

Test for Equality of Baseline Hazard Functions for Correlated Survival Data using Frailty Models

Polyna Khudyakov¹, Malka Gorfine² and Paul Feigin²

¹*Harvard School of Public Health, USA* and ²*Technion, Israel*

Abstract: In this work we provide a new class of test statistics for hypothesis testing of the equality of the baseline hazard functions for correlated survival data under frailty models. The asymptotic distribution of the test statistics is investigated theoretically under the null hypothesis and certain local alternatives. We also provide a simple variance estimator. The properties of the test statistics, under finite sample size, is studied by an extensive simulation study and we verify the control of Type I error and our proposed sample size formula. To the best of our knowledge, this is the first work for comparing the baseline hazard functions of correlated survival outcomes with covariates and frailty models. The utility of our proposed estimating technique is illustrated by the analysis of the call center data of an Israeli commercial company that processes up to 100,000 calls per day and the analysis of the breast cancer data of the Washington Ashkenazi Kin-Cohort family study.

Key words and phrases: Customer Patience, Frailty Model, Hypothesis Testing, Local Alternatives, Multivariate Survival Analysis.

1. Introduction

Call centers are intended to provide and improve customer service, marketing, technical support, etc. For a customer, addressing the call center actually means addressing the company itself, and any negative experience on the part of the customer can lead to the rejection of company products and services. Hence, it is very important to ensure that a call center works effectively and provides high quality service to its customers. Call centers collect a huge amount of data, and this provides a great opportunity for companies to use this information for the analysis of customer needs, desires, and intentions. Such data analysis can improve the quality of customer service and decrease costs.

This work was motivated by the analysis of customer patience, which we define as a willingness of customer to endure waiting in a queue before receiving service. The complication of customer patience analysis is that in most cases customers receive the required service before they lose their patience and we do not observe the customer patience. Such incomplete data motivates us to use statistical model under the setting of survival analysis with right censoring. In our context, an event is the customer abandonment of the system before being served. For a customer who receives service, his/her patience time is not fully observed and is considered as censored. Hence, for each customer, at each call, the observed time is the time until abandonment (patience time) or time until being served, whichever comes first. The data consists of customer calls with possibly multiple calls for a customer. We believe that the observed times of the same customer are not independent. Therefore, the Cox proportional hazard model cannot be used directly, and we use a well-known and popular approach that deals with clustered data - the frailty model approach (Ripatti and

Palmgren, 1978; Klein, 1992; Nielsen et al., 1992; Murphy, 1994; Parner, 1998, Hougaard, 2000 and references therein). The frailty model to be used in this work is the extended Cox model where the frailty variate acts multiplicatively on the baseline hazard function. In addition, different unspecified baseline hazard functions are being determined for each call, since it could be that customer behavior changes as s/he becomes more experienced with the system. Then, the main objective of this work is to provide a test statistic for comparing the cumulative baseline hazard functions.

A detailed review of estimation methods and frailty model testing can be found in Hougaard (2000). In particular, Nielsen et al. (1992) and Klein (1992) considered non-parametric maximum likelihood estimators under the gamma frailty. Murphy (1994) showed the consistency and asymptotic normality. Later, Parner (1998) extended these results to the model with covariates. Zeng and Lin (2007) presented an estimation technique for a class of semi-parametric regression models, which also includes random effects with any random effect distribution. They provided a semi-parametric maximum likelihood estimators, based on the EM algorithm, together with their asymptotic properties. A noniterative estimation procedure for estimating the hazard functions under any frailty distribution with finite moments was proposed by Gorfine et al. (2006). The detailed proof of the asymptotic properties of their proposed estimators was provided by Zucker et al. (2007). In this work we extend the method of Gorfine et al. (2006) to address the case of different baseline hazard functions.

The most popular test statistic for testing the equality of two hazard functions is the weighted log-rank test. Often the weighted log-rank statistics were constructed for independent samples (Lawless and Nadeau, 1995; Cook et al., 1996; Eng and Kosorok, 2005). Comparison of two treatments based on clustered data with no covariates was presented by Gangnon and Kosorok (2004). They used the weighted log-rank test statistic and presented a simple sample size formula. Song et al. (2008) studied a covariate-adjusted weighted log-rank statistic for recurrent events data while comparing between two independent treatment groups. To the best of our knowledge, so far there is no published work for comparing the baseline hazard functions of correlated survival outcomes with covariates and frailty models.

One of the most widely used sample size formulas for the log-rank test under the setting of two independent samples is that of Schoenfeld (1983). This formula was developed under the assumption that the hazard functions are not time varying. A sample-size formula while adopting the idea of Schoenfeld (1983) and extending the class of alternatives was presented by Fleming and Harrington (1991). Later, Kosorok and Lin (1999) proposed a class of contiguous alternatives for the power and sample size calculations. This class was used for sample size calculations for clustered survival data, with no covariates (Gangnon and Kosorok, 2004), for the supremum log-rank statistic (Eng and Kosorok, 2005) and for covariate-adjusted log-rank statistic for independent samples (Song et al., 2008). In all the above works, the sample size formula was developed under simplifying assumptions, such as assuming identical censoring distributions and consistent difference between the two hazard functions.

The rest of the article is organized as follows. Section 2 presents the notation and the model formulation. The estimation procedure and the asymptotic properties of the estimators are presented in Section 3. A new test for comparing two or more baseline hazard

functions in the case of clustered data is provided in Section 4. In Section 5 we propose a sample size formula for given significance level and power. The proofs and technical details are presented in the Appendix. The performance and utility of our proposed estimation technique, the test statistic and the sample size formula are illustrated in Section 6, by extensive simulation study. Then, in Section 7, we apply our approach to a real call center data set, and a breast cancer data set from the Washington Ashkenazi Kin-Cohort Study. Our conclusions are set out in Section 8.

2. Notation and formulation of the model

In what follows we use the terminology of call center data, although, it is obvious that this work is of practical importance in other fields, as is demonstrated in Section 7. Consider n customers, where customer i has m_i calls ($m_i \leq m$ for all $i = 1, \dots, n$). Later, we consider real data analysis with a maximum of 5 calls for each customer ($m = 5$). We assume that the waiting behavior of each customer does not depend on the waiting behavior of other customers. Let T_{ij}^0 and C_{ij} denote the failure and censoring times, respectively, for call j of individual i ($i = 1, \dots, n$, $j = 1, \dots, m_i$). The observed follow-up time is $T_{ij} = \min(T_{ij}^0, C_{ij})$, and the failure indicator is $\delta_{ij} = I(T_{ij}^0 \leq C_{ij})$. For call j of customer i we observe a vector of covariates Z_{ij} and assume that the waiting behavior of customer i ($i = 1, \dots, n$) is influenced by some additional unobservable subject-dependent properties which are represented by the frailty variate w_i .

