to the true distributions of all random variables involved. Let $\|\cdot\|$ denote the Euclidean norm and let $\|\cdot\|_\infty$ denote the supremum norm on $[0, \tau]$. Let $BV[0, \tau]$ denote the class of functions with bounded total variation on $[0, \tau]$. Let single superscript dots denote first derivatives and double superscript dots denote second derivatives.

Appendix B. Proof and Associated Lemmas for Theorem 1

For ease of presentation we state several lemmas used in the proof of Theorem 1, but defer their proof until the end of the Appendix.

To account for the fact that plug-in estimates of $\boldsymbol{\beta}$ and Λ are used in the likelihood for σ, we will need to take a Taylor series expansion of the likelihood of σ around $\boldsymbol{\beta}_0$ and Λ_0. Since Λ_0 is an unspecified function, this expansion will need to include a functional expansion term. An expansion using Hadamard derivatives is appropriate for this situation. Hence, in order to use the functional expansion, we need to verify that the log-likelihood is Hadamard differentiable with respect to Λ, which is done in Lemma 1.

Lemma 1. *Under conditions C.1–C.5, the log-likelihood $l_m(\sigma)$ is Hadamard differentiable with respect to Λ.*

After we have an expansion of the log-likelihood we will need the first order terms to be bounded by a random variable with finite expectation. We provide this verification in Lemma 2.

Lemma 2. *Write the Hadamard derivative of $l_m(\sigma)$ with respect to Λ at $\Pi \in BV[0,\tau]$ as $\int_0^\tau \zeta_m(\Lambda,\sigma)(u)\, d\Pi(u)$ and let $\boldsymbol{\zeta}_m(\boldsymbol{\beta},\sigma) = \partial l_m(\sigma)/\partial\boldsymbol{\beta}$. Under conditions C.1–C.5, $\|\zeta_m(\Lambda,\sigma)\|_\infty$ and $\|\boldsymbol{\zeta}_m(\boldsymbol{\beta},\sigma)\|$ are bounded. Expressions for $\boldsymbol{\zeta}_m(\boldsymbol{\beta},\sigma)$ and $\zeta_m(\Lambda,\sigma)$ are provided in the proof.*

In order to prove consistency of $\hat{\sigma}$ we will need to verify the uniform convergence of the log-likelihood with plug-in estimates of $\boldsymbol{\beta}$ and Λ to the expected value of the log-likelihood evaluated at the truth. We accomplish this, using the results of Lemmas 1 and 2, in Lemma 3.

Lemma 3. *Under conditions C.1–C.5, as $m \to \infty$*

$$\sup_{\sigma\in\nu} |\hat{l}_m(\sigma) - El_{m0}(\sigma)| = o_p(1).$$

Finally, in order to verify that $\hat{\sigma}$ is consistent, we will need to show that the expected log-likelihood is maximized at the truth, which is done in Lemma 4.

Lemma 4. *Under conditions C.1–C.6,*

$$El_{m0}(\sigma) - El_{m0}(\sigma_0) < 0.$$

Proof of Theorem 1

The results for $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}$ follow from arguments along the lines of Spiekerman and Lin (1998). We use the results of Lemmas 3 and 4 to prove the result for $\hat{\sigma}$.

Since $\hat{\sigma}$ maximizes $\hat{l}_m(\sigma)$, Lemma 3 implies that

$$\begin{aligned} 0 \leq \hat{l}_m(\hat{\sigma}) - \hat{l}_m(\sigma_0) &= \hat{l}_m(\hat{\sigma}) - \hat{l}_m(\sigma_0) + El_{m0}(\sigma_0) - El_{m0}(\sigma_0) \\ &= \hat{l}_m(\hat{\sigma}) - El_{m0}(\sigma_0) + o_p(1). \end{aligned}$$

Therefore $El_{m0}(\sigma_0) \leq \hat{l}_m(\hat{\sigma}) + o_p(1)$. Subtract $El_{m0}(\hat{\sigma})$ from each side of the inequality to write

$$\begin{aligned} El_{m0}(\sigma_0) - El_{m0}(\hat{\sigma}) &\leq \hat{l}_m(\hat{\sigma}) - El_{m0}(\hat{\sigma}) + o_p(1) \\ &\leq \sup_{\sigma \in \nu} |\hat{l}_m(\sigma) - El_{m0}(\sigma)| + o_p(1) = o_p(1), \end{aligned} \tag{8}$$

where the last equality comes from Lemma 3.

