



Published in final edited form as:

*Stat Biosci.* 2010 December ; 2(2): 154–179. doi:10.1007/s12561-010-9026-x.

## A Gaussian Copula Model for Multivariate Survival Data

**Megan Othus** and

Fred Hutchinson Cancer Research Center, Seattle, WA 98109, Tel.: 206-667-5749

**Yi Li**

Harvard University and Dana Farber Cancer Institute, Boston, MA 02115

Megan Othus: mothus@fhcrc.org

### Abstract

We consider a Gaussian copula model for multivariate survival times. Estimation of the copula association parameter is easily implemented with existing software using a two-stage estimation procedure. Using the Gaussian copula, we are able to test whether the association parameter is equal to zero. When the association term is positive, the model can be extended to incorporate cluster-level frailty terms. Asymptotic properties are derived under the two-stage 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

Copula model; Correlated survival data; Proportional hazards model; Semiparametric normal transformation

## 1 Introduction

In some correlated survival data settings, practitioners have two primary interests: determining the effect of treatment and assessing potential dependence between subjects. For example, in many multi-center clinical trials, data are clustered within treatment center. Although institutions participating in clinical trials follow trial-specific protocols, differences can still exist in outcomes between institutions. The dependence among patients treated at the same institution is an important component of a multi-center clinical trial analysis (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).

Our motivating data arise from a large multi-center clinical trial for children with acute lymphoblastic leukemia. The goal of the clinical trial was to test whether either an increase in the strength or an increase in the duration of the standard chemotherapy regimen was associated with improved survival. We are interested in evaluating whether there exists correlation between survival outcomes within an institution while concurrently assessing the efficacy of the new treatment regimens.

A variety of statistical models are available for correlated survival data. Marginal models treat within-cluster correlations as a nuisance (Wei, Lin, and Weissfeld, 1989; Prentice and Cai, 1992; Cai and Prentice, 1995; Cai, Wei, and Wilcox, 2000, among others). Parameters from marginal models have population-average interpretations. Frailty models are used when within-cluster inferences are desired because the parameters from these models have interpretations conditional on the value of the frailty (Clayton, 1978; Oakes, 1989; Murphy,

1994, 1995; Parner, 1998; Cai, Cheng, and Wei, 2002; Lam, Lee, and Leung, 2002; Glidden and Vittinghoff, 2004; Zeng and Lin, 2007, among others). The positive stable frailty model (Fine, Glidden, and Lee, 2003) allows for marginal proportional hazards interpretation of parameters within a frailty model framework. Copula models (Shih and Louis, 1995; Glidden, 2000; Li, Prentice, and Lin, 2008, among others) embed marginal survival functions within a copula function parametrized by an association term.

For our application, we would like to: (1) test whether either of the two new chemotherapy schedules is associated with improved marginal survival and (2) test whether there is a non-zero correlation between survival outcomes within institutions while controlling for known prognostic factors. Few existing models for multivariate survival data accommodate both semiparametric marginal distributions and unrestricted pairwise dependence.

Consider, for example, Clayton (1978)'s model for a pair of survival times. The dependence term for this model,  $\theta$ , takes values in  $(0, \infty)$ . When  $\theta = 1$  the model reduces to the independence model, while  $\theta > 1$  induces positive association and  $\theta < 1$  induces negative association. When  $\theta \geq 0.5$ , the distribution is absolutely continuous, but for  $\theta \leq 0.50$ , a singular distribution is concentrated on a curve (Oakes, 1989). Hougaard (2000) noted that frailty models cannot yield unrestricted marginal distributions with unrestricted pairwise parameters. Hence, it will be of substantial interest to specify a semiparametric likelihood model that allows for arbitrary modeling of the marginal survival functions and a flexible and interpretable correlation structure.

The goal of this project is to develop a model for multivariate survival data that addresses points (1) and (2) above. To this end we use a semiparametric normal transformation that establishes a Gaussian copula for survival data. The marginal survival function follows a proportional hazards model. The Gaussian copula includes a parameter that summarizes the within-cluster correlation. The correlation parameter can take positive and negative values, which allows for straightforward testing of whether the correlation parameter is equal to zero.

We note that there have been two previous articles using the semiparametric normal transformation model, but neither is applicable to our setting and the proposed model extends the range of data to which the idea can be applied. In contrast to Li et al (2008), our model can accommodate varying cluster sizes and allows for covariates. Li and Lin (2006) assume a specific spatial correlation structure on the entire dataset. In contrast, our method explicitly allows for correlated survival times within independent clusters.

The rest of the paper is structured as follows: in Section 2 we define notation and describe the model; Section 3 summarizes inference procedures; Section 4 outlines an extension of the model when the correlation term is positive; we provide a summary of asymptotic results in Section 5; simulations are presented in Section 6; Section 7 contains an analysis of a Children's Oncology Group multi-center clinical trial; and we finish with a brief discussion in Section 8. Regularity conditions and 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  $j = 1, \dots, n_i$  and  $i = 1, \dots, m$ . The observed data are  $X_{ij} = \min(T_{ij}, C_{ij})$  and  $\Delta_{ij} = I(T_{ij} \leq C_{ij})$ . Let  $Z_{ij}(t)$  denote an external time-dependent covariate vector (Kalbfleisch and Prentice, 2002, page 197) of length  $p$  and write its covariate path up to time  $t$  as  $\bar{Z}_{ij}(t) = \{Z_{ij}(s) \mid 0 \leq s \leq t\}$ . Assume that  $T_{ij}$ , conditional on the covariate process  $\bar{Z}_{ij}(T_{ij})$ , is independent of  $C_{ij}$ . Also, assume that, conditional on each individual's covariate path, the hazard of  $T_{ij}$ , denoted  $\lambda\{t \mid \bar{Z}_{ij}(t)\}$ , follows a proportional hazards model:

$$\lim_{h \rightarrow 0} h^{-1} P\{t \leq T_{ij} < t+h | T_{ij} \geq t, \bar{Z}_{ij}(t)\} = \lambda(t) \exp\{\beta' Z_{ij}(t)\}. \quad (1)$$

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

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

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

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

Denote the correlation of  $(\tilde{T}_{i1}, \dots, \tilde{T}_{in_i})$  with  $\Sigma_i$ . We consider an exchangeable correlation structure for  $\Sigma_i$ , where the diagonal terms are equal to 1 and off-diagonal terms are equal to  $\sigma$ . In this model  $\sigma$  can take positive and negative values. The value zero is an interior point of the parameter space for  $\sigma$  so this model can be used to test whether potentially clustered survival data have non-zero correlation. The Gaussian copula model of Li et al (2008) is a special case of this model with no covariates and cluster size fixed at two.

The term  $\sigma$  can be considered a summary measure for the correlation between two subjects within the same cluster after controlling for the covariates included in model (1). Known prognostic factors can be included in the proportional hazards model, and the estimate of the correlation will be based on Cox-Snell type residuals as defined with Equation (2). The term  $\sigma$  can be viewed as a generalization of Kendall's  $\tau$  and Spearman's  $\rho$  to allow for covariates. For bivariate data, a direct relationship between  $\sigma$  and Kendall's  $\tau$  and Spearman's  $\rho$  is straightforward to establish (Li et al, 2008). We can relate  $\sigma$  to the original time scale using the cross-ratio, a local dependence measure (Kalbfleisch and Prentice, 2002). A derivation of this result can be found in Li and Lin (2006, Section 3.1).

## 3 Inference

### 3.1 Likelihood Development

Let  $Y_{ij}$  be a potentially censored version of  $\tilde{T}_{ij}$  (Equation (2)). The semiparametric normal transformation is monotone and thus preserves censoring patterns. To simplify the

presentation, define  $\Delta_i = \sum_{j=1}^{n_i} \Delta_{ij}$  and order the observations such that  $\Delta_{i1} = \dots = \Delta_{i\Delta_i} = 1$ .

First consider  $1 \leq \Delta_i \leq n_i - 1$ . Let  $Y_i^{\Delta_i} = (Y_{i1}, \dots, Y_{i\Delta_i})$  and  $Y_i^{n_i - \Delta_i} = (Y_{i\Delta_i+1}, \dots, Y_{in_i})$ .

Let  $\Sigma_i$  be the covariance matrix for the transformed failure times. Write  $\Sigma_i$  as a partitioned matrix:

$$\Sigma_i = \begin{pmatrix} \Sigma_{i11} & \Sigma_{i12} \\ \Sigma_{i21} & \Sigma_{i22} \end{pmatrix},$$

where  $\Sigma_{i11}$  has dimension  $\Delta_i \times \Delta_i$ . The vector  $Y_i^{\Delta_i}$  follows a multivariate normal distribution with mean 0 and covariance matrix  $\Sigma_{i11}$ . It follows that  $Y_i^{n_i-\Delta_i}|Y_i^{\Delta_i}$  is a censored observation from a normal distribution with mean  $\sum_{i21} \sum_{i11}^{-1} Y_i^{\Delta_i}$  and covariance matrix  $\sum_{i22} - \sum_{i21} \sum_{i11}^{-1} \sum_{i12}$ .