The conditional hazard function of the patience of customer i at the j -th call given the frailty w_i , is assumed to take the form

$$\lambda_{ij}(t) = \lambda_{0j}(t)w_i e^{\beta^T Z_{ij}} \quad i = 1, \dots, n \quad j = 1, \dots, m_i, \quad (2.1)$$

where $\lambda_{0j}(t)$ is an unspecified baseline hazard function of call j and β is a p -dimensional vector of unknown regression coefficients. In this model, the baseline hazard functions are assumed to be different at each call, since it could be that customer behavior changes as s/he becomes more experienced with the system. It is also possible to consider a model with different regression coefficient vectors β_j , but for simplicity of presentation we assume that $\beta_j = \beta$, for all j . We also assume the following standard assumptions:

- (A.1) The frailty variate w_i is independent of m_i and Z_{ij} $j = 1, \dots, m_i$.
- (A.2) The frailty variates w_i $i = 1, \dots, n$ are independent and identically distributed random variables with a density of known parametric form, $f(w) \equiv f(w; \theta)$, where θ is an unknown vector of parameters.
- (A.3) The vector of covariates Z_{ij} is bounded.
- (A.4) The random vectors $(m_i, T_{i1}^0, \dots, T_{im_i}^0, C_{i1}, \dots, C_{im_i}, Z_{i1}, \dots, Z_{im_i}, w_i)$, $i = 1, \dots, n$, are independent and identically distributed, and the model will be built conditional on m_i $i = 1, \dots, n$.
- (A.5) Given Z_{ij} $j = 1, \dots, m_i$ and w_i , calls of customer i are independent.

(A.6) Given Z_{ij} $j = 1, \dots, m_i$ and w_i , the censoring is independent and noninformative for w_i and $(\beta, \{\Lambda_{0j}\}_{j=1}^{m_i})$.

3. Estimation

The main goal of this work is to provide a test for comparing two or more baseline hazard functions. However, our proposed test requires estimators of the unknown parameters: β , θ as well as $\{\Lambda_{0j}(t)\}_{j=1}^m$. A simple estimation procedure that provides consistent estimators is given in the next section.

3.1. The Proposed Estimation Procedure

Our estimation procedure is a direct extension of Gorfine et al. (2006) which handles any frailty distribution with finite moments. We extend this method to the case of different baseline hazard functions, $\lambda_{0j}(t)$, and describe it in short so the current paper will be self-contained.

According to our model (2.1), the full likelihood can be written as

$$L = \prod_{i=1}^n \prod_{j=1}^{m_i} \{\lambda_{0j}(T_{ij}) e^{\beta^T Z_{ij}}\}^{\delta_{ij}} \prod_{i=1}^n \int_0^\infty w^{N_{i\cdot}(\tau)} e^{-w H_{i\cdot}(\tau)} f(w) dw,$$

where τ is the maximal follow-up time, $H_{ij}(t) = \Lambda_{0j}(T_{ij} \wedge t) e^{\beta^T Z_{ij}}$, $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(s) ds$, $a \wedge b = \min(a, b)$, $H_{i\cdot}(t) = \sum_{j=1}^{m_i} H_{ij}(t)$, $N_{ij}(t) = \delta_{ij} I(T_{ij} \leq t)$ and $N_{i\cdot}(t) = \sum_{j=1}^{m_i} N_{ij}(t)$.

As in Gorfine et al. (2006), let $\gamma = (\beta^T, \theta)^T$, and for simplicity assume that θ is a scalar. If θ is a vector, the calculation can be derived in a similar way. The score vector, denoted by $U(\gamma, \{\Lambda_{0j}\}_{j=1}^m) = (U_1, \dots, U_p, U_{p+1})^T$, is determined as follows

$$U_r = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \left[Z_{ijr} \{\delta_{ij} - H_{ij}(T_{ij})\} \frac{\int_0^\infty w^{N_{i\cdot}(\tau)+1} \exp\{-w H_{i\cdot}(\tau)\} f(w) dw}{\int_0^\infty w^{N_{i\cdot}(\tau)} \exp\{-w H_{i\cdot}(\tau)\} f(w) dw} \right]$$

for $r = 1, \dots, p$, and

$$U_{p+1} = \frac{1}{n} \sum_{i=1}^n \frac{\int_0^\infty w^{N_{i\cdot}(\tau)} \exp\{-w H_{i\cdot}(\tau)\} f'(w) dw}{\int_0^\infty w^{N_{i\cdot}(\tau)} \exp\{-w H_{i\cdot}(\tau)\} f(w) dw},$$

where $f'(w) = df(w)/d\theta$. The estimation procedure consist of two main steps. One is to estimate γ by substituting estimators of $\{\Lambda_{0j}\}_{j=1}^m$ into the score equations $U(\gamma, \{\Lambda_{0j}\}_{j=1}^m) = 0$. The other is to estimate $\{\Lambda_{0j}\}_{j=1}^m$ given the estimated value of γ . To this end, we provide here the estimators of $\{\Lambda_{0j}\}_{j=1}^m$. Define $Y_{ij}(t) = I(T_{ij} \geq t)$ and the entire observed history $\mathcal{F}_t$ up to time t as

$$\mathcal{F}_t = \sigma\{N_{ij}(u), Y_{ij}(u), Z_{ij}, \quad i = 1, \dots, n; \quad j = 1, \dots, m_i; \quad 0 \leq u \leq t\}.$$

To simplify notation, we define $Z_{ij} = 0$ and $N_{ij}(t) = Y_{ij}(t) = 0$ for all $t \in [0, \tau]$ for each $m_i < j \leq m$ and $i = 1, \dots, n$. As shown in Parner (1998), applying the innovation theorem

(Andersen et al., 1993) to the observed history $\mathcal{F}_t$, the stochastic intensity process of $N_{ij}(t)$ with respect to $\mathcal{F}_t$ is given by

$$\lambda_{0j}(t) \exp(\beta^T Z_{ij}) Y_{ij}(t) \psi_i(t), \quad (3.1)$$

where

$$\psi_i(t) = E(w_i | \mathcal{F}_{t-}) = \frac{\int_0^\infty w^{N_{i \cdot}(t-)+1} e^{-w H_{i \cdot}(t-)} f(w) dw}{\int_0^\infty w^{N_{i \cdot}(t-)} e^{-w H_{i \cdot}(t-)} f(w) dw}.$$