Take σ such that $|\sigma - \sigma_0| \geq \varepsilon$ for any fixed $\varepsilon > 0$. By Lemma 4 there must exist some $\gamma_\varepsilon > 0$ such that $El_{m0}(\sigma) + \gamma_\varepsilon < El_{m0}(\sigma_0)$. It follows that $P(|\hat{\sigma} - \sigma_0| \geq \varepsilon) \leq P\{El_{m0}(\hat{\sigma}) + \gamma_\varepsilon < El_{m0}(\sigma_0)\}$. Equation (8) implies that $P\{El_{m0}(\hat{\sigma}) + \gamma_\varepsilon < El_{m0}(\sigma_0)\}$ converges to 0 as $m \to \infty$. Therefore $P(|\hat{\sigma} - \sigma_0| \geq \varepsilon)$ converges to 0 as $m \to \infty$. □

Appendix C. Proof and Associated Lemmas for Theorem 2

For ease of presentation we state several lemmas used in the proof of Theorem 2, but defer their proof until the end of the Appendix.

To account for the fact that plug-in estimates of $\boldsymbol{\beta}$ and Λ are used in the likelihood and score function for σ, we will need to take a Taylor series expansion of the score function for σ around $\boldsymbol{\beta}_0$ and Λ_0. We verify that the score function is Hadamard

differentiable with respect to Λ, which is done in Lemma 5.

Lemma 5. *Under conditions C.1–C.5, the score function $U_m(\sigma)$ is Hadamard differentiable with respect to Λ.*

After we have an expansion of the score function for σ, we will need the first order terms to be bounded by a random variable with finite expectation. We provide this verification in Lemma 6.

Lemma 6. *Write the Hadamard derivative of $U_m(\sigma)$ with respect to Λ at $\Pi \in BV[0,\tau]$ as $\int_0^\tau \xi_m(\sigma,\Lambda)(u)\, d\Pi(u)$ and let $\boldsymbol{\xi}_m(\boldsymbol{\beta},\sigma) = \partial U_m(\sigma)/\partial\boldsymbol{\beta}$. Under conditions C.1–C.5, $\|\xi_m(\sigma,\Lambda)\|_\infty$ and $\|\boldsymbol{\xi}_m(\sigma,\boldsymbol{\beta})\|$ are bounded. Expressions for $\boldsymbol{\xi}_m(\sigma,\boldsymbol{\beta})$ and $\xi_m(\sigma,\Lambda)$ are provided in the proof.*

Proof of Theorem 2

The result that $\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges to mean zero normal distribution and that $\sqrt{m}(\hat{\Lambda} - \Lambda_0)$ converges to mean zero Guassian process follows from arguments along the lines of Spiekerman and Lin (1998). This proof needs to verify that $\sqrt{m}(\hat{\sigma} - \sigma_0)$ converges to a normal distribution with mean zero after accounting for the extra variance induced by the profile estimation procedure. The variance of $\hat{\sigma}$ should be inflated over a model where $\boldsymbol{\beta}_0$ and Λ_0 are used to take into account the estimation of $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}$.

First we will show that the score equation associated with $\hat{l}_m$ evaluated at σ_0 follows a normal distribution. This result coupled with a first order expansion of the score equation associated with $\hat{l}_m$ around σ_0 will finish the proof.

Using Lemma 5, a Taylor series expansion of $\hat{U}_m(\sigma)$ around $\boldsymbol{\beta}_0$ and Λ_0 gives

$$\hat{U}_m(\sigma_0) = U_m(\sigma_0) + \int_0^\tau \xi_m(\sigma_0,\Lambda)(t)\, d\{\hat{\Lambda}(t) - \Lambda_0(t)\} + \boldsymbol{\xi}_m(\sigma_0,\boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + G_m,$$