Because the semiparametric normal transformation is monotonic, the likelihood for the observed data,  $X_{ij}$ , can be written in terms of the transformed terms,  $Y_{ij}$ . To do so, we use the fact that  $P(X_{ij} < x) = P(Y_{ij} < y)$ , where  $y$  is the semiparametric normal transformation of  $x$ . Write the likelihood based on the observed data as  $L(\sigma, \beta, \lambda)$ . The likelihood contribution from cluster  $i$  is

$$\varphi^{\Delta_i}(Y_i^{\Delta_i}) \tilde{\Phi}^{n_i-\Delta_i}(Y_i^{n_i-\Delta_i}|Y_i^{\Delta_i}) \prod_{j=1}^{n_i} [f\{X_{ij}|\bar{Z}_{ij}(X_{ij})\}/\varphi(Y_{ij})]^{\Delta_{ij}}, \tag{3}$$

where  $\varphi^{\Delta_i}$  is the multivariate normal density corresponding to its argument,  $\tilde{\Phi}^{n_i-\Delta_i}$  is the multivariate normal survival function corresponding to its argument,  $f$  is the density corresponding to Equation (1), and  $\varphi$  is the standard normal density function. If  $\sigma = 0$ ,  $L(\sigma, \beta, \lambda)$  reduces to the usual proportional hazards likelihood. A derivation of this likelihood is provided in the Appendix.

When  $\Delta_i = 0$  (all subjects in cluster  $i$  are censored) define  $\varphi^{\Delta_i} = 1$  and

$\tilde{\Phi}^{n_i-\Delta_i}(Y_i^{n_i-\Delta_i}|Y_i^{\Delta_i}) = \tilde{\Phi}^{n_i-\Delta_i}(Y_i^{n_i-\Delta_i})$ . When  $\Delta_i = n_i$  (all subjects in cluster  $i$  have been observed to fail) define  $\tilde{\Phi}^{n_i-\Delta_i} = 1$ . With these conventions, Equation (3) holds also for  $\Delta_i = 1$  and  $\Delta_i = n_i$ .  $L(\sigma, \beta, \lambda)$  is the product of Equation (3) over  $i = 1, \dots, m$ .

To make the likelihood  $L(\sigma, \beta, \lambda)$  more transparent, 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  $Y_{A1} = \Phi^{-1}[1 - S\{C_{A1} | \bar{Z}_{A1}(C_{A1})\}]$  and  $Y_{A2} = \Phi^{-1}[1 - S\{T_{A2} | \bar{Z}_{A2}(T_{A2})\}]$ . In this case

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

The first term of  $L(\sigma, \beta, \lambda)$  can be written

$$\varphi_{Y_{A2}}^1(y_{A2}) = (2\pi)^{-1/2} \exp(-y_{A2}^2/2) \tag{4}$$

while the second term can be written

$$\Phi_{Y_{A1}|Y_{A2}}^1(y_{A1}|y_{A2}) = \int_{y_{A1}}^{\infty} \{2\pi(1 - \sigma^2)\}^{-1/2} \exp\{-(x - \sigma y_{A2})^2/2(1 - \sigma^2)\} dx. \tag{5}$$

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

### 3.2 Estimation

We propose a two-stage method to estimate  $(\sigma, \beta, A)$ . First we estimate  $\hat{\beta}$  and  $\hat{A}$  from the marginal proportional hazards model. We then solve  $\max_{\sigma} L(\sigma, \hat{\beta}, \hat{A})$  for  $\hat{\sigma}$ . Formulas for the standard errors of  $\hat{\beta}$  and  $\hat{A}$  that account for clustering can be found using a sandwich formula (Spiekerman and Lin, 1998). The formula for  $\beta$  has been implemented in many statistical programs. The analytic standard error of  $\hat{\sigma}$  is complicated because it needs to account for the variability from  $\hat{\beta}$  and  $\hat{A}$ . In practice the standard error can be estimated using a resampling scheme. To maintain the correlated structure of the failure times, the clusters should be the unit of removal for the resampling calculations (Cai et al, 1997; Cai and Shen, 2000). We choose to use the jackknife for resampling because theoretical validation of the method exists (Lipsitz, Dear, and Zhao, 1994; Lipsitz and Parzen, 1996).

This estimation procedure is computationally straightforward. Marginal estimates of the survival function are available in all standard computing programs, while the likelihood for  $\sigma$ ,  $L(\sigma, \hat{\beta}, \hat{A})$ , is proportional to a product of multivariate normal terms, quickly computable using existing software (e.g., R package MVTNORM).

## 4 A Frailty Model Extension

### 4.1 A marginalized frailty model

When  $\sigma > 0$ , the model for  $\tilde{T}_{ij}$  can be extended to allow for a frailty term:

$$\tilde{T}_{ij} = \sqrt{\sigma} b_i + \varepsilon_{ij}, \quad (6)$$

where  $b_i$  is a cluster-level frailty and  $\varepsilon_{ij}$  is an error term. We assume that cluster-level frailties  $b_i$  have a standard normal distribution and the error terms  $\varepsilon_{ij}$  are independent and identically distributed  $N(0, 1 - \sigma)$  random variables that are independent of  $b_i$ . The cluster-level frailties  $b_i$  can be used to assess cluster-level differences. The  $\beta$  parameters in Equation (1) have marginal interpretation, while  $\sigma$  and  $b_i$  from Equation (6) characterize the cluster effect.

Equation (6) merges elements of frailty models with the marginal model, and so will take a moment to review the interpretation of the components of this model. Larger values of  $\sigma$  imply that the frailty terms explain a larger portion of the variance in the  $T_{ij}$  compared to smaller values for  $\sigma$ . Larger values of  $\sigma$  provide evidence for a stronger cluster effect compared to smaller values for  $\sigma$ . In the context of a multi-center clinical trial, the cluster-level frailties ( $b_i$ ) characterize the center effect. For a fixed set of covariates, smaller or more negative values for  $b_i$  are associated with shorter survival times (on the original untransformed scale) compared to larger or more positive values for  $b_i$ .

### 4.2 Prediction of the frailty terms

If  $\hat{\sigma} > 0$ , prediction of  $b_i$  and the associated standard error can be found using Laplace approximations to the  $b_i$ 's first two moments. Denote the observed data for the  $i^{\text{th}}$  cluster,  $(X_{i1}, \dots, X_{in_i}, \Delta_{i1}, \dots, \Delta_{in_i}, Z_{i1}(X_{i1}), \dots, Z_{in_i}(X_{in_i}))$ , with  $\Psi_i$ . The conditional density of  $b_i$  given the observed data  $\Psi_i$ , denoted  $g(b_i | \Psi_i; \sigma, \beta, A)$ , can be written

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

where  $L_i$  is the likelihood for  $\Psi_i | \sigma, \beta, \Lambda$ . Define  $k_i$  such that

$g(b_i|\Psi_i; \sigma, \beta, \Lambda) = L_i^{-1} \exp\{k_i(b_i|\Psi_i; \sigma, \beta, \Lambda)\}$ . Using the Laplace approximations to the first two moments of  $g(b_i | \Psi_i; \sigma, \beta, \Lambda)$  (Booth and Hobert, 1998), the predicted estimate and variance of  $b_i$  are taken to be:

$$\widehat{b}_i = E(b_i|\Psi_i) \approx \arg \max_{b_i} k_i(b_i|\Psi_i; \widehat{\sigma}, \widehat{\beta}, \widehat{\Lambda}) \quad (7)$$

$$V(b_i|\Psi_i) \approx -\ddot{k}_i(\widehat{b}_i|\Psi_i; \widehat{\sigma}, \widehat{\beta}, \widehat{\Lambda})^{-1}, \quad (8)$$

where double superscript dots denote second derivatives.

The prediction of the shared frailties is straightforward. The expression for  $k_i(b_i | \widehat{\sigma}, \widehat{\beta}, \widehat{\Lambda}, \Psi_i)$  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.

## 5 Theoretical Results

The following theorems establish the theoretical properties of  $(\widehat{\sigma}, \widehat{\beta}, \widehat{\Lambda})$  where their true values are denoted with  $(\sigma_0, \beta_0, \Lambda_0)$ .

### Theorem 1

Under Conditions C.1 – C.6 in the Appendix,  $(\widehat{\sigma}, \widehat{\beta}, \widehat{\Lambda})$  converges in probability to  $(\sigma_0, \beta_0, \Lambda_0)$  as  $m \rightarrow \infty$ .

### Theorem 2

Under Conditions C.1 – C.7 in the Appendix Report, as  $m \rightarrow \infty$ ,  $\sqrt{m}(\widehat{\sigma} - \sigma_0)$  and  $\sqrt{m}(\widehat{\beta} - \beta_0)$  converge to zero-mean normal distributions and  $\sqrt{m}\{\widehat{\Lambda}(t) - \Lambda_0(t)\}$  converges to a zero-mean Gaussian process.

Proofs of Theorems 1 and 2 can be found in the Appendix. The proofs of both theorems for  $\widehat{\sigma}$  adjust for the two-stage estimating procedure. These theorems verify that  $\widehat{\sigma}$  is consistent and asymptotically normal when plug-in estimates of  $\beta$  and  $\Lambda$  are used in the likelihood function.

## 6 Simulation Results

Simulations were conducted 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  $\beta_1$ ) and one Uniform(0,1) covariate with parameter equal to 0.75 (denoted  $\beta_2$ ). Censoring times were taken from the Exponential(median=3) distribution and produced about a 25% censoring rate. Correlated

survival times were created by first generating random correlated multivariate normal values. The normal values were transformed to the survival scale using Equation (2). Each simulation is based on 250 replications.

In order to focus on the novel elements of the method, we present results for  $\beta_1$  and  $\beta_2$  in the Appendix. As has been shown by other authors, our simulations verify that estimates of  $\beta_1$  and  $\beta_2$  have little bias and appropriate coverage probability.

For our results for  $\sigma$ , we summarize scenarios with 45, 60, and 90 clusters. Within each replication, clusters varied in size between 2 and 7 units. Standard errors (SEs) for  $\sigma$  were found using the jackknife. We chose to use the jackknife for resampling because theoretical validation of the method exists for multivariate survival data (Lipsitz, Dear, and Zhao, 1994; Lipsitz and Parzen, 1996). Bias, SEs, and coverage probabilities are summarized in Table 1. Power results are summarized in Figure 1.