It should be noted that $\psi_i(t)$ is a function of the unknown parameters γ and $\{\Lambda_{0j}\}_{j=1}^m$. Now, let $h_{ij}(t) = \psi_i(t) \exp(\beta^T Z_{ij})$ and note that given (3.1) it can be considered as a time-dependent covariate effect. Hence, the estimator of each Λ_{0j} is provided by using a Breslow-type (Breslow, 1974) estimator as follows. Let $\hat{\Lambda}_{0j}$ be a step function with jumps at the observed failure times τ_{jk} ($k = 1, \dots, K_j$, $j = 1, \dots, m$), where K_j is the total number of the distinct observed times of type j . Then, the jump size of $\hat{\Lambda}_{0j}$ at τ_{jk} given the value of $\hat{\gamma}$ is defined by

$$\Delta \hat{\Lambda}_{0j}(\tau_{jk}) = \left[\sum_{i=1}^n dN_{ij}(\tau_{jk}) \right] / \left[\sum_{i=1}^n \hat{h}_{ij}(\tau_{jk}) Y_{ij}(\tau_{jk}) \right], \quad (3.2)$$

where $\hat{h}_{ij}(t) = \hat{\psi}_i(t) \exp(\hat{\beta}^T Z_{ij})$ and in $\hat{\psi}_i(t)$ we substitute $\hat{\gamma}$ and $\{\hat{\Lambda}_{0j}(t)\}_{j=1}^m$ into $\psi_i(t)$. It is important to note that each value $\Delta \hat{\Lambda}_{0j}(\tau_{jk})$ is a function of $\{\hat{\Lambda}_{0j}(t)\}_{j=1}^m$, where $t < \tau_{jk}$. Therefore, the estimation procedure is based on ordering the observed failure times of all the calls in increasing order and estimating $\{\Lambda_{0j}\}_{j=1}^m$ sequentially, according to the order of the observed failure times.

To summarize, the following is our proposed estimation procedure. Provide initial value of γ , and proceed as follows:

Step 1 : Given the value of γ estimate $\{\Lambda_{0j}\}_{j=1}^m$ by using (3.2).

Step 2 : Given the value of $\{\Lambda_{0j}\}_{j=1}^m$, estimate γ by solving $U(\gamma, \{\hat{\Lambda}_{0j}\}_{j=1}^m) = 0$.

Step 3 : Repeat Steps 1 and 2 until convergence is reached with respect to $\{\hat{\Lambda}_{0j}\}_{j=1}^m$ and $\hat{\gamma}$.

For the choice of initial values for β we propose to use the naive Cox regression model, and for θ , take its value under the independence case. If the above integrals are not of closed analytical form, one can use numerical integration. As was already shown by Gorfine et al. (2006), such an approach avoids the use of iterative processes in estimating the cumulative baseline hazard functions.

3.2. Asymptotic properties

In this section, we formulate and summarize the asymptotic results of our proposed estimators. We denote by $\gamma^o = (\beta^{oT}, \theta^o)^T$ and $\Lambda_0^o(t) = \{\Lambda_{0j}^o(t)\}_{j=1}^m$ the true values of β ,

θ and $\Lambda_0(t) = \left\{ \Lambda_{0j}(t) \right\}_{j=1}^m$, respectively.

Claim 3.1. The estimator $\hat{\Lambda}_0(t) = \{\hat{\Lambda}_{0j}(t)\}_{j=1}^m$ converges almost surely to a limit $\Lambda_0(t, \gamma)$ uniformly in t and γ , with $\Lambda_0(t, \gamma) = \Lambda_0^o(t)$, and $n^{1/2}[\hat{\Lambda}_0(t) - \Lambda_0^o(t)]$ converges weakly to a Gaussian process.

Claim 3.2. The function $U[\gamma, \hat{\Lambda}_0(\cdot)]$ converges almost surely in t and γ to a limit $u[\gamma, \Lambda_0(\cdot)]$.

Claim 3.3. There exists a unique consistent root to $U[\hat{\gamma}, \hat{\Lambda}_0(\cdot)] = 0$.

Claim 3.4. The asymptotic distribution of $n^{1/2}(\hat{\gamma} - \gamma^o)$ is normal with mean zero and with a covariance matrix that can be consistently estimated by a sandwich estimator.

The proofs of Claims 3.1 - 3.4 along with all the required additional conditions are almost identical to those presented in Gorfine et al. (2006) and Zucker et al. (2007), since the only minor difference is the use of $\{\hat{\Lambda}_{0j}(t)\}_{j=1}^m$ instead of a global estimator based on all the calls together. Hence, the proofs and a detailed list of the additional required assumptions are omitted. It should be noted that although a consistent variance estimator of $\hat{\gamma}$ and $\left\{ \hat{\Lambda}_{0j}(t) \right\}_{j=1}^m$ can be provided, its form is very complicated. Hence, we recommend on using the bootstrap approach.

4. Family of weighted tests for correlated samples

4.1. Introduction and preliminaries

Our main objective is to provide a test statistic for comparing the cumulative baseline hazard functions corresponding to different calls. Namely, we are interested in testing the hypothesis

$$H_0 : \quad \Lambda_{01} = \Lambda_{02} = \dots = \Lambda_{0m} = \Lambda_0 \quad (4.1)$$

where Λ_0 is some unspecified cumulative hazard with $\Lambda_0(t) < \infty$. As noted earlier, the intensity processes of the counting processes $N_{ij}(t)$ $i = 1, \dots, n$, $j = 1, \dots, m_i$, with respect to $\mathcal{F}_t$ has the form $h_{ij}(t)Y_{ij}(t)\lambda_{0j}(t)$. However, given the frailty variate w_i , the intensity processes of $N_{ij}(t)$ $i = 1, \dots, n$, $j = 1, \dots, m_i$ take the form $\tilde{h}_{ij}(t)Y_{ij}(t)\lambda_{0j}(t)$ with $\tilde{h}_{ij}(t) = w_i \exp(\beta^T Z_{ij})$.

Let $\bar{Y}_j(t, \gamma) = \sum_{i=1}^n h_{ij}(t)Y_{ij}(t)$ and $\tilde{Y}_j(t, \gamma) = \sum_{i=1}^n \tilde{h}_{ij}(t)Y_{ij}(t)$, and note that $E\left[\sum_{i=1}^n w_i Y_{ij}(t) e^{\beta^T Z_{ij}}\right] = E\left[\sum_{i=1}^n \psi_i Y_{ij}(t) e^{\beta^T Z_{ij}}\right]$. Then, by the uniform strong law of large numbers Andersen and Gill (1982), the functions $n^{-1}\bar{Y}_j(t, \gamma)$ and $n^{-1}\tilde{Y}_j(t, \gamma)$ converge to the same function, if one of them converges.

For deriving the asymptotic properties of our proposed test statistic, we make the following assumptions:

A.7 $\hat{W}_n(s)$ is nonnegative, cadlag or caglad, with bounded total variation, and converges in probability to some uniformly bounded integrable function $W(s)$, that is

$$\sup_{s \in [0, \tau]} \left| \hat{W}_n(s) - W(s) \right| \rightarrow 0.$$

A.8 There exist positive deterministic functions $\bar{y}_j(s)$, $j = 1, \dots, m$, such that

$$\sup_{s \in [0, \tau]} |n^{-1} \bar{Y}_j(s, \gamma^o) - \bar{y}_j(s)| \rightarrow 0 \quad \sup_{s \in [0, \tau]} |n^{-1} \tilde{Y}_j(s, \gamma^o) - \bar{y}_j(s)| \rightarrow 0,$$

$j = 1, \dots, m$ almost surely, as $n \rightarrow \infty$.