where G_m is a remainder term for the Taylor series. Since $\hat{\Lambda}$ and $\hat{\boldsymbol{\beta}}$ are $\sqrt{m}$-consistent it can be shown that $G_m = o_p(m^{-1/2})$. Define the pointwise limit of $\xi_m(\sigma, \Lambda)(t)$ as $\xi(\sigma, \Lambda)(t)$ and let $\boldsymbol{\xi}(\sigma, \boldsymbol{\beta}) = E\boldsymbol{\xi}_m(\sigma, \boldsymbol{\beta})$. From Lemma 6, $\|\boldsymbol{\xi}(\sigma_0, \Lambda)\|_\infty$ and $\|\boldsymbol{\xi}(\sigma, \boldsymbol{\beta})\|$ are bounded. It follows that

$$\begin{aligned}\sqrt{m}\hat{U}_m(\sigma_0) = \sqrt{m}U_m(\sigma_0) + \sqrt{m}\int_0^\tau \xi(\sigma_0, \Lambda)(t)\, d\{\hat{\Lambda}(t) - \Lambda_0(t)\} \\ +\sqrt{m}\,\boldsymbol{\xi}(\sigma_0, \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1). \qquad (9)\end{aligned}$$

Using the results of Spiekerman and Lin (1998), we can write Equation (9) as a sum of independent and identically distributed random variables, $\sqrt{m}\sum_{i=1}^m \Xi_i$, where $E\Xi_1 = 0$ and $V\Xi_1 < \infty$. The central limit theorem implies that $\sqrt{m}\hat{U}_m(\sigma_0)$ converges to a normally distributed random variable with mean zero and variance equal to the variance of Ξ_1.

Next, we take a first order Taylor series expansion of $\hat{U}_m(\hat{\sigma})$ around σ_0:

$$\hat{U}_m(\hat{\sigma}) = \hat{U}_m(\sigma_0) + (\hat{\sigma} - \sigma_0)\hat{W}_m(\sigma^*),$$

where $\hat{W}_m(\sigma) = \partial\hat{U}_m(\sigma)/\partial\sigma$ and σ^* is between $\hat{\sigma}$ and σ_0. It must be the case that $\hat{U}_m(\hat{\sigma}) = 0$ since $\hat{\sigma}$ was taken to be the maximum of $L(\sigma, \hat{\boldsymbol{\beta}}, \hat{\Lambda})$. Theorem 1 showed that $\hat{\sigma}$ consistently estimates σ_0, so the the law of large numbers implies that $\hat{W}_m(\sigma^*)$ converges in probability to $W(\sigma_0) = \lim_{m\to\infty} W_m(\sigma_0)$. Finally, using the central limit theorem and Slutsky's theorem, $\sqrt{m}(\hat{\sigma} - \sigma_0)$ converges to a normal distribution with mean zero and variance equal to $W(\sigma_0)^{-2}V(\Xi_1)$. □

Appendix D. Proofs of Lemmas

Proof of Lemma 1

Define $Y_{ij}(t) = I(X_{ij} \geq t)$. The log-likelihood can be written

$$l_m(\sigma) = m^{-1} \sum_{i=1}^{m} \log \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i}) + \log \tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{\mathbf{X}}_i^{n_i - \Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})$$

where $\tilde{X}_{ij} = \tilde{\Phi}^{-1}(\exp[-\int_0^\tau Y_{ij}(u) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(u)\} \, d\Lambda(u)])$. By condition C.5 the term

$$\int_0^\tau Y_{ij}(u) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(u)\} \, d\Lambda(u)$$

is Hadamard differentiable. Using multiple iterations of the chain rule for Hadamard derivatives (van der Vaart 1998, Theorem 20.9), we conclude that $l_m(\sigma)$ is Hadamard differentiable. □

Proof of Lemma 2

First we find expressions for $\zeta_m(\sigma, \Lambda)$ and $\boldsymbol{\zeta}_m(\boldsymbol{\beta}, \sigma)$, starting with $\zeta_m(\sigma, \Lambda)$. To make the argument more concrete express $l_m(\sigma)$ as a function of Λ by writing $l_m(\sigma, \Lambda) = l_m(\sigma)$. Let $\Gamma \in BV[0, \tau]$. Denote

$$H_{ij} = \exp[-\int_0^\tau Y_{ij}(u) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(u)\} \, d\Lambda(u)].$$

By conditions C.1 and C.2, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$, $H_{ij} > 0$ and $|\tilde{X}_{ij}| < B^* < \infty$ for some constant B^*.