The estimates of  $\sigma$  have little bias across the simulations. The jackknife SE is close to the Monte Carlo SE across the scenarios. In clusters of size 45, parameter estimates remain unbiased, but the coverage probability drops below nominal levels. As expected, as sample size and correlation increases, the power for testing whether  $\sigma \neq 0$  increases.

We also conducted simulations to assess the performance of our frailty estimation method (Equations (7) and (8)). In these simulations, we took  $\sigma = 0.5$  to ensure all estimates of  $\sigma$  were positive so that frailties could be predicted. Data were generated using Equation (6), to provide true values for the frailty terms. The parameters  $\sigma$  and  $\beta$  were estimated using our proposed two-stage method and their estimates were used in Equations (7) and (8) to predict frailty values and calculate standard errors. We summarize results for simulations with 90 clusters and a range of cluster sizes: cluster sizes varying between 3 and 10 with median size 5, cluster sizes varying between 7 and 20 with median size 10, cluster sizes varying between 10 and 33 with median size 15, and cluster sizes varying between 13 and 50 with median size 20. Results are summarized in Table 2.

The method performs well even with small cluster sizes. The relative bias is small. As the cluster sizes increase, the likelihood SEs approach 1, their true value.

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

### 7.1 Is there evidence of correlation between the survival times of patients within the same institution?

We applied our method to a Children's Oncology Group (COG) study (protocol number 1961) (Seibel et al, 2008). We analyzed 460 children with enlarged livers from 104 institutions. The goal of the clinical trial was to test whether either an increase in the strength or an increase in the duration of the standard chemotherapy regimen was associated with improved survival for "higher risk" acute lymphoblastic leukemia patients. A  $2 \times 2$  factorial design was used. The distribution of subjects with enlarged livers among the four arms is presented in Table 3. The number of subjects in each of the 104 institutions is summarized in the histogram in Figure 2. We analyzed the overall survival endpoint.

We present regression results in Table 4. Standard errors for the marginal survival covariates were found using three methods: a sandwich formula, the jackknife, and a naïve estimate ignoring potential clustering. Standard errors for  $\sigma$  were found using the jackknife. P-values were found using the jackknife standard errors.

In a model with two covariates for treatment, there was some evidence that increased duration of chemotherapy was associated with worse overall survival compared to the

standard duration. There was also evidence that overall survival was improved for patients with increased strength of chemotherapy compared to standard strength. The estimate for the correlation of the transformed failure times was 0.173 with a standard error of 0.159.

We also considered a larger model including the two treatment covariates, their interaction, and the prognostic factors of age and platelet count at registration. In this larger model, none of the covariates appeared to be associated with overall survival. The correlation term of this larger model was of similar magnitude as the treatment only model: the estimate was 0.207 with a standard error of 0.200.

## 7.2 Is there further information that can be gained by considering frailty terms?

The standard errors for the correlation term ( $\hat{\sigma}$ ) were large in both models summarized in Table 4, so we felt it would be useful to investigate whether there were any outliers contributing to the large standard error. In both regression models presented, the estimate of the correlation ( $\hat{\sigma}$ ) was positive, so we were able to predict frailty terms for each institution. The predicted frailty values were very similar between the two models, so we only present the frailties from the larger model.

A qq-plot for frailty values standardized by their standard errors indicated several potential outliers (Figure 3). The largest positive standardized frailties were from institutions with only one patient, so we disregarded those institutions as potential outliers. The largest negative frailties were from institutions with 10 and 22 patients, so we investigated these institutions further.

A plot of the predicted frailty values by institution size is provided in Figure 4. The two institutions identified as potential outliers in the qq-plot are marked in Figure 4, and the institutions still appear to be outliers. Among the other institutions there appears to be a positive trend, where institutions that contributed more patients had larger frailty values. Both of the institutions that are potential outliers do not follow this trend.

Given these results, we were interested in whether these trends could be explained by the type of patients at the two institutions. We used Fisher's Exact test to compare known prognostic factors between each of the potential outlying institutions and all other patients. Given the limited sample sizes, we recognized that the power of the tests may be limited. We compared the following baseline characteristics: spleen enlargement, race, nodes (normal, moderately enlarged, significantly enlarged), mediastinal mass, white blood cell count (less than 50, 50–199, 200 or more; all  $\times 10^3/\text{mm}^2$ ), hemoglobin count (1 to 7.9, 8.0 to 10.9, 11.0 or more; all g/dL), platelet count (1 to 49, 50 to 149, 150 or more; all  $\times 10^3/\text{mm}^3$ ), and age (1–9, 10–15, 16 and older). Only one test had a p-value  $< 0.05$ , the comparison of age with the institution with ten patients.

The larger regression model in Table 4 controls for age, and age was not significantly associated with survival in the model. Given this result, there may be other factors beyond the available patient data that explain why these institutions have unusual frailty values.

To evaluate the impact of these two institutions on the estimate of the correlation, we refit our models excluding patients from the two institutions marked in Figures 3 and 4. A summary of the original and refit correlation estimates ( $\hat{\sigma}$ ) is provided in Table 5. The correlation estimate was smaller in the subset compared to the full dataset, though the change was more pronounced in the treatment only model.

The correlation estimates for the subset data are positive and so frailty terms were able to be predicted for each of the institutions. A qq-plot of the frailties standardized by their standard



errors and a plot of the frailty terms versus the institution size is provided in Figure 5. There do not appear to be any outliers in the model fit with the subset of the data. The frailty results look very similar to the results in Figures 3 and 4 excluding the two marked values.

In this subset of the data, there is not evidence that survival times are correlated within institution. Our available data does not explain why two institutions may be outliers.

### 7.3 An alternative analysis using a gamma frailty model

We applied the gamma frailty model to the COG dataset. Results from an analysis with the two treatment covariates and an analysis including a treatment interaction and controlling for age and platelet count is provided in Table 6. The estimate for covariates is the log hazard ratio.

In the gamma frailty model, the strength of the correlation is measured by the variance of the gamma frailties. In both the treatment only and larger models, the estimate for the frailty variance is not significant. As with our results, these two gamma frailty models do not show evidence of clustering within institution.

The covariate parameter values from the proposed model are the same as would be found using a marginal proportional hazards model treating the potential clustering as a nuisance. Therefore the parameters from the proposed model have population interpretation as an average effect across all the patients in the study.

In contrast, due to the functional specification of the gamma frailty hazard function, covariate parameters values from a gamma frailty model are interpreted as conditional on the value of the frailty and do not represent an average effect. When there is no correlation within clusters, the proposed model and the gamma frailty model both reduce to the proportional hazards model. Given the low evidence of clustering in this dataset, it is not surprising that the covariate parameter estimates are close in Tables 4 and 6. Given that the COG study was interested in population averaged hazard ratios, the parameters in Table 4 provide a more appropriate interpretation.

As with the proposed method, we looked plotted the log frailty values from the gamma frailty model versus the institution size in Figure 6. We note that the with the proposed method large, positive frailties are associated with improved survival, in contrast to the gamma frailty model in which large frailties are associated with increased hazards. The two institutions identified as potential outliers with the proposed method are marked with crosses. In Figure 6 there does not appear to be a relationship between log frailty value and institution size. The two marked institutions do not appear to have unusual frailty values given the trends in the data.

## 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 the correlation between survival times is described by a Gaussian copula. When the correlation between the transformed survival times is positive, the model can be extended to incorporate frailties. We believe this model can provide an easy graphical method to identify potentially “abnormal” clusters.

When the model was applied to a Children's Oncology Group dataset, we were able to identify two potential outlying institutions. Baseline patient characteristics were unable to explain the observed trends. More information on patients or on the institutions may be needed to understand the underlying issues.

One area of future research is investigating the relative efficiency of using a non-parametric maximum likelihood approach in contrast to the two-stage approach used in this paper. Also, it might be useful to develop a more flexible frailty model that allows for covariates so that factors influencing correlation can be examined within a regression framework.

## Acknowledgments

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.

## References

- Anello C, O'Neill R, Dubey S. Multicentre Trials: A US Regulatory Perspective. *Statistical Methods in Medical Research*. 2005; 14(3):303–318. [PubMed: 15969305]
- Booth J, Hobert J. Standard Errors of Prediction in Generalized Linear Mixed Models. *Journal of the American Statistical Association*. 1998; 93(441):262–272.
- Cai J, Prentice R. Estimating Equations for Hazard Ratio Parameters Based on Correlated Failure Time Data. *Biometrika*. 1995; 82(1):151–164.
- Cai J, Shen Y. Permutation Tests for Comparing Marginal Survival Functions with Clustered Failure Time Data. *Statistics in Medicine*. 2000; 19(21):2963–2973. [PubMed: 11042626]
- Cai J, Zhou H, Davis C. Estimating the Mean Hazard Ratio Parameters for Clustered Survival Data with Random Clusters. *Statistics in Medicine*. 1997; 16(17):2009–2020. [PubMed: 9304770]
- Cai T, Wei L, Wilcox M. Semiparametric Regression Analysis for Clustered Failure Time Data. *Biometrika*. 2000; 87(4):867–878.
- Cai T, Cheng S, Wei L. Semiparametric Mixed-effects Models for Clustered Failure Time Data. *Journal of the American Statistical Association*. 2002; 97(458):514–522.
- Chen X, Fan Y, Tsyrennikov V. Efficient Estimation of Semiparametric Multivariate Copula Models. *Journal of the American Statistical Association*. 2006; 101(475):1228–1240.
- Clayton D. A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*. 1978:141–151.
- Fine J, Glidden D, Lee K. A Simple Estimator for a Shared Frailty Regression Model. *Journal of the Royal Statistical Society Series B, Statistical Methodology*. 2003; 65(1):317–329.
- Fleiss J. Analysis of Data from Multiclinic Trials. *Controlled Clinical Trials*. 1986; 7(4):267–275. [PubMed: 3802849]
- Glidden D. A Two-Stage Estimator of the Dependence Parameter for the Clayton-Oakes Model. *Lifetime Data Analysis*. 2000; 6(2):141–156. [PubMed: 10851839]
- Glidden D, Vittinghoff E. Modelling Clustered Survival Data from Multicentre Clinical Trials. *Statistics in medicine*. 2004; 23(3):369–388. [PubMed: 14748034]
- Gray R. A Bayesian Analysis of Institutional Effects in a Multicenter Cancer Clinical Trial. *Biometrics*. 1994; 50(1):244–253. [PubMed: 8086608]
- Hougaard, P. *Analysis of Multivariate Survival Data*. Springer Verlag; 2000.
- Jones B, Teather D, Wang J, Lewis J. A Comparison of Various Estimators of a Treatment Difference for a Multi-centre Clinical Trial. *Statistics in Medicine*. 1998; 17:1767–1777. [PubMed: 9749446]
- Kalbfleisch, J.; Prentice, R. *The Statistical Analysis of Failure Time Data*. 2. Wiley; 2002.
- Klaassen C, Wellner J. Efficient Estimation in the Bivariate Normal Copula Model: Normal Margins are Least Favourable. *Bernoulli*. 1997; 3(1):55–77.