A.9 $Q_{lj}(s, \gamma^o) = \frac{\partial}{\partial \gamma_l} [\bar{Y}_j(s, \gamma) / \bar{Y}_\cdot(s, \gamma)]_{\gamma=\gamma^o}$ $l = 1, \dots, p+1$ $j = 1, \dots, m$ are bounded over $[0, \tau]$ where $\bar{Y}_\cdot(s, \gamma^o) = \sum_{j=1}^m \bar{Y}_j(s, \gamma^o)$.

A.10 There exist deterministic functions $g_{lj}(s)$, $l = 1, \dots, p+1$ $j = 1, \dots, m$, such that

$$\sup_{s \in [0, \tau]} |Q_{lj}(s, \gamma^o) - g_{lj}(s)| \rightarrow 0$$

almost surely, as $n \rightarrow \infty$.

4.2. Test for equality of two hazard functions

We start by comparing the cumulative baseline hazard functions of two calls. In this subsection we use indices 1 and 2 for comparing any two baseline hazard functions out of the m possible functions. The extension to more than two calls will follow. Assume we are interested in testing the hypothesis

$$H_0 : \quad \Lambda_{01} = \Lambda_{02} = \Lambda_0. \quad (4.1)$$

We propose to use the weighted log-rank statistic (Fleming and Harrington, 1991) that takes the form

$$\begin{aligned} S_n(t, \hat{\gamma}) &= \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \frac{\bar{Y}_1(s, \hat{\gamma}) \bar{Y}_2(s, \hat{\gamma})}{\bar{Y}_\cdot(s, \hat{\gamma})} \left\{ d\hat{\Lambda}_{01}(s) - d\hat{\Lambda}_{02}(s) \right\} \\ &= \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \frac{\bar{Y}_1(s, \hat{\gamma}) \bar{Y}_2(s, \hat{\gamma})}{\bar{Y}_\cdot(s, \hat{\gamma})} \left\{ \frac{d\bar{N}_1(s)}{\bar{Y}_1(s, \hat{\gamma})} - \frac{d\bar{N}_2(s)}{\bar{Y}_2(s, \hat{\gamma})} \right\}, \end{aligned} \quad (4.2)$$

for $t \in [0, \tau]$ where $\bar{Y}_\cdot(s, \hat{\gamma}) = \bar{Y}_1(s, \hat{\gamma}) + \bar{Y}_2(s, \hat{\gamma})$, $d\bar{N}_j(s) = \sum_{i=1}^n dN_{ij}(s)$ and the estimators $\hat{\gamma}$ and $\{\hat{\Lambda}_{0j}\}_{j=1}^m$ are as defined in Section 3. In practice, one can choose t to be the smallest s such that $\prod_{j=1}^m \bar{Y}_j(s, \hat{\gamma}) = 0$ or any value that is of practical importance.

For deriving the asymptotic distribution of $S_n(t, \hat{\gamma})$ it is important to note that given w_i and the intensity process $\tilde{h}_{ij}(t)Y_{ij}(t)\lambda_{0j}(t)$, the process $M_{ij}(t) = N_{ij}(t) - w_i \int_0^t \lambda_{0j}(u) e^{\beta^T Z_{ij}} Y_{ij}(u) du$ is a mean-zero martingale with respect to $\mathcal{F}_{t-}$. Then, given $w_\cdot = \{w_i\}_{i=1}^n$, the sum of these martingales $\bar{M}_j(t) = \sum_{i=1}^n M_{ij}(t)$ is also a mean-zero martingale with respect to $\mathcal{F}_{t-}$. Since $N_{i1}(t)$ and $N_{i2}(t)$ are conditionally independent given w_i for all $i = 1, \dots, n$, then, given $w_\cdot$, $\bar{M}_1(t)$ and $\bar{M}_2(t)$ are uncorrelated martingales.

To simplify the notation, we define

$$\mathcal{G}(s, \gamma) = \frac{\bar{Y}_1(s, \gamma) \bar{Y}_2(s, \gamma)}{\bar{Y}_\cdot(s, \gamma)}, \quad D_n(s, \gamma) = \frac{\hat{W}_n(s)}{\sqrt{n}} \mathcal{G}(s, \gamma), \quad D_j^n(s, \gamma) = \frac{\hat{W}_n(s)}{\sqrt{n}} \frac{\mathcal{G}(s, \gamma)}{\bar{Y}_j(s, \gamma)}$$

for $j = 1, 2$. For the asymptotic distribution of our test statistic $S_n(\tau, \hat{\gamma})$ and its variance estimator, we start with the following theorem. The proof is presented in Appendix 1.1.

Theorem 4.1 *Given Assumptions A.1-A.6 and A.9-A.10, the test statistic $S_n(t, \hat{\gamma})$ has the same asymptotic distribution as $\tilde{S}_n(t, \gamma^o) + S_n^{**}(t)$, where*

$$\begin{aligned}\tilde{S}_n(t, \gamma^o) &= \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \mathcal{G}(s, \gamma^o) \left\{ \frac{d\bar{M}_1(s)}{\bar{Y}_1(s, \gamma^o)} - \frac{d\bar{M}_2(s)}{\bar{Y}_2(s, \gamma^o)} \right\}, \\ S_n^{**}(t) &= \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \mathcal{G}(s, \hat{\gamma}) \left\{ \frac{\tilde{Y}_1(s, \gamma^o) d\Lambda_{01}(s)}{\bar{Y}_1(s, \hat{\gamma})} - \frac{\tilde{Y}_2(s, \gamma^o) d\Lambda_{02}(s)}{\bar{Y}_2(s, \hat{\gamma})} \right\}.\end{aligned}$$

Now, consider $S_n^{**}(t)$. By the first order Taylor expansion about γ^o we get

$$\begin{aligned}S_n^{**}(t) &\approx \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \left[\frac{\bar{Y}_2(s, \gamma^o) \tilde{Y}_1(s, \gamma^o)}{\bar{Y}(s, \gamma^o)} d\Lambda_{01}(s) - \frac{\bar{Y}_1(s, \gamma^o) \tilde{Y}_2(s, \gamma^o)}{\bar{Y}(s, \gamma^o)} d\Lambda_{02}(s) \right] \\ &+ \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \left\{ \tilde{Y}_1(s, \gamma^o) \mathbf{Q}_1^T(s, \gamma^o) d\Lambda_{01}(s) - \tilde{Y}_2(s, \gamma^o) \mathbf{Q}_2^T(s, \gamma^o) d\Lambda_{02}(s) \right\} (\hat{\gamma} - \gamma^o),\end{aligned}\tag{4.3}$$

where $\mathbf{Q}_j^T(s, \gamma^o) = (Q_{1j}, \dots, Q_{(p+1)j})$, $j = 1, 2$. The second term of the right-hand side of (4.3) represents the additional variability of $S_n(t, \hat{\gamma})$ due to $\hat{\gamma}$. Based on Claim 3.4 it is easy to see that it is asymptotically normal with mean zero. However, this term is expected to be of a negligible contribution to the total variance, since, $\hat{\gamma}$ is being estimated parametrically (Acar et al., 2010, Section 2.3). It should be noted that our extensive simulation study, presented in Section 6, also supports this argument, as will be discussed there. To summarize, we formulate the following conclusion.