To find the expression for the derivative, take a Taylor series expansion of $l_m\{\sigma, \Lambda +$

$t(\Gamma - \Lambda)\}$ around $t = 0$ and evaluate the result at $t = 1$. The final expression is

$$l_m(\sigma, \Gamma) = l_m(\sigma, \Lambda) + \int_0^\tau \zeta_m(\sigma, \Lambda)(u)\, d(\Lambda - \Gamma)(u),$$

where $\zeta_m(\sigma, \Lambda)(u)$ is equal to $m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} D_{ij}^l Y_{ij}(u) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(u)\} H_{ij}$ and D_{ij}^l is equal to

$$\begin{aligned}(\Delta_{ij}[\phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})^{-1}\{\partial \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})/\partial \tilde{X}_{ij}\}] + (1 - \Delta_{ij})[\tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{\mathbf{X}}_i^{n_i - \Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})^{-1} \\ \times \{\partial \tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{\mathbf{X}}_i^{n_i - \Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})/\partial \tilde{X}_{ij}\}]) \sum_{j=1}^{n_i} \partial \Phi^{-1}(H_{ij})/\partial H_{ij}.\end{aligned}$$

Therefore the Hadamard derivative for $\Pi \in BV[0, \tau]$ is $\int_0^\tau \zeta_m(\sigma, \Lambda)(u)\, d\Pi(u)$. Direct calculation verifies that $\boldsymbol{\zeta}_m(\sigma, \boldsymbol{\beta})$ is equal to

$$m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} D_{ij}^l [\int_0^\tau Y_{ij}(u) \mathbf{Z}_{ij}(u) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(u)\}\, d\Lambda(u)] H_{ij}.$$

We need to check whether each of the terms in D_{ij}^l is bounded and also that the terms unique to $\boldsymbol{\zeta}_m(\sigma, \boldsymbol{\beta})$ and $\zeta_m(\sigma, \Lambda)$ are bounded. First,

$$\phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i}) = (2\pi)^{-\Delta_i/2} \det(\boldsymbol{\Sigma}_{i11})^{-1/2} \exp(-\tilde{\mathbf{X}}_i^{\Delta_i'} \boldsymbol{\Sigma}_{i11}^{-1} \tilde{\mathbf{X}}_i^{\Delta_i}/2) > 1/B_1 > 0$$

for some constant B_1 since for $\tilde{X}_{ij} \in \tilde{\mathbf{X}}_i^{\Delta_i}$, $|\tilde{X}_{ij}| < B^*$. Therefore, for $i = 1, \ldots, m$, $\phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})^{-1} < B_1 < \infty$.

Let $\mathbf{w}_\alpha(j)$ denote the vector of length α where the j^{th} element is 1 and the rest of the vector is 0. Using the chain rule, for $j = 1, \ldots, \Delta_i$ and $i = 1, \ldots, m$,

$$\partial \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})/\partial \tilde{X}_{ij} = -\tilde{\mathbf{X}}_i^{\Delta_i'} \boldsymbol{\Sigma}_{i11}^{-1} \mathbf{w}_{\Delta_i}(j) \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i}).$$

The multivariate normal density $\phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})$ is bounded and for $\tilde{X}_{ij} \in \tilde{\mathbf{X}}_i^{\Delta_i}$, $|\tilde{X}_{ij}| < B^*$. Hence, for $j = 1, \ldots, \Delta_i$ and $i = 1, \ldots, m$, $|\partial \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})/\partial \tilde{X}_{ij}| < B_2 < \infty$ for some constant B_2.