- Lam K, Lee Y, Leung T. Modeling Multivariate Survival Data by a Semiparametric Random Effects Proportional Odds Model. *Biometrics*. 2002; 58(2):316–323. [PubMed: 12071404]
- Li Y, Lin X. Semiparametric Normal Transformation Models for Spatially Correlated Survival Data. *Journal of the American Statistical Association*. 2006; 101(474):591–603.
- Li Y, Prentice R, Lin X. Semiparametric Maximum Likelihood Estimation in Normal Transformation Models for Bivariate Survival Data. *Biometrika*. 2008; 95(4):947–960. [PubMed: 19079778]
- Lipsitz S, Parzen M. A Jackknife Estimator of Variance for Cox Regression for Correlated Survival Data. *Biometrics*. 1996; 52:291–298. [PubMed: 8934598]
- Lipsitz S, Dear K, Zhao L. Jackknife Estimators of Variance for Parameter Estimates from Estimating Equations with Applications to Clustered Survival Data. *Biometrics*. 1994; 50:842–846. [PubMed: 7981404]
- Logan B, Nelson G, Klein J. Analyzing Center Specific Outcomes in Hematopoietic Cell Transplantation. *Lifetime Data Analysis*. 2008; 14:389404.
- Murphy S. Consistency in a Proportional Hazards Model Incorporating a Random Effect. *Annals of Statistics*. 1994; 22(2):712–731.
- Murphy S. Asymptotic Theory for the Frailty Model. *Annals of Statistics*. 1995; 23:182–198.
- Oakes D. Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association*. 1989; 84(406):487–493.
- Parner E. Asymptotic Theory for the Correlated Gamma-frailty Model. *Annals of Statistics*. 1998; 26(2):183–214.
- Prentice R, Cai J. Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data. *Biometrika*. 1992; 79(3):495–512.
- Seibel N, Steinherz P, Sather H, Nachman J, Delaat C, Ettinger L, Freyer D, Mattano L Jr, Hastings C, Rubin C, Bertolone K, Franklin J, Heerema N, Mitchell T, Pysemany A, La M, Edens C, Gaynon P. 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*. 2008; 11(5):2548–2555. [PubMed: 18039957]
- Senn S. Some Controversies in Planning and Analysing Multi-centre Trials. *Statistics in Medicine*. 1998; 17(15–16):1753–1765. [PubMed: 9749445]
- Shih J, Louis T. Assessing Gamma Frailty Models for Clustered Failure Time Data. *Lifetime Data Analysis*. 1995; 1(2):205–220. [PubMed: 9385102]
- Spiekerman C, Lin D. Marginal Regression Models for Multivariate Failure Time Data. *Journal of the American Statistical Association*. 1998; 93(443):1164–1175.
- Therneau, T.; Grambsch, P. *Modeling Survival Data: Extending the Cox Model*. Springer Verlag; 2000.
- van der Vaart, A. *Asymptotic Statistics*. Cambridge University Press; 1998.
- Vierron E, Giraudeau B. Sample Size Calculation for Multicenter Randomized Trial: Taking the Center Effect into Account. *Contemporary Clinical Trials*. 2007; 28(4):451–458. [PubMed: 17188941]
- Wei L, Lin D, Weissfeld L. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association*. 1989; 84(408):1065–1073.
- Zeng D, Lin D. Maximum Likelihood Estimation in Semiparametric Regression Models with Censored Data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69(4):507–564.
- Zheng L, Zelen M. Multi-center Clinical Trials: Randomization and Ancillary Statistics. *Annals of Applied Statistics*. 2008; 2(2):582–600.

## A Regularity Conditions and Notation

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

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

- C.2  $\Lambda(\tau) < \infty$
- C.3  $\sigma \in \nu$ , where  $\nu$  is a compact subset of  $(-1, 1)$
- C.4  $P(C_{ij} \geq t \forall t \in [0, \tau] | Z_{ij}) > \delta_c > 0$  for  $j = 1, \dots, n_i$  and  $i = 1, \dots, m$
- C.5 Write  $Z_{ij}(t) = \{Z_{ij1}(t), \dots, Z_{ijp}(t)\}$ .  $|Z_{ijk}(0)| + \int_0^\tau |dZ_{ijk}(t)| \leq B_Z < \infty$  almost surely for some constant  $B_Z$  and  $i = 1, \dots, m, j = 1, \dots, n_i, k = 1, \dots, p$
- C.6  $E[\log\{L(\sigma_1; \beta, \Lambda)/L(\sigma_2; \beta, \Lambda)\}]$  exists for all  $\sigma_1, \sigma_2 \in (-1, 1)$
- C.7 Let  $Y_{ij}(t) = I(X_{ij} \geq t)$ ,  $K = \max_i n_i$ ,  $a^{\otimes 0} = 1$ ,  $a^{\otimes 1} = a$ ,  $a^{\otimes 2} = a'a$ ,

$$Q_j^{(k)}(\beta, t) = m^{-1} \sum_{i=1}^m Y_{ij}(t) \exp\{\beta' Z_{ij}(t)\} Z_{ij}(t)^{\otimes k}, \quad q_j^{(k)}(\beta, t) = E\{Q_j^{(k)}(\beta, t)\},$$

$$\eta_j(\beta, t) = \frac{q_j^{(1)}(\beta, t)}{q_j^{(0)}(\beta, t)}, \quad \varrho_j(\beta, t) = \frac{q_j^{(2)}(\beta, t)}{q_j^{(0)}(\beta, t)} - \eta_j(\beta, t)^{\otimes 2} \text{ for } j=1, \dots, K.$$

Assume  $\sum_{j=1}^K \int_0^\tau \varrho_j(\beta_0, t) q_j^{(0)}(\beta_0, t) \lambda_0(t) dt$  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  $\sigma_0$ . Condition C.7 is a technical condition from Spiekerman and Lin (1998) that is needed for the results for  $\hat{\beta}$  and  $\hat{\Lambda}$ .

Before providing technical results, we will provide a brief description of how the likelihood in Equation (3) can be derived. Let  $S_{ij}(x) = S\{x | \bar{Z}_{ij}(X_{ij})\}$ . We consider a cluster of size  $n_i$  with  $\Delta_i$  subjects who are not censored. The contribution to the likelihood will be the density function for  $X_{i1}, \dots, X_{i\Delta_i}$  and the survival function for  $X_{i\Delta_i+1}, \dots, X_{in_i}$ . We have specified a copula structure, so we can write the survival function for  $(X_{i1}, \dots, X_{in_i})$  as follows:

$$\begin{aligned} P(X_{i1} > x_{i1}, \dots, X_{in_i} > x_{in_i}) &= P\{1 - S_{i1}(X_{i1}) > 1 - S_{i1}(x_{i1}), \dots, 1 - S_{in_i}(X_{in_i}) > 1 - S_{in_i}(x_{in_i})\} \\ &= P[\Phi^{-1}\{1 - S_{i1}(X_{i1})\} > \Phi^{-1}\{1 - S_{i1}(x_{i1})\}, \dots, \Phi^{-1}\{1 - S_{in_i}(X_{in_i})\} > \Phi^{-1}\{1 - S_{in_i}(x_{in_i})\}] \\ &= P(Y_{i1} > y_{i1}, \dots, Y_{in_i} > y_{in_i}) \\ &= P(Y^{\Delta_i} > y^{\Delta_i}, Y^{n_i - \Delta_i} > y^{n_i - \Delta_i}) \end{aligned}$$

We can write the density of  $(X_{i1}, \dots, X_{i\Delta_i})$ , denoted  $f_{\Delta_i}(x_{i1}, \dots, x_{i\Delta_i}) = f(x_{i1}, \dots, x_{i\Delta_i} | \bar{Z}_{i1}(x_{i1}), \dots, \bar{Z}_{in_i}(x_{in_i}))$ , in terms of  $Y^{\Delta_i}$  using the chain-rule:

$$\begin{aligned} f_{\Delta_i}(x_{i1}, \dots, x_{i\Delta_i}) &= \partial^2 P(X_{i1} > x_{i1}, \dots, X_{i\Delta_i} > x_{i\Delta_i}) / \partial x_{i1}, \dots, \partial x_{i\Delta_i} \\ &= \partial^2 P(Y^{\Delta_i} > y^{\Delta_i}) / \partial x_{i1}, \dots, \partial x_{i\Delta_i} \\ &= \frac{\partial^2 P(Y^{\Delta_i} > y^{\Delta_i})}{\partial y_{i1}, \dots, \partial y_{i\Delta_i}} \frac{\partial y_{i1}}{\partial x_{i1}}, \dots, \frac{\partial y_{i\Delta_i}}{\partial x_{i\Delta_i}} \\ &= \varphi^{\Delta_i}(y^{\Delta_i}) \frac{\partial y_{i1}}{\partial x_{i1}}, \dots, \frac{\partial y_{i\Delta_i}}{\partial x_{i\Delta_i}}. \end{aligned}$$

For  $j = 1, \dots, \Delta_i$ :

$$\begin{aligned} \partial y_j / \partial x_j &= \partial \Phi^{-1}\{1 - S_j(x_j)\} / \partial x_j \\ &= \frac{1}{\varphi[\Phi^{-1}\{1 - S_j(x_j)\}]} \frac{\partial \{1 - S_j(x_j)\}}{\partial x_j} \\ &= \frac{1}{\varphi(y_j)} \frac{\partial \{1 - S_j(x_j)\}}{\partial x_j} \\ &= \frac{1}{\varphi(y_j)} f_j(x_j). \end{aligned}$$

The rest of Equation (3) can be found by combining these calculations with survival terms for subjects who are censored.