Conclusion 1. An approximation of the asymptotic distribution of $S_n(t, \hat{\gamma})$ is the asymptotic distribution of $\tilde{S}_n^*(t, \gamma^o) = \tilde{S}_n(t, \gamma^o) + S_n^*(t, \gamma^o)$, where

$$S_n^*(t, \gamma^o) = \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \mathcal{G}(s, \gamma^o) \left[\frac{\tilde{Y}_1(s, \gamma^o)}{\bar{Y}_1(s, \gamma^o)} d\Lambda_{01}(s) - \frac{\tilde{Y}_2(s, \gamma^o)}{\bar{Y}_2(s, \gamma^o)} d\Lambda_{02}(s) \right].$$

We derive the asymptotic null distribution of $\tilde{S}_n^*(t, \gamma^o)$ by the asymptotic distribution of each of the above two terms. For this end, consider the following theorem. The proof is sketched in Appendix 1.2.

Theorem 4.2 *Given Assumptions A.1-A.8 and under the null hypothesis,*

- (1) $\tilde{S}_n(t, \gamma^o)$ converges to a zero-mean normally distributed random variable with finite variance $\sigma_{\tilde{S}}^2(t)$, as $n \rightarrow \infty$, where

$$\sigma_{\tilde{S}}^2(t) = \int_0^t W^2(s) \frac{\bar{y}_1(s) \bar{y}_2(s)}{\bar{y}_1(s) + \bar{y}_2(s)} d\Lambda_0(s).\tag{4.4}$$

- (2) $S_n^*(t, \gamma^o)$ converges to a zero-mean random variable with finite variance $\sigma_{S^*}^2(t)$ as $n \rightarrow \infty$.

- (3) The two random variables, $\tilde{S}_n(t, \gamma^o)$ and $S_n^*(t, \gamma^o)$, are uncorrelated.

Summarizing the results of Conclusion 1 and Theorem 4.2, under the null hypothesis our test statistic $S_n(t, \hat{\gamma})$ is asymptotically zero-mean normally distributed random variable, and its asymptotic variance can be approximated by $Var\{\tilde{S}_n(t, \gamma^o)\} + Var\{S_n^*(t, \gamma^o)\}$. Thus, based on direct calculations of the variances, as presented in Appendix 1.3, we provide the following variance estimator of $S_n(t, \hat{\gamma})$

$$\begin{aligned} \hat{\sigma}_I^2(t) = & \int_0^t \hat{W}_n^2(s) \sum_{j=1}^2 \left\{ D_j^n(s, \hat{\gamma}) \right\}^2 \sum_{i=1}^n e^{\hat{\beta}^T Z_{ij}} Y_{ij}(s) \hat{E}(w_i) \frac{d\bar{N}_j(s)}{\bar{Y}_j(s, \hat{\gamma})} \\ & + \sum_{j=1}^2 \int_0^t \int_0^t D_j^n(s, \hat{\gamma}) D_j^n(u, \hat{\gamma}) \sum_{i=1}^n Y_{ij}(s \vee u) e^{2\hat{\beta}^T Z_{ij}} \widehat{Var}(w_i \mid \mathcal{F}_{s \vee u-}) d\hat{\Lambda}_{0j}(s) d\hat{\Lambda}_{0j}(u) \\ & - 2 \int_0^t \int_0^t D_1^n(s, \hat{\gamma}) D_2^n(u, \hat{\gamma}) \sum_{i=1}^n Y_{i1}(s) Y_{i2}(u) e^{\hat{\beta}^T Z_{i\cdot}} \widehat{Var}(w_i \mid \mathcal{F}_{s \vee u-}) d\hat{\Lambda}_{01}(s) d\hat{\Lambda}_{02}(u). \end{aligned}$$

The first component of $\hat{\sigma}_I^2(t)$ is the estimator of $Var\{\tilde{S}_n(t, \gamma^o)\}$, and the two other components are the estimator of $Var\{S_n^*(t, \gamma^o)\}$. For the unconditional expectation estimator $\hat{E}(w_i)$ and the conditional variance estimator $\widehat{Var}(w_i \mid \mathcal{F}_t)$ one can use $\hat{\gamma}$ and $\{\hat{\Lambda}_{0j}(\cdot)\}_{j=1}^m$. Also, it should be noted that often $E(w_i)$ is set to be 1 for the model (2.1) to be identifiable. In these cases $\hat{E}(w_i) = 1$ $i = 1, \dots, n$ (for a comprehensive discussion of identifiability in frailty models, the reader is referred to Hougaard (2000, Section 7.2)). However, as we show later by extensive simulation study, $Var\{S_n^*(t, \gamma^o)\}$ is of a negligible contribution to the total variance (less than 10%). Hence, we recommend to estimate the variance of the test statistic $S_n(t, \hat{\gamma})$ by the estimator of $Var\{\tilde{S}_n(t, \gamma^o)\}$. Specifically,

$$\hat{\sigma}_{II}^2(t) = \int_0^t \hat{W}_n^2(s) \sum_{j=1}^2 \left\{ D_j^n(s, \hat{\gamma}) \right\}^2 \sum_{i=1}^n e^{\hat{\beta}^T Z_{ij}} Y_{ij}(s) \hat{E}(w_i) d\hat{\Lambda}_{0j}(s). \quad (4.5)$$

In conclusion, our proposed test statistic is defined by $S_n(t, \hat{\gamma})/\hat{\sigma}_{II}(t)$ (or $S_n(t, \hat{\gamma})/\hat{\sigma}_I(t)$) and the rejection region corresponding to the null hypothesis (4.1) should be defined by the standard normal distribution.