Next consider $\tilde{\Phi}_c^{n_i-\Delta_i}(\tilde{\mathbf{X}}_i^{n_i-\Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})$, which for $i = 1, \ldots, m$ is equal to

$$\int_{\mathbf{M}_i} (2\pi)^{n_i-\Delta_i} \det(\tilde{\boldsymbol{\Sigma}}_i) \exp\{(\mathbf{t}^{n_i-\Delta_i} - \tilde{\boldsymbol{\mu}}_i)' \tilde{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{t}^{n_i-\Delta_i} - \tilde{\boldsymbol{\mu}}_i)/2\} \, d\mathbf{t}^{n_i-\Delta_i}$$

where $\mathbf{M}_i = \{t_{(\Delta_i+1)} > \tilde{X}_{i,(\Delta_i+1)}, \ldots, t_{n_i} > \tilde{X}_{i,n_i}\}$, $\mathbf{t}^{n_i-\Delta_i} = (t_{(\Delta_i+1)}, \ldots, t_{n_i})$, $\tilde{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}_{i22} - \boldsymbol{\Sigma}'_{i21} \boldsymbol{\Sigma}_{i11}^{-1} \boldsymbol{\Sigma}_{i12}$, and $\tilde{\boldsymbol{\mu}}_i = \boldsymbol{\Sigma}_{i21} \boldsymbol{\Sigma}_{i11}^{-1} \tilde{\mathbf{X}}_i^{\Delta_i}$. Since $|\tilde{X}_{ij}| < B^*$ for $\tilde{X}_{ij} \in \tilde{\mathbf{X}}_i^{n_i-\Delta_i}$, it must be the case that for $i = 1, \ldots, m$. $|\tilde{\Phi}_c^{n_i-\Delta_i}(\tilde{\mathbf{X}}_i^{n_i-\Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})^{-1}| < B_3 < \infty$ for some constant B_3.

Let $\mathbf{t}_j^{n_i-\Delta_i}$ be equal to $\mathbf{t}^{n_i-\Delta_i}$ but with the component corresponding to $(j - \Delta_i)^{th}$ component replaced by $\tilde{X}_{ij}$. Let $\mathbf{t}_{-j}^{n_i-\Delta_i}$ be equal to $\mathbf{t}^{n_i-\Delta_i}$ but with the $(j - \Delta_i)^{th}$ element removed. Let $\mathbf{M}_{i,-j}$ denote $\mathbf{M}_i$ but with the $(j - \Delta_i^{th})$ inequality removed. Consider $|\partial \tilde{\Phi}_c^{n_i-\Delta_i}(\tilde{\mathbf{X}}_i^{n_i-\Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})/\partial \tilde{X}_{ij}|$, which, for $j = \Delta_i + 1, \ldots, n_i$, $i = 1 \ldots, m$, can be written

$$\left| \int_{\mathbf{M}_i} -(2\pi)^{n_i-\Delta_i} \det(\tilde{\boldsymbol{\Sigma}}_i) \exp\{(\mathbf{t}_j^{n_i-\Delta_i} - \tilde{\boldsymbol{\mu}}_i)' \tilde{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{t}_j^{n_i-\Delta_i} - \tilde{\boldsymbol{\mu}}_i)/2\} \, d\mathbf{t}_{-j}^{n_i-\Delta_i} \right| < B_4$$

for some constant $B_4 < \infty$ since $|\tilde{X}_{ij}| < B^*$ for $\tilde{X}_{ij} \in \tilde{\mathbf{X}}_i^{n_i-\Delta_i}$.

Using the definition of the derivative of an inverse function, $\partial \Phi^{-1}(H_{ij})/\partial H_{ij} = -[\phi\{\Phi^{-1}(H_{ij})\}]^{-1}$, where ϕ is the density of the standard normal distribution and Φ^{-1} is the inverse of the distribution function of the standard normal distribution. Since $|\tilde{X}_{ij}| < B^*$, $0 < B_5 < H_{ij} < B_6 < 1$ for some constants B_5 and B_6. Therefore, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$, $|\partial \Phi^{-1}(H_{ij})/\partial H_{ij}| < B_7 < \infty$ for some constant B_7. By condition C.5, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$, $\|Y_{ij} \exp(\boldsymbol{\beta}' \mathbf{Z}_{ij})\|_\infty < B8 <$

∞ and $\| \int_0^\tau Y_{ij}(u)\mathbf{Z}_{ij}(u)\exp\{\boldsymbol{\beta}'\mathbf{Z}_{ij}(u)\}\, d\Lambda(u)]\| < B_9 < \infty$ for some constants B_8 and B_9. Hence $\|\zeta_m(\sigma,\Lambda)\|_\infty$ and $\|\boldsymbol{\zeta}_m(\boldsymbol{\beta},\sigma)\|$ are bounded are bounded by $(B_1B_2 + B_3B_4)B_7(B_8+B_9) < \infty$. □