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

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

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

$$\begin{aligned} l_{m0}(\sigma) &= m^{-1} \log L^*(\sigma, \beta_0, \Lambda_0), \quad l_m(\sigma) = m^{-1} \log L^*(\sigma, \beta, \Lambda), \\ \widehat{l}_m(\sigma) &= m^{-1} \log L^*(\sigma, \widehat{\beta}, \widehat{\Lambda}), \quad U_{m0}(\sigma) = \partial l_{m0}(\sigma) / \partial \sigma, \quad \text{and } U_m(\sigma) = \partial l_m(\sigma) / \partial \sigma. \end{aligned}$$

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.

## 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 Appendix D.

To account for the fact that plug-in estimates of  $\beta$  and  $\Lambda$  are used in the likelihood for  $\sigma$ , we will need to take a Taylor series expansion of the likelihood of  $\sigma$  around  $\beta_0$  and  $\Lambda_0$ . Since  $\Lambda_0$  is an unspecified function, this expansion will need to include a functional expansion term. An expansion using Hadamard derivatives is appropriate for this situation. In order to use the functional expansion, we first need to verify that the log-likelihood is Hadamard differentiable with respect to  $\Lambda$ , 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  $\Lambda$ .

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  $\Lambda$  at  $\mathcal{T} \in BV[0, \tau]$  as

$\int_0^\tau \zeta_m(\Lambda, \sigma)(u) dY(u)$  and let  $\zeta_m(\beta, \sigma) = \partial l_m(\sigma) / \partial \beta$ . Under conditions C.1–C.5,  $\|\zeta_m(\Lambda, \sigma)\|_\infty$  and  $\|\zeta_m(\beta, \sigma)\|$  are bounded. Expressions for  $\zeta_m(\beta, \sigma)$  and  $\zeta_m(\Lambda, \sigma)$  are provided in the proof.

In order to prove  $\hat{\sigma}$  is consistent we will need to verify the uniform convergence of the log-likelihood with plug-in estimates of  $\beta$  and  $\Lambda$  to the expected value of the log-likelihood evaluated at the true values of  $\beta$  and  $\Lambda$ , denoted  $l_{m0}(\sigma)$ . We accomplish this, using the results of Lemmas 1 and 2, in Lemma 3.

## Lemma 3

Under conditions C.1–C.5, as  $m \rightarrow \infty$ ,

$$\sup_{\sigma \in \mathcal{V}} |\hat{l}_m(\sigma) - E\{l_{m0}(\sigma)\}| = o_p(1).$$

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, for any  $\sigma \neq \sigma_0$ ,

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

## Proof of Theorem 1

The results for  $\hat{\beta}$  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

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

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

$$E\{l_{m0}(\sigma_0)\} - E\{l_{m0}(\hat{\sigma})\} \leq \hat{l}_m(\hat{\sigma}) - E\{l_{m0}(\hat{\sigma})\} + o_p(1) \leq \sup_{\sigma \in \mathcal{V}} |\hat{l}_m(\sigma) - E\{l_{m0}(\sigma)\}| + o_p(1) = o_p(1),$$

(9)

where the last equality comes from Lemma 3.

Take  $\sigma$  such that  $|\sigma - \sigma_0| \geq \varepsilon$  for any fixed  $\varepsilon > 0$ . By Lemma 4 there must exist some  $\gamma_\varepsilon > 0$  such that  $E\{l_{m0}(\sigma)\} + \gamma_\varepsilon < E\{l_{m0}(\sigma_0)\}$ . It follows that  $P(|\hat{\sigma} - \sigma_0| \geq \varepsilon) \leq P[E\{l_{m0}(\sigma)\} + \gamma_\varepsilon < E\{l_{m0}(\sigma_0)\}]$ . Equation (9) implies that  $P[E\{l_{m0}(\hat{\sigma})\} + \gamma_\varepsilon < E\{l_{m0}(\sigma_0)\}]$  converges to 0 as  $m \rightarrow \infty$ . Therefore  $P(|\hat{\sigma} - \sigma_0| \geq \varepsilon)$  converges to 0 as  $m \rightarrow \infty$ .

## 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 Appendix D.

To account for the fact that plug-in estimates of  $\beta$  and  $A$  are used in the likelihood and score function for  $\sigma$ , we will need to take a Taylor series expansion of the score function for  $\sigma$  around  $\beta_0$  and  $A_0$ . To do so we first need to verify that the score function is Hadamard differentiable with respect to  $A$ , 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  $A$ .

After we have an expansion of the score function for  $\sigma$ , 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  $A$  at  $T \in BV[0, \tau]$  as

$\int_0^\tau \xi_m(\sigma, \Lambda)(u) d\Upsilon(u)$  and let  $\xi_m(\beta, \sigma) = \partial U_m(\sigma) / \partial \beta$ . Under conditions C.1–C.5,  $\|\xi_m(\sigma, A)\|_\infty$  and  $\|\xi_m(\sigma, \beta)\|$  are bounded. Expressions for  $\xi_m(\sigma, \beta)$  and  $\xi_m(\sigma, A)$  are provided in the proof.

### Proof of Theorem 2

The result that  $\sqrt{m}(\hat{\beta} - \beta)$  converges to mean zero normal distribution and that  $\sqrt{m}(\hat{\Lambda} - \Lambda_0)$  converges to mean zero Gaussian 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 two-stage estimation procedure. The variance of  $\hat{\sigma}$  should be adjusted compared to a model where  $\beta_0$  and  $A_0$  are used to take into account the estimation of  $\beta$  and  $A$ .

First we will show that the score equation associated with  $\hat{l}_m$  evaluated at  $\sigma_0$  follows a normal distribution. This result coupled with a first order expansion of the score equation associated with  $\hat{l}_m$  around  $\sigma_0$  will finish the proof.

Using Lemma 5, a Taylor series expansion of  $\hat{U}_m(\sigma)$  around  $\beta_0$  and  $A_0$  gives

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

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

$$\sqrt{m}\widehat{U}_m(\sigma_0) = \sqrt{m}U_m(\sigma_0) + \sqrt{m}\int_0^\tau \xi(\sigma_0, \Lambda)(t) d\{\widehat{\Lambda}(t) - \Lambda_0(t)\} + \sqrt{m}\xi(\sigma_0, \beta)(\widehat{\beta} - \beta) + o_p(1). \quad (10)$$

Using the results of Spiekerman and Lin (1998), we can write Equation (10) 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  $\sqrt{m}\widehat{U}_m(\sigma_0)$  converges to a normally distributed random variable with mean zero and variance equal to the variance of  $\Xi_1$ .

Next, we take a first order Taylor series expansion of  $\widehat{U}_m(\widehat{\sigma})$  around  $\sigma_0$ :

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

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

## D Proofs of Lemmas

### Proof of Lemma 1

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

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

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

$$\int_0^\tau Y_{ij}(u) \exp\{\beta' Z_{ij}(u)\} d\Lambda(u)$$

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  $\zeta_m(\beta, \sigma)$ , starting with  $\zeta_m(\sigma, \Lambda)$ . To make the argument more concrete express  $l_m(\sigma)$  as a function of  $\Lambda$  by writing  $l_m(\sigma, \Lambda) = l_m(\sigma)$ . Let  $\Gamma \in BV[0, \tau]$ . Denote

$$H_{ij} = \exp[-\int_0^\tau Y_{ij}(u) \exp\{\beta' Z_{ij}(u)\} d\Lambda(u)].$$



By conditions C.1 and C.2, for  $j = 1, \dots, n_i$  and  $i = 1, \dots, m$ ,  $H_{ij} > 0$  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  $m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} D_{ij}^l Y_{ij}(u) \exp\{\beta' Z_{ij}(u)\} H_{ij}$  and  $D_{ij}^l$  is equal to

$$(\Delta_{ij} [\varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i})^{-1} \{\partial \varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i}) / \partial \tilde{X}_{ij}\}] + (1 - \Delta_{ij}) [\tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{X}_i^{n_i - \Delta_i} | \tilde{X}_i^{\Delta_i})^{-1} \times \{\partial \tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{X}_i^{n_i - \Delta_i} | \tilde{X}_i^{\Delta_i}) / \partial \tilde{X}_{ij}\}]) \sum_{j=1}^{n_i} \partial \Phi^{-1}(H_{ij}) / \partial H_{ij}.$$

Therefore the Hadamard derivative for  $T \in BV[0, \tau]$  is  $\int_0^\tau \zeta_m(\sigma, \Lambda)(u) dY(u)$ . Direct calculation verifies that  $\zeta_m(\sigma, \beta)$  is equal to