4.3. Test for equality of m hazard functions

Now we extend the test proposed in the previous section to test the null hypothesis (4.1) with $m > 2$ baseline hazard functions. Namely, we compare each of the m estimators of the cumulative baseline hazard functions $\left\{ \hat{\Lambda}_{0j} \right\}_{j=1}^m$ with an estimator of the common cumulative baseline hazard function constructed under the null hypothesis. Let $\hat{\Lambda}_0$ be the estimated cumulative baseline hazard function under the null hypothesis (Gorfine et al., 2006) in which the jump size of $\hat{\Lambda}_0$ at time s is defined by $\Delta \hat{\Lambda}_0(s) = \sum_{j=1}^m d\bar{N}_j(s)/\bar{Y}(s, \hat{\gamma})$. We define $\mathbf{S}_n(t, \hat{\gamma}) = (S_{n1}(t, \hat{\gamma}), \dots, S_{nm}(t, \hat{\gamma}))^T$ to be the m -sample statistic. In the spirit of (4.2), we define

$$S_{nj}(t, \hat{\gamma}) = \frac{1}{\sqrt{n}} \int_0^t \hat{W}_{nj}(s) \frac{\bar{Y}_j(s, \hat{\gamma}) \bar{Y}(s, \hat{\gamma})}{\bar{Y}_j(s, \hat{\gamma}) + \bar{Y}(s, \hat{\gamma})} \left\{ d\hat{\Lambda}_{0j}(s) - d\hat{\Lambda}_0(s) \right\} \quad j = 1, \dots, m,$$

where $\hat{W}_{nj}(s)$ are nonnegative cadlag or caglad with total bounded variation. However, the special choice of weight processes such as $\hat{W}_{nj}(s) = \hat{W}_n(s)\{\bar{Y}_j(s, \hat{\gamma}) + \bar{Y} \cdot(s, \hat{\gamma})\}/\bar{Y} \cdot(s, \hat{\gamma})$ $j = 1, \dots, m$, where $\hat{W}_n(s)$ is nonnegative cadlag or caglad with total bounded variation, covers a wide variety of interesting cases (Andersen et al., 1993, Section V.2). Hence, the above choice of weight process will be considered here. Then,

$$S_{nj}(t, \hat{\gamma}) = \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \bar{Y}_j(s, \hat{\gamma}) \left\{ d\hat{\Lambda}_{0j}(s) - d\hat{\Lambda}_0(s) \right\} \quad j = 1, \dots, m,$$

and $\sum_{j=1}^m S_{nj}(t, \hat{\gamma}) = 0$. It is easy to verify that for $m = 2$, $S_{n1}(t, \hat{\gamma})$ equals (4.2). Similar arguments used in the case of comparing two baseline hazard functions can be used here, such that we arrive to the following conclusion.

Conclusion 2. An approximation of the asymptotic distribution of $\mathbf{S}_n(t, \hat{\gamma})$ is the asymptotic distribution of $\tilde{\mathbf{S}}_n^*(t, \gamma^o) = \tilde{\mathbf{S}}_n(t, \gamma^o) + \mathbf{S}_n^*(t, \gamma^o)$, where the respective j -th components of $\tilde{\mathbf{S}}_n(t, \gamma^o)$ and $\mathbf{S}_n^*(t, \gamma^o)$ are

$$\tilde{S}_{nj}(t, \gamma^o) = \int_0^t \frac{\hat{W}_n(s)}{\sqrt{n}} \bar{Y}_j(s, \gamma^o) \left\{ \frac{d\bar{M}_j(s)}{\bar{Y}_j(s, \gamma^o)} - \frac{d\bar{M} \cdot(s)}{\bar{Y} \cdot(s, \gamma^o)} \right\}, \quad (4.2)$$

$$S_{nj}^*(t, \gamma^o) = \int_0^t \frac{\hat{W}_n(s)}{\sqrt{n}} \bar{Y}_j(s, \gamma^o) \left\{ \frac{\tilde{Y}_j(s, \gamma^o) d\Lambda_{0j}(s)}{\bar{Y}_j(s, \gamma^o)} - \frac{\tilde{Y} \cdot(s, \gamma^o) d\Lambda_0(s)}{\bar{Y} \cdot(s, \gamma^o)} \right\}, \quad (4.3)$$

where $\bar{M} \cdot(s) = \sum_{i=1}^n \bar{M}_i(s)$.

For the asymptotic distribution of $\tilde{\mathbf{S}}_n^*(t, \gamma^o)$ we present the following theorem. A sketch of the proof can be found in the Appendix 1.4.

Theorem 4.3 *Given Assumptions A.1-A.8 and under the null hypothesis,*

- (1) $\tilde{\mathbf{S}}_n(t, \gamma^o)$ converges to a zero-mean multivariate normally distributed random variable with variance matrix $\mathbf{V}(t)$ and its jk -th component is defined by

$$V_{jk}(t) = \begin{cases} \int_0^t W^2(s) \frac{\bar{y}_j(s) \sum_{r \neq j, r=1}^m \bar{y}_r(s)}{\bar{y} \cdot(s)} \lambda_0(s) ds & k = j \\ - \int_0^t W^2(s) \frac{\bar{y}_j(s) \bar{y}_k(s)}{\bar{y} \cdot(s)} \lambda_0(s) ds & k \neq j. \end{cases}$$

- (2) $\mathbf{S}_n^*(t, \gamma^o)$ converges to a zero-mean multivariate normal random variable with covariance matrix having finite diagonal entries and zero valued non-diagonal entries.

- (3) $\tilde{\mathbf{S}}_n(t, \gamma^o)$ and $\mathbf{S}_n^*(t, \gamma^o)$ are uncorrelated.

Summarizing our results so far, under the null hypothesis (4.1), $\mathbf{S}_n(t, \hat{\gamma})$ is asymptotically normal with mean zero. Using similar arguments as for the case of testing the equality of two hazard functions, we estimate the variance of $\mathbf{S}_n(t, \hat{\gamma})$ based on the variance estimator of $\tilde{\mathbf{S}}_n(t, \gamma^o)$. Hence our proposed estimator, denoted by $\hat{\mathbf{V}}(t)$, is given by

$$\hat{V}_{jj}(t) = \frac{1}{n} \int_0^t \hat{W}_n^2(s) \sum_{i=1}^n \left[\left\{ 1 - \frac{\bar{Y}_j(s, \hat{\gamma})}{\bar{Y} \cdot(s, \hat{\gamma})} \right\}^2 E_{ij}(s) + \left\{ \frac{\bar{Y}_j(s, \hat{\gamma})}{\bar{Y} \cdot(s, \hat{\gamma})} \right\}^2 \sum_{l \neq j}^m E_{il}(s) \right] \quad j = 1, \dots, m \quad (4.4)$$

and for $k \neq j$

$$\begin{aligned} \hat{V}_{kj}(t) = \frac{1}{n} \int_0^t \hat{W}_n^2(s) & \left[\sum_{l \neq j, k} \frac{\bar{Y}_j(s, \hat{\gamma}) \bar{Y}_k(s, \hat{\gamma})}{\bar{Y}^2(s, \hat{\gamma})} \hat{E}_{il}(s) - \frac{\bar{Y}_j(s, \hat{\gamma})}{\bar{Y}(s, \hat{\gamma})} \sum_{i=1}^n \hat{E}_{ik}(s) \right. \\ & \left. - \frac{\bar{Y}_k(s, \hat{\gamma})}{\bar{Y}(s, \hat{\gamma})} \sum_{i=1}^n \hat{E}_{ij}(s) \right], \quad k, j = 1, \dots, m, \end{aligned} \quad (4.5)$$