Proof of Lemma 3

An expansion of $\hat{l}_m(\sigma)$ around Λ_0 and $\boldsymbol{\beta}_0$ can be written:

$$\hat{l}_m(\sigma) = l_{m0}(\sigma) + \boldsymbol{\zeta}_m(\boldsymbol{\beta},\sigma)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \int_0^\tau \zeta_m(\sigma,\Lambda)(t)\, d(\hat{\Lambda} - \Lambda_0)(t) + R,$$

where R is a remainder term of order $o_p\{\max(\|\hat{\Lambda} - \Lambda_0\|_\infty, \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|)\}$ and $\boldsymbol{\zeta}_m(\boldsymbol{\beta},\sigma)$ and $\zeta_m(\sigma,\Lambda)(t)$ are defined in Lemma 2. Since $\hat{\Lambda}$ is uniformly consistent and $\hat{\boldsymbol{\beta}}$ is consistent (Spiekerman and Lin 1998), $R = o_p(1)$. The result follows from the law of large numbers, the uniform consistency of $\hat{\Lambda}$, the consistency of $\hat{\boldsymbol{\beta}}$, and the fact that $\|\boldsymbol{\zeta}_m(\boldsymbol{\beta},\sigma)\|$ and $\|\zeta_m(\sigma,\Lambda)\|_\infty$ are bounded (Lemma 2). □

Proof of Lemma 4

The log-likelihood, $l_m(\sigma)$, can be written as a sum of independent and identically distributed random variables $m^{-1}\sum_{i=1}^m \varphi_i(\sigma)$. Take $\sigma \neq \sigma_0$. The law of large numbers and Jensen's inequality imply that $El_{m0}(\sigma) - El_{m0}(\sigma_0) = \lim_{m\to\infty} l_{m0}(\sigma) - l_{m0}(\sigma_0)$ which is strictly less than $\log[E(L^*(\sigma,\boldsymbol{\beta}_0,\Lambda_0)/L^*(\sigma_0,\boldsymbol{\beta}_0,\Lambda_0))] = 0$. □

Proof of Lemma 5

Let $N(\mathbf{t}, d, \boldsymbol{\mu}, \boldsymbol{\Sigma}^\dagger)$ be defined as $(2\pi)^{-d/2}\det(\boldsymbol{\Sigma}^\dagger)^{-1/2}\exp\{-(\mathbf{t}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}^\dagger)^{-1}(\mathbf{t}-\boldsymbol{\mu})/2\}$ $[tr\{(\boldsymbol{\Sigma}^\dagger)^{-1}\tilde{\mathbf{W}}_d\} - \{-(\mathbf{t}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}^\dagger)^{-1}\tilde{\mathbf{W}}_d(\boldsymbol{\Sigma}^\dagger)^{-1}(\mathbf{t}-\boldsymbol{\mu})/2\}]/2$, where $\tilde{\mathbf{W}}_d$ is the d dimensional square matrix with zeros along the diagonal and ones off the diagonal. Let

$\mathbf{0}^d$ denote a vector of length d of zeros. The score function can be written

$$U_m(\sigma) = m^{-1} \sum_{i=1}^{m} \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})^{-1} N(\tilde{\mathbf{X}}_i^{\Delta_i}, \Delta_i, \mathbf{0}^{\Delta_i}, \boldsymbol{\Sigma}_{i11})$$
$$+\tilde{\Phi}_c^{n_i-\Delta_i}(\tilde{\mathbf{X}}_i^{n_i-\Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})^{-1} \int_{\mathbf{M}_i} N(\mathbf{t}^{n_i-\Delta_i}, n_i - \Delta_i, \tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i)\, d\mathbf{t}^{n_i-\Delta_i}.$$

Using the results of Lemma 1 and multiple iterations of the chain rule for Hadamard derivatives (van der Vaart 1998, Theorem 20.9), we conclude that $U_m(\sigma)$ is Hadamard differentiable. □

Proof of Lemma 6

First we find expressions for $\xi_m(\sigma, \Lambda)$ and $\boldsymbol{\xi}_m(\sigma, \boldsymbol{\beta})$, starting with $\xi_m(\sigma, \Lambda)$. To make the argument more concrete express $U_m(\sigma)$ as a function of Λ by writing $U_m(\sigma, \Lambda) = U_m(\sigma)$. Let $\Gamma \in BV[0, \tau]$.