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

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

$$\varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i}) = (2\pi)^{-\Delta_i/2} \det(\sum_{i|1}^{-1})^{-1/2} \exp(-\tilde{X}_i^{\Delta_i} \sum_{i|1}^{-1} \tilde{X}_i^{\Delta_i} / 2) > 1/B_1 > 0$$

for some constant  $B_1$  since for  $\tilde{X}_{ij} \in \tilde{X}_i^{\Delta_i}$ ,  $|\tilde{X}_{ij}| < B^*$ . Therefore, for  $i = 1, \dots, m$ ,

$$\varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i})^{-1} < B_1 < \infty.$$

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

$$\partial \varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i}) / \partial \tilde{X}_{ij} = -\tilde{X}_i^{\Delta_i} \sum_{i|1}^{-1} w_{\Delta_i}(j) \varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i}).$$

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

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

$$\int_{M_i} (2\pi)^{n_i-\Delta_i} \det(\tilde{\Sigma}_i) \exp\{(t^{n_i-\Delta_i} - \tilde{\mu}_i)' \tilde{\Sigma}_i^{-1} (t^{n_i-\Delta_i} - \tilde{\mu}_i)/2\} dt^{n_i-\Delta_i}$$

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

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

$$|\int_{M_{i,-j}} - (2\pi)^{n_i-\Delta_i} \det(\tilde{\Sigma}_i) \exp\{(t_j^{n_i-\Delta_i} - \tilde{\mu}_i)' \tilde{\Sigma}_i^{-1} (t_j^{n_i-\Delta_i} - \tilde{\mu}_i)/2\} dt_{-j}^{n_i-\Delta_i}| < B_4$$

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

Using the definition of the derivative of an inverse function,

$$\partial \Phi^{-1}(H_{ij}) / \partial H_{ij} = -[\varphi\{\Phi^{-1}(H_{ij})\}]^{-1},$$

where  $\varphi$  is the density of the standard normal distribution and  $\Phi^{-1}$  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  $j = 1, \dots, n_i$  and  $i = 1, \dots, m$ ,  $|\partial \Phi^{-1}(H_{ij}) / \partial H_{ij}| < B_7 < \infty$  for some constant  $B_7$ . By condition C.5, for  $j = 1, \dots, n_i$  and  $i = 1, \dots, m$ ,  $\|Y_{ij} \exp\{\beta' Z_{ij}\}\|_\infty < B_8 < \infty$  and  $\|\int_0^\tau Y_{ij}(u) Z_{ij}(u) \exp\{\beta' Z_{ij}(u)\} d\Lambda(u)\| < B_9 < \infty$  for some constants  $B_8$  and  $B_9$ . Hence  $\|\zeta_m(\sigma, \Lambda)\|_\infty$  and  $\|\zeta_m(\beta, \sigma)\|$  are bounded by  $(B_1 B_2 + B_3 B_4) B_7 (B_8 + B_9) < \infty$

### Proof of Lemma 3

An expansion of  $\hat{l}_m(\sigma)$  around  $\Lambda_0$  and  $\beta_0$  can be written:

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

where  $R$  is a remainder term of order  $o_p\{\max(\|\hat{\Lambda} - \Lambda_0\|_\infty, \|\hat{\beta} - \beta_0\|\}$  and  $\zeta_m(\beta, \sigma)$  and  $\zeta_m(\sigma, \Lambda)(t)$  are defined in Lemma 2. Since  $\hat{\Lambda}$  is uniformly consistent and  $\hat{\beta}$  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  $\hat{\beta}$ , and the fact that  $\|\zeta_m(\beta, \sigma)\|$  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  $m^{-1} \sum_{i=1}^m \phi_i(\sigma)$ . Take  $\sigma \neq \sigma_0$ . The law of large numbers and Jensen's inequality imply that  $E\{l_{m0}(\sigma)\} - E\{l_{m0}(\sigma_0)\} = \lim_{m \rightarrow \infty} l_{m0}(\sigma) - l_{m0}(\sigma_0)$  which is strictly less than  $\log[E\{L^*(\sigma, \beta_0, A_0)/L^*(\sigma_0, \beta_0, A_0)\}] = 0$ .

**Proof of Lemma 5**

Let  $N(t, d, \mu, \Sigma^\dagger)$  be defined as  $(2\pi)^{-d/2} \det(\Sigma^\dagger)^{-1/2} \exp\{-\frac{1}{2}(t - \mu)'(\Sigma^\dagger)^{-1}(t - \mu)/2\} [tr\{(\Sigma^\dagger)^{-1} W_d\} - \frac{1}{2}(t - \mu)'(\Sigma^\dagger)^{-1} W_d(\Sigma^\dagger)^{-1}(t - \mu)/2]$ , where  $W_d$  is the  $d$  dimensional square matrix with zeros along the diagonal and ones off the diagonal. Let  $0^d$  denote a vector of length  $d$  of zeros. The score function can be written

$$U_m(\sigma) = m^{-1} \sum_{i=1}^m \varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i})^{-1} N(\tilde{X}_i^{\Delta_i}, \Delta_i, 0^{\Delta_i}, \sum_{i11}) + \tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{X}_i^{n_i - \Delta_i} | \tilde{X}_i^{\Delta_i})^{-1} \int_{M_i} N(t^{n_i - \Delta_i}, n_i - \Delta_i, \tilde{\mu}_i, \tilde{\Sigma}_i) dt^{n_i - \Delta_i}.$$

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  $\xi_m(\sigma, A)$  and  $\xi_m(\sigma, \beta)$ , starting with  $\xi_m(\sigma, A)$ . To make the argument more concrete express  $U_m(\sigma)$  as a function of  $A$  by writing  $U_m(\sigma, A) = U_m(\sigma)$ . Let  $\Gamma \in BV[0, \tau]$ .

To find the expression for the derivative, take a Taylor series expansion of  $U_m\{\sigma, A + \tau(\Gamma - A)\}$  around  $t = 0$  and evaluate the result at  $t = 1$ . The final expression is

$$U_m(\sigma, \Gamma) = U_m(\sigma, \Lambda) + \int_0^\tau \xi_m(\sigma, \Lambda)(u) d(\Lambda - \Gamma)(u), \text{ where } \xi_m(\sigma, \Lambda)(u) \text{ is equal to}$$

$$m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} D_{ij}^U Y_{ij}(u) \exp\{\beta' Z_{ij}(u)\} H_{ij}$$

and

$$\begin{aligned} D_{ij}^U &= (\Delta_{ij} [\partial \varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i})^{-1} / \partial \tilde{X}_{ij}] N(\tilde{X}_i^{\Delta_i}, \Delta_i, 0^{\Delta_i}, \sum_{i11}) + \varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i})^{-1} \\ &\times \{\partial N(\tilde{X}_i^{\Delta_i}, \Delta_i, 0^{\Delta_i}, \sum_{i11}) / \partial \tilde{X}_{ij}\}) + (1 - \Delta_{ij}) [\partial \tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{X}_i^{n_i - \Delta_i} | \tilde{X}_i^{\Delta_i})^{-1} / \partial \tilde{X}_{ij}] \\ &\times \int_{M_i} N(t^{n_i - \Delta_i}, n_i - \Delta_i, \tilde{\mu}_i, \tilde{\Sigma}_i) dt^{n_i - \Delta_i} + \tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{X}_i^{n_i - \Delta_i} | \tilde{X}_i^{\Delta_i})^{-1} \\ &\times \{\partial \int_{M_i} N(t^{n_i - \Delta_i}, n_i - \Delta_i, \tilde{\mu}_i, \tilde{\Sigma}_i) dt^{n_i - \Delta_i} / \partial \tilde{X}_{ij}\}) \sum_{j=1}^{n_i} \partial \Phi^{-1}(H_{ij}) / \partial H_{ij} \end{aligned}$$

Therefore the Hadamard derivative for  $T \in BV[0, \tau]$  is  $\int_0^\tau \xi_m(\sigma, \Lambda)(u) dT(u)$ . Direct calculation verifies that  $\xi_m(\sigma, \beta)$  is equal to

$$m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} D_{ij}^U \left[ \int_0^\tau Y_{ij}(u) Z_{ij}(u) \exp\{\beta' Z_{ij}(u)\} d\Lambda(u) \right] H_{ij}.$$

In Lemma 2 we showed that, for  $i = 1, \dots, m$ ,  $|\varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i})^{-1}| < B_1 < \infty$  and  $|\tilde{\Phi}_c^{n_i - \Delta_i}(\tilde{X}_i^{n_i - \Delta_i})^{-1}| < B_3 < \infty$ . Also, for  $j = 1, \dots, n_i$ ,  $i = 1, \dots, m$ ,  $|\partial \Phi^{-1}(H_{ij}) / \partial H_{ij}| < B_7 < \infty$ ,  $\|Y_{ij} \exp(\beta' Z_{ij})\|_\infty < B_8 < \infty$  and  $\left\| \int_0^\tau Y_{ij}(u) Z_{ij}(u) \exp\{\beta' Z_{ij}(u)\} d\Lambda(u) \right\| < B_9 < \infty$ .

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

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

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

$$|\tilde{X}_i^{\Delta_i} \sum_{i11}^{-1} w_{\Delta_i}(j) N(\tilde{X}_i^{\Delta_i}, \Delta_i, 0^{\Delta_i}, \sum_{i11}) + \tilde{X}_i^{\Delta_i} \sum_{i11}^{-1} \tilde{W}_{\Delta_i} \sum_{i11}^{-1} w_{\Delta_i}(j) \varphi_u^{\Delta_i}(\tilde{X}_i^{\Delta_i})|,$$

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  $j = \Delta_i + 1, \dots, n_i$  and  $i = 1, \dots, m$ ,

$|\partial \Phi_c^{n_i - \Delta_i}(\tilde{X}_i^{n_i - \Delta_i} | \tilde{X}_i^{\Delta_i})^{-1} / \partial \tilde{X}_{ij}|$  is equal to

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

for some constant  $B_{14}$ .