where $E_{ij}(s) = \hat{E}(w_i)Y_{ij}(s)e^{\hat{\beta}^T Z_{ij}}d\hat{\Lambda}_{0j}(s)$. The details of the derivation of $\hat{\mathbf{V}}(\tau)$ are presented in Appendix 1.5. Clearly, $\mathbf{V}(t)$ has a rank of $m - 1$. Hence, we define $\hat{\mathbf{V}}^o(t)$ as a $(m - 1) \times (m - 1)$ matrix obtained by deleting the last row and column of $\hat{\mathbf{V}}(t)$. Also, let $\mathbf{S}_n^o(t) = (S_{n1}(t, \hat{\gamma}), \dots, S_{n(m-1)}(t, \hat{\gamma}))^T$. Then, our proposed test statistic is defined by $\mathbf{S}_n^o(t, \hat{\gamma})^T [\hat{\mathbf{V}}^o(t)]^{-1} \mathbf{S}_n^o(t, \hat{\gamma})$ and the rejection region should be defined by the $\chi^2(m - 1)$ distribution.

5. Sample size formula for equality of two hazard functions

In this section, we present a sample size formula under proportional means local alternative and certain simplifying assumptions for testing the equality of two baseline hazard functions. Specifically, let

$$H_1 : \quad \Lambda_{0j}^n(s) = \int_0^s \exp\{(-1)^{j-1} \varphi(u)/(2\sqrt{n})\} d\Lambda_0(u) \quad j = 1, 2 \quad \text{for all } s \in [0, \tau],$$

where Λ_0 is some unspecified cumulative hazard function with $\Lambda_0(s) < \infty$ and $\varphi(s) \neq 0$ for all $s \in [0, \tau]$. The above local alternatives formulation was originally proposed by Kosorok and Lin (1999) and also can be found in Gangnon and Kosorok (2004).

It is easy to verify that the above Λ_{0j}^n $j = 1, 2$ satisfies the following assumptions:

$$A.11 \quad \text{For } j = 1, 2 \quad \sup_{s \in [0, \tau]} |d\Lambda_{0j}^n(s)/d\Lambda_0(s) - 1| \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

$$A.12 \quad \text{As } n \rightarrow \infty, \quad \sup_{s \in [0, \tau]} |\sqrt{n} \{d\Lambda_{01}^n(s)/d\Lambda_{02}^n(s) - 1\} - \varphi(s)| \rightarrow 0, \quad \text{where } \varphi \text{ is either} \\ \text{cadlag or caglad with bounded total variation.}$$

We start with the asymptotic distribution of $\tilde{S}_n^*(t, \gamma^o)$, under the above local alternatives.

Theorem 5.1 *Given Assumptions A.1-A.12, $S_n(t, \gamma^o)$ converges in distribution to a normal random variable with mean $\mu_1(t)$ and variance $\sigma^2(t)$, where*

$$\mu_1(t) = \int_0^t W(s) \varphi(s) \frac{\bar{y}_1(s) \bar{y}_2(s)}{\bar{y}_1(s) + \bar{y}_2(s)} d\Lambda_0(s) \quad (5.1)$$

and $\sigma^2(t) = \sigma_{\tilde{S}}^2(t)$ as defined in (4.4).

A sketched proof of Theorem 5.1 is given in the Appendix 1.6.

Under the above contiguous alternative, we can approximate the power calculation as follows. For a fixed alternative, set $\varphi(t) = \sqrt{n} \varphi^*(t)$. Then, by (5.1) and the first

order Taylor expansion we get that the expectation of $S_n(t, \hat{\gamma})$ under the alternative equals $\sqrt{n}\mu_1^*(t) + o(\sqrt{n})$, where $\mu_1^*(t) = \int_0^t W(s)\varphi^*(s) \frac{\bar{y}_1(s)\bar{y}_2(s)}{\bar{y}_1(s) + \bar{y}_2(s)} d\Lambda_0(s)$. Now, based on the limiting distribution of $S_n(\tau, \hat{\gamma})$, and given significance level α and power π , we get the following sample size formula

$$n = \left(Z_{1-\alpha/2} + Z_\pi \right)^2 \sigma^2(t) / \{\mu_1^*(t)\}^2 \quad (5.2)$$

where Z_p is the p -quantile of the standard normal distribution.

However, in order to calculate the required sample size based on (5.2) one should estimate $\sigma^2(t)$ and $\mu_1^*(t)$ based on a pilot study or existing relevant data sets. In what follows, we propose simple estimators under simplifying assumptions, similar to those of Song et al. (2008). These simple estimators provide a practical sample size formula.

Assume that the baseline hazard functions are continuous and the local alternatives satisfy $\varphi^*(s) = \varepsilon$ for all $s \in [0, \tau]$, where $\varepsilon \in \mathbb{R}$ and the weight function is constant $\hat{W}_n(s) \equiv 1$. We also assume that the limiting values of $\bar{Y}_j(s, \gamma)/n_j$ are $\pi_j(s)$, $j = 1, 2$ and the proportion of customers making the j -th call, n_j/n , converges to $p_j \in (0, 1]$, $j = 1, 2$. Then, based on Assumption A.8, we replace $\bar{y}_j(s)$ by $p_j\pi_j(s)$. In addition, we assume that $\pi_1(s) = \pi_2(s) = \pi(s)$. Hence, $\mu_1^*(t) = \varepsilon R(t)p_1p_2/(p_1 + p_2)$, where $R(t) = \int_0^t \pi(s)d\Lambda_0(s)$. A simple estimator of $R(t)$ can be obtained by $\hat{R}(t) = \int_0^t \{\hat{p}_1\hat{\pi}_1(s)d\hat{\Lambda}_{01}(s) + \hat{p}_2\hat{\pi}_2(s)d\hat{\Lambda}_{02}(s)\} = n^{-1} \sum_{j=1}^2 \bar{N}_j(t)$. Thus, a simplified sample size formula is given by

$$n = \left(Z_{1-\alpha/2} + Z_\pi \right)^2 \hat{\sigma}_{II}^2(t) / \{\varepsilon \hat{p}_1 \hat{p}_2 \hat{R}(t) / (\hat{p}_1 + \hat{p}_2)\}^2, \quad (5.3)$$

where $\hat{\sigma}_{II}^2(t)$ is given by (4.5).

6. Simulation

In this section we present our simulation study aimed to investigate the finite sample properties of our proposed procedures. The simulations were carried out under the popular Gamma frailty model with mean 1 and variance θ . We consider three levels of dependence: independence ($\theta = 0.01005$), mild dependence ($\theta = 1$) and strong dependence ($\theta = 4$). These values of the frailty parameters were defined based on the Kendall's τ coefficient (Kendall, 1938). Under the Gamma frailty distribution Kendall's τ equals $\theta/(\theta + 2)$. Therefore, the respective values of Kendall's τ for the above values of θ are: 1/200, 1/3 and 2/3. We assume that each cluster is of size 2 and we consider four scenarios: (I) Constant baseline hazards, $\lambda_{01}(s) = \lambda_{02}(s) = 1$, and the covariates Z_{i1} and Z_{i2} are independent and each was generated from $Unif\{1, 2, 3\}$. (II) Constant baseline hazards as above and Z_{i1} was generated from $Unif\{1, 2, 3\}$ and Z_{i2} from $Bin(2, 0.25)$. (III) Weibull baseline hazard functions with $\lambda_{01}(s) = \lambda_{02}(s) = 2s$, and covariates as in Scenario (I). (IV) Weibull baseline hazard functions as in Scenario (III) and covariates as in scenario (II). For each covariate we generated two dummy variables. In all the above Scenarios, $\beta = (1, 2)^T$. Censoring times were generated from exponential distribution yielding 70% – 80% censoring rate, and for the test statistic we used $t = 0.1$ for Scenarios (I) and (II) and $t = 0.3$ for Scenarios (III)

and (IV). We consider $n = 250$ or 500 and the results are based on 1000 random samples for each configuration.

Tables 1-2 summarize the results of our proposed point estimators $\{\hat{\theta}, \hat{\beta}, \hat{\Lambda}_{01}, \hat{\Lambda}_{02}\}$ and present the true parameters' values, the empirical mean and standard deviation of the estimates. For the cumulative baseline hazard functions $\hat{\Lambda}_{0j}(t)$ $j = 1, 2$ we consider the values at $(t_1, t_2, t_3, t_4) = (0.005, 0.01, 0.05, 0.1)$. Tables 1-2 verify that our estimating procedure performs very well in terms of bias.

Table 3 compares the two variance estimators of our test statistic $S_n(t, \hat{\gamma})$, $\hat{\sigma}_I^2(t)$ and $\hat{\sigma}_{II}^2(t)$, by presenting the following descriptive statistics: the minimum, 1-st quartile, median, mean, 3-rd quartile and the maximum. It is evident that the differences between the two estimators are very small even under a strong dependency such as $\theta = 4$. These results support our recommendation to use the simplified estimator $\hat{\sigma}_{II}^2(t)$ rather than $\hat{\sigma}_I^2(t)$.

Now, we are comparing between our proposed variance estimator of $S_n(t, \hat{\gamma})$ and other naive variance estimators. One is an estimator that does not take into account the dependence between the samples. We denote this estimator by $\hat{\sigma}_1^2(t)$ and it is easy to verify that

$$\hat{\sigma}_1^2(t) = \frac{1}{n} \int_0^t \hat{W}_n(s) \sum_{j=1}^2 \sum_{i=1}^n \left\{ \frac{\bar{Y}_{3-j}(s, \hat{\gamma})}{\bar{Y}_j(s, \hat{\gamma})} \right\}^2 d\bar{N}_j(s).$$

The second estimator is the robust estimator of Song et al. (2008) and is given by

$$\hat{\sigma}_2^2(t) = \frac{1}{n} \sum_{j=1}^2 \sum_{i=1}^n \left\{ \int_0^t \hat{W}_n(s) \mathcal{G}(s, \hat{\gamma}) d\hat{M}_{ij}(s) \right\}^2,$$

where $\hat{M}_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s) e^{\hat{\beta}^T Z_{ij}} d\bar{N}_j(s) / \bar{Y}_j(s, \hat{\gamma})$ $i = 1, \dots, n$, $j = 1, 2$. This estimator was proposed for repeated events where the two baseline hazard functions were estimated based on independent samples. In Table 4 we present the mean of each variance estimator and the empirical significance level of a test with Type I error $\alpha = 0.05$. The empirical significance level is the percent of tests such that the null was rejected. The results show that under the independent setting the four methods provide similar results, but as the dependence increases, the differences between our methods and the two other naive methods, tend to increase as well. The empirical significance level for the other estimators increases with θ . It is evident, that only our methods perform reasonably well under any dependency level. In some cases, such as Scenario II with the sample size of 500, the empirical Type I error of the naive is about 11%. In addition, there are small differences between the empirical Type I error provided by our two proposed estimators $\hat{\sigma}_I^2$ and $\hat{\sigma}_{II}^2$. Hence our recommendation of using (4.5) for the variance estimate of $S_n(t, \hat{\gamma})$, is again being justified.

Now, we provide simulation results to evaluate the proposed sample size formula. All the three levels of dependence were examined under a two-sided test, with $\alpha = 0.05$, and $\pi = 0.80$. The baseline hazard functions corresponding to the local alternatives of the form $\lambda_{01}(s) = \exp\{\varepsilon/2\sqrt{n}\}\lambda_0(s)$ and $\lambda_{02}(s) = \exp\{-\varepsilon/2\sqrt{n}\}\lambda_0(s)$, where $\lambda_0(s) = 1$ and ε takes the values of 0.3, 0.5 or 0.6.

We first generated 100 random samples for each configuration, and based on these simulated data we calculated the average sample size based on (5.3). These results serve

as the required sample size with $\alpha = 0.05$, and $\pi = 0.80$. Then, for each configuration we generated 1000 random samples with the respective sample size. For each sample we calculated the test statistic $S_n(t, \hat{\gamma})$ and its variance estimate $\hat{\sigma}_{II}^2(t)$. Finally, we calculated the empirical power based on a two-sided test with $\alpha = 0.05$, to be compared with the nominal power of 0.80. The results are presented in Table 5. It is evident that our sample size formula performs well since the empirical power is reasonably close to the nominal power 0.80. The results demonstrate that as the difference between the two baseline hazard functions increases, less observations are required. The sample size formula of Schoenfeld (1983), $(Z_{1-\alpha/2} + Z_\pi)^2 / (2\varepsilon^2)$, presented in the last column of Table 5, gives similar values to that of our formula in the case of independence ($\theta = 0.01005$), as expected. In all other cases, Schoenfeld's formula underestimate the required sample size.

7. Data analysis

7.1. Call center data of a financial company

The data we analyze here are provided by a call center belonging to an Israeli financial company. The data cover a period of almost three years, October 2006 - June 2009. In this call center a customer requesting service from an agent is redirected to a pool of agents. If all the agents are busy, the customer waits in a queue. Otherwise, s/he is served immediately. The customer is not always ready to wait in a queue, and s/he can choose to abandon the system at any point during the waiting period. After the abandonment the customer may make an additional call. Customers who have been served may also call again to get an additional service or in continuation of the previously requested service. In our context, an event is the customer abandonment of the system before being served. The waiting time of a customer in a call ended after being served is considered as a censoring