To find the expression for the derivative, take a Taylor series expansion of $U_m\{\sigma, \Lambda + t(\Gamma - \Lambda)\}$ around $t = 0$ and evaluate the result at $t = 1$. The final expression is $U_m(\sigma, \Gamma) = U_m(\sigma, \Lambda) + \int_0^\tau \xi_m(\sigma, \Lambda)(u)\, d(\Lambda - \Gamma)(u)$, where $\xi_m(\sigma, \Lambda)(u)$ is equal to $m^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} D_{ij}^U Y_{ij}(u) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(u)\} H_{ij}$ and

$$D_{ij}^U = (\Delta_{ij}[\{\partial \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})^{-1} / \partial \tilde{X}_{ij}\} N(\tilde{\mathbf{X}}_i^{\Delta_i}, \Delta_i, \mathbf{0}^{\Delta_i}, \boldsymbol{\Sigma}_{i11}) + \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})^{-1}$$
$$\times \{\partial N(\tilde{\mathbf{X}}_i^{\Delta_i}, \Delta_i, \mathbf{0}^{\Delta_i}, \boldsymbol{\Sigma}_{i11}) / \partial \tilde{X}_{ij}\}] + (1 - \Delta_{ij})[\{\partial \tilde{\Phi}_c^{n_i-\Delta_i}(\tilde{\mathbf{X}}_i^{n_i-\Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})^{-1} / \partial \tilde{X}_{ij}\}$$
$$\times \int_{\mathbf{M}_i} N(\mathbf{t}^{n_i-\Delta_i}, n_i - \Delta_i, \tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i)\, d\mathbf{t}^{n_i-\Delta_i} + \tilde{\Phi}_c^{n_i-\Delta_i}(\tilde{\mathbf{X}}_i^{n_i-\Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})^{-1}$$
$$\times \{\partial \int_{\mathbf{M}_i} N(\mathbf{t}^{n_i-\Delta_i}, n_i - \Delta_i, \tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i)\, d\mathbf{t}^{n_i-\Delta_i} / \partial \tilde{X}_{ij}\}]) \sum_{j=1}^{n_i} \partial \Phi^{-1}(H_{ij}) / \partial H_{ij}$$

Therefore the Hadamard derivative for $\Pi \in BV[0, \tau]$ is $\int_0^\tau \xi_m(\sigma, \Lambda)(u)\, d\Pi(u)$. Direct

calculation verifies that $\boldsymbol{\xi}_m(\sigma, \boldsymbol{\beta})$ is equal to

$$m^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} D_{ij}^U [\int_0^\tau Y_{ij}(u) \mathbf{Z}_{ij}(u) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(u)\} \, d\Lambda(u)] H_{ij}.$$

In Lemma 2 we showed that, for $i = 1, \ldots, m$, $|\phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})^{-1}| < B1 < \infty$ and $|\tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{\mathbf{X}}_i^{n_i - \Delta_i})^{-1}| < B_3 < \infty$. Also, for $j = 1, \ldots, n_i$, $i = 1, \ldots, m$, $|\partial \Phi^{-1}(H_{ij})/\partial H_{ij}| < B_7 < \infty$, $\|Y_{ij} \exp(\boldsymbol{\beta}' \mathbf{Z}_{ij})\|_\infty < B8 < \infty$ and $\| \int_0^\tau Y_{ij}(u) \mathbf{Z}_{ij}(u) \exp\{\boldsymbol{\beta}' \mathbf{Z}_{ij}(u)\} \, d\Lambda(u)]\| < B_9 < \infty$.

We tackle each of the remaining terms. First, using results from Lemma 2, for $j = 1, \ldots, \Delta_i$, $i = 1, \ldots, m$, $|\partial \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})^{-1} / \partial \tilde{X}_{ij}|$ is equal to $|-\phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})^{-2} \{\partial \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i}) / \partial \tilde{X}_{ij}\}| < B_{11} = B_1^2 B_2 < \infty$ for some constant B_{11}.