Using similar arguments as above one can directly show that for  $i = 1, \dots, m$ ,

$$\int_{M_i} N(t^{n_i - \Delta_i}, n_i - \Delta_i, \tilde{\mu}_i, \sum_i) dt^{n_i - \Delta_i} < B_{15} < \infty$$

for some constant  $B_{15}$ .

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

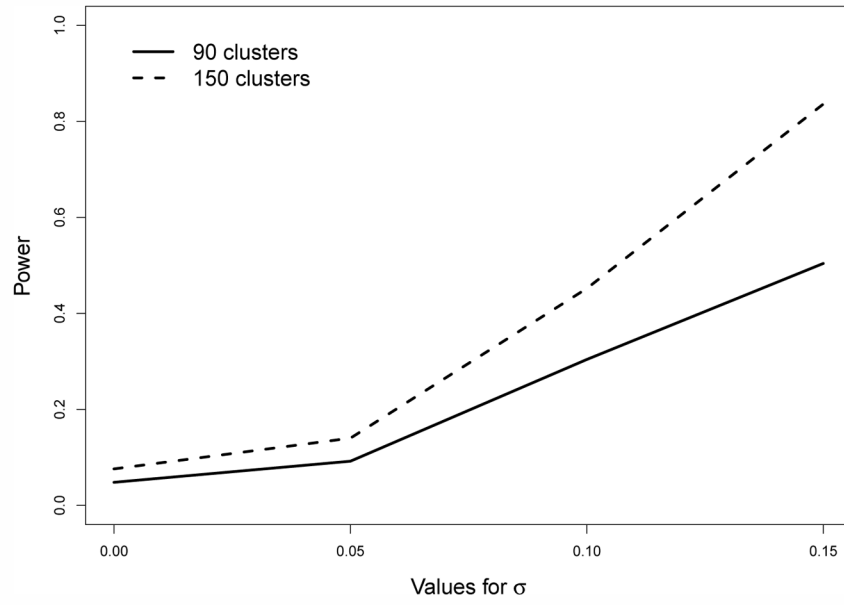
$$|\partial \int_{M_i} N(t^{n_i-\Delta_i}, n_i-\Delta_i, \tilde{\mu}_i, \sum_i) dt^{n_i-\Delta_i} / \partial \tilde{X}_{ij} = \int_{M_{i-j}} N(t_j^{n_i-\Delta_i}, n_i-\Delta_i, \tilde{\mu}_i, \sum_i) dt_{-j}^{n_i-\Delta_i}| < B_{15} < \infty$$

for some constant  $B_{16}$ .

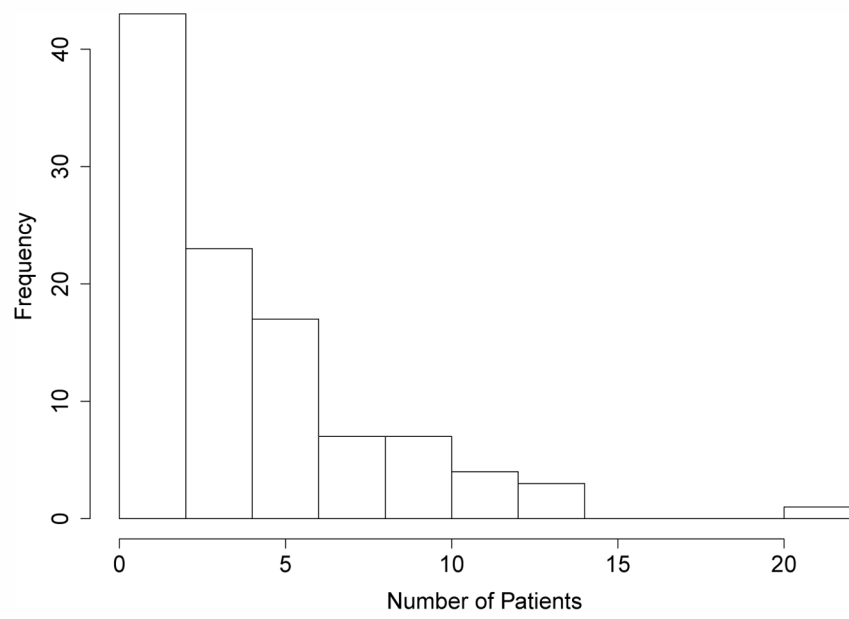
Hence  $\|\zeta_m(\sigma, A)\|_\infty$  and  $\|\zeta_m(\beta, \sigma)\|$  are bounded by  $(B_{11}B_{12}+B_1B_{13}+B_{14}B_{15}+B_3B_{16})B_7(B_8+B_9) < \infty$ .

## E Extended simulation results

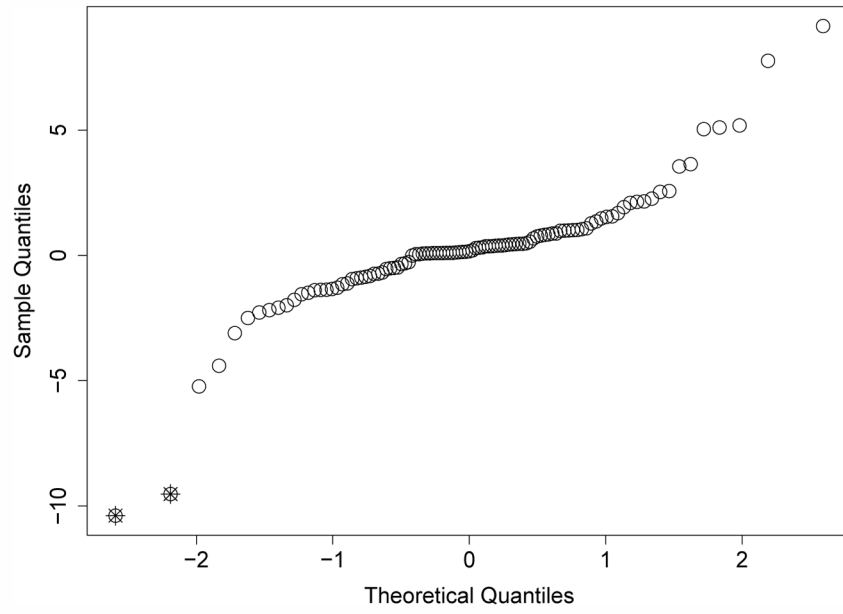
In the interest of space, the simulation results in the main body of the manuscript focus on the novel results for  $\sigma$ . We summarize the performance of  $\beta$  in Table 7. The true values for  $\beta_1$  and  $\beta_2$  are  $\log(0.5)$  and  $0.75$ , respectively. In all scenarios investigated, both the robust sandwich standard error and the jackknife standard error perform well. The jackknife standard error (SE) appears to match the Monte Carlo SE more closely compared to the robust SE, but the coverage probabilities are very similar and indicate appropriate coverage. Power for a Wald-type test with the jackknife SE is high across all the scenarios.



**Fig. 1.** Power curves as a function of  $\sigma$  and number of cluster.

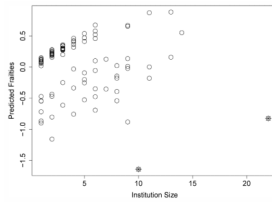


**Fig. 2.**  
A histogram of the number of patients in each institution.

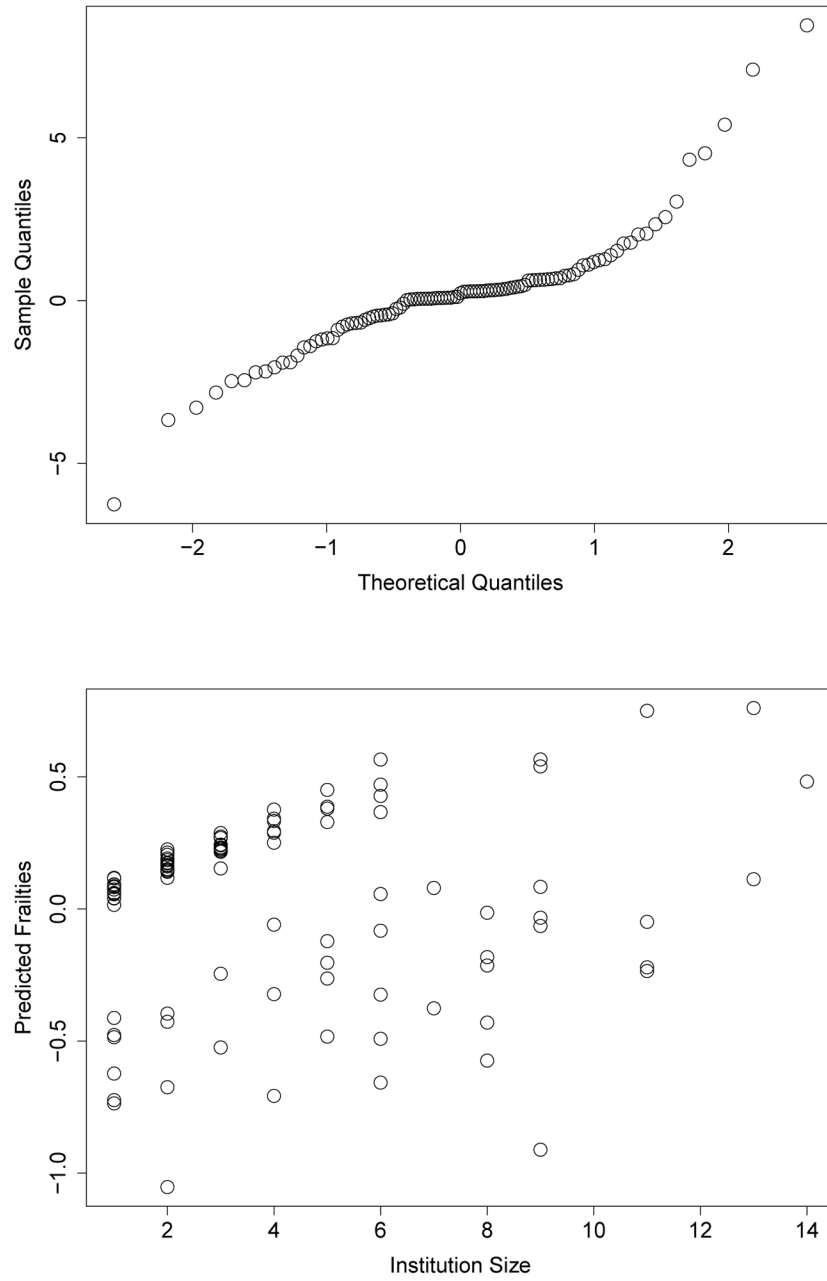


**Fig. 3.** A qq-plot of frailties from proposed model standardized by their standard errors. Two potential outlier frailties are marked with crosses.