Since $\boldsymbol{\Sigma}_{i11}$ has an exchangeable structure, $tr\{\boldsymbol{\Sigma}_{i11}^{-1} \tilde{\mathbf{W}}_{\Delta_i}\}$ and $\det(\boldsymbol{\Sigma}_{i11})^{-1/2}$ are both bounded by some constant $B_{10} < \infty$. Therefore for $i = 1, \ldots, m$, $|N(\tilde{\mathbf{X}}_i^{\Delta_i}, \Delta_i, \mathbf{0}^{\Delta_i}, \boldsymbol{\Sigma}_{i11})| < B_{12} < \infty$ for some constant B_{12}.

Next, we consider $|\partial N(\tilde{\mathbf{X}}_i^{\Delta_i}, \Delta_i, \mathbf{0}^{\Delta_i}, \boldsymbol{\Sigma}_{i11}) / \partial \tilde{X}_{ij}|$ for $j = 1, \ldots, \Delta_i$ and $i = 1, \ldots, m$, which is equal to

$$|\tilde{\mathbf{X}}_i^{\Delta_i'} \boldsymbol{\Sigma}_{i11}^{-1} \mathbf{w}_{\Delta_i}(j) N(\tilde{\mathbf{X}}_i^{\Delta_i}, \Delta_i, \mathbf{0}^{\Delta_i}, \boldsymbol{\Sigma}_{i11}) + \tilde{\mathbf{X}}_i^{\Delta_i'} \boldsymbol{\Sigma}_{i11}^{-1} \tilde{\mathbf{W}}_{\Delta_i} \boldsymbol{\Sigma}_{i11}^{-1} \mathbf{w}_{\Delta_i}(j) \phi_u^{\Delta_i}(\tilde{\mathbf{X}}_i^{\Delta_i})|,$$

and, by the results of the previous paragraph and the results of Lemma 2, is bounded by some constant $B_{13} < \infty$.

Using results from Lemma 2, for $j = \Delta_i + 1, \ldots, n_i$ and $i = 1, \ldots, m$, $|\partial \Phi_c^{n_i - \Delta_i}(\tilde{\mathbf{X}}_i^{n_i - \Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})^{-1} / \partial \tilde{X}_{ij}|$ is equal to

$$|-\Phi_c^{n_i - \Delta_i}(\tilde{\mathbf{X}}_i^{n_i - \Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i})^{-2} \{\partial \Phi_c^{n_i - \Delta_i}(\tilde{\mathbf{X}}_i^{n_i - \Delta_i} \mid \tilde{\mathbf{X}}_i^{\Delta_i}) / \partial \tilde{X}_{ij}\}| < B_{14} = B_3^2 B_4 < \infty$$

for some constant B_{14}.

Using similar arguments as above one can directly show that for $i = 1, \ldots, m$, $\int_{\mathbf{M}_i} N(\mathbf{t}^{n_i-\Delta_i}, n_i - \Delta_i, \tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i)\, d\mathbf{t}^{n_i-\Delta_i} < B_{15} < \infty$ for some constant B_{15}.

Also, for $j = \Delta_i + 1, \ldots, n_i$ and $i = 1, \ldots, m$,

$$
\begin{aligned}
&|\partial \int_{\mathbf{M}_i} N(\mathbf{t}^{n_i-\Delta_i}, n_i - \Delta_i, \tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i)\, d\mathbf{t}^{n_i-\Delta_i} / \partial \tilde{X}_{ij} \\
&\quad = \int_{\mathbf{M}_{i,-j}} N(\mathbf{t}_j^{n_i-\Delta_i}, n_i - \Delta_i, \tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i)\, d\mathbf{t}_{-j}^{n_i-\Delta_i}| < B_{15} < \infty
\end{aligned}
$$

for some constant B_{16}.

Hence $\|\xi_m(\sigma, \Lambda)\|_\infty$ and $\|\boldsymbol{\xi}_m(\boldsymbol{\beta}, \sigma)\|$ are bounded are bounded by $(B_{11}B_{12}+B_1B_{13}+B_{14}B_{15} + B_3B_{16})B_7(B_8 + B_9) < \infty$. □