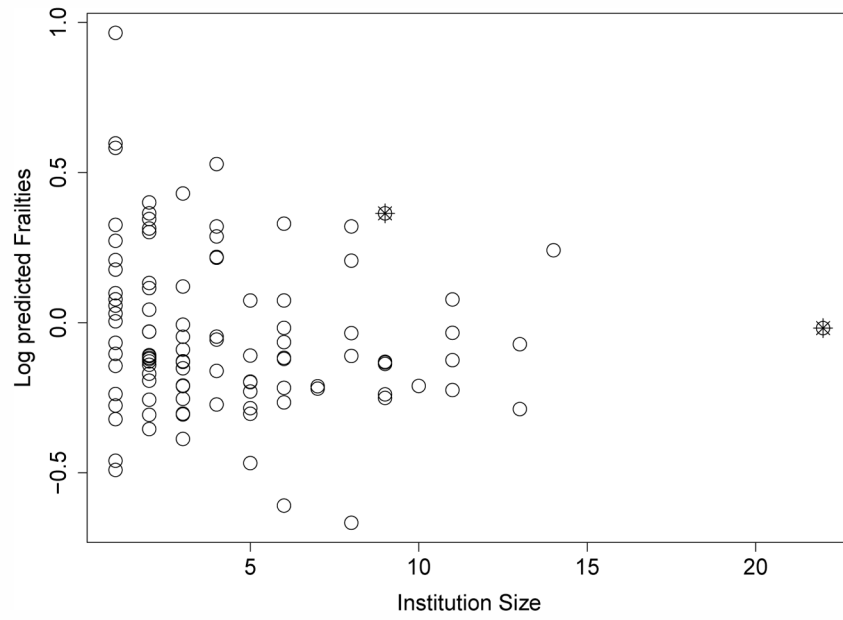




**Fig. 4.** A plot of institution size by frailty value from proposed model. The two frailties marked with crosses in Figure 3 are marked with crosses in this Figure.



**Fig. 5.** A qq-plot of standardized frailties from proposed model (left) and a plot of institution size by frailty value from proposed model (right) for the subset data.



**Fig. 6.** A plot of institution size by the log frailty value from gamma frailty model. Institutions identified as outliers in proposed methods are marked with crosses.

**Table 1**

Simulation results for the correlation terms

	Estimate	Jackknife SE	Monte Carlo SE	Coverage Probability
<b>45 clusters</b>				
$\sigma = 0$	-0.001	0.091	0.091	0.904
$\sigma = 0.05$	0.058	0.104	0.099	0.924
$\sigma = 0.10$	0.108	0.110	0.100	0.932
$\sigma = 0.15$	0.160	0.112	0.118	0.888
<b>60 clusters</b>				
$\sigma = 0$	0.000	0.076	0.078	0.942
$\sigma = 0.05$	0.049	0.086	0.082	0.928
$\sigma = 0.10$	0.104	0.094	0.083	0.944
$\sigma = 0.15$	0.154	0.096	0.091	0.964
<b>90 clusters</b>				
$\sigma = 0$	0.002	0.060	0.059	0.952
$\sigma = 0.05$	0.049	0.067	0.067	0.900
$\sigma = 0.10$	0.108	0.072	0.068	0.960
$\sigma = 0.15$	0.151	0.075	0.077	0.944
$\sigma = 0.50$	0.523	0.045	0.044	0.924
$\sigma = -0.10$	-0.109	0.067	0.068	0.930

**Table 2**

Summary of frailty simulations. Relative bias=bias/predicted value

	<b>Median relative bias</b>	<b>SE</b>
Median cluster size = 5	0.06	1.07
Median cluster size = 10	-0.05	1.02
Median cluster size = 15	-0.05	1.01
Median cluster size = 20	-0.06	1.01

**Table 3**

Distribution of patients with enlarged livers across treatment arms

<b>Strength</b>	<b>Duration</b>	<b>Number of Patients</b>
Standard	Standard	119
Standard	Double	104
Increased	Standard	117
Increased	Double	120

**Table 4**

Data analysis results using proposed model

Parameter	Estimate	SE <sub>J</sub>	P-value	SE <sub>S</sub>	Naive SE
<b>Treatment Only Model</b>					
Correlation	0.173	0.159	0.279	-	-
Increased Strength	-0.430	0.231	0.063	0.225	0.247
Increased Duration	0.419	0.231	0.070	0.226	0.245
<b>Larger Model</b>					
Correlation	0.207	0.200	0.300	-	-
Increased Strength	-0.249	0.482	0.605	0.440	0.374
Increased Duration	0.551	0.387	0.154	0.360	0.326
Interaction	-0.321	0.704	0.648	0.646	0.498
Age					
1-9	(ref)				
10-15	0.353	0.234	0.132	0.227	0.256
16+	0.485	0.739	0.512	0.565	0.451
Platelets ( $\times 10^3/\text{mm}^2$ )					
1-49	(ref)				
50-150	0.397	0.305	0.193	0.290	0.258
150+	-0.145	0.651	0.824	0.574	0.528

Estimate = log hazard ratios and estimates of  $\sigma$

SE<sub>J</sub> = jackknife based standard errors

P-value = two-sided p-value using SE<sub>J</sub>

SE<sub>S</sub> = sandwich-formula based standard errors (only for  $\beta$ )

Naive SE = SE assuming independence (only for  $\beta$ )

**Table 5**

Correlation estimates from proposed model for the full data and for data excluding two potential outliers.

	Treatment Only Model		Larger Model	
	All Data	Subset	All Data	Subset
Correlation	0.17	0.03	0.21	0.12
SE	0.16	0.13	0.20	0.16
P-value	0.28	0.81	0.30	0.44



**Table 6**

Data analysis results using gamma frailty model

Parameter	Estimate	SE	P-value
<b><u>Treatment Only Model</u></b>			
Frailty variance	0.440	0.374	0.12
Increased Strength	-0.455	0.251	0.071
Increased Duration	0.498	0.252	0.048
<b><u>Larger Model</u></b>			
Frailty variance	0.443	0.361	0.11
Increased Strength	-0.324	0.383	0.40
Increased Duration	0.594	0.337	0.08
Interaction	-0.256	0.512	0.62
Age			
1-9	(ref)		
10-15	0.418	0.264	0.11
16+	0.558	0.457	0.24
Platelets ( $\times 10^3/\text{mm}^2$ )			
1-49	(ref)		
50-150	0.361	0.266	0.17
150+	-0.087	0.541	0.11

**Table 7**

Simulation results for marginal parameters from the proportional hazards model. CP denotes coverage probability and SE denotes standard error.

	Estimate	Jackknife SE	Monte Carlo SE	Robust SE	Jackknife CP	Robust CP	Power	
<b>90 clusters</b>								
$\sigma = 0$	$\beta_1$	-0.700	0.140	0.136	0.182	0.964	0.960	0.996
	$\beta_2$	0.765	0.237	0.236	0.185	0.956	0.944	0.912
$\sigma = 0.05$	$\beta_1$	-0.711	0.138	0.124	0.183	0.964	0.964	1.000
	$\beta_2$	0.800	0.237	0.212	0.198	0.956	0.952	0.948
$\sigma = 0.10$	$\beta_1$	-0.696	0.140	0.140	0.185	0.952	0.944	1.000
	$\beta_2$	0.769	0.241	0.241	0.197	0.948	0.944	0.884
$\sigma = 0.15$	$\beta_1$	-0.710	0.141	0.127	0.182	0.972	0.960	1.000
	$\beta_2$	0.756	0.240	0.236	0.178	0.960	0.948	0.884
<b>150 clusters</b>								
$\sigma = 0$	$\beta_1$	-0.689	0.106	0.108	0.139	0.944	0.936	1.000
	$\beta_2$	0.743	0.182	0.173	0.141	0.956	0.956	0.988
$\sigma = 0.05$	$\beta_1$	-0.679	0.106	0.101	0.138	0.952	0.952	1.000
	$\beta_2$	0.770	0.182	0.172	0.141	0.968	0.968	0.996
$\sigma = 0.10$	$\beta_1$	-0.699	0.106	0.107	0.146	0.956	0.948	1.000
	$\beta_2$	0.753	0.183	0.178	0.144	0.964	0.952	0.988
$\sigma = 0.15$	$\beta_1$	-0.702	0.107	0.109	0.143	0.940	0.936	1.000
	$\beta_2$	0.742	0.183	0.178	0.138	0.948	0.940	0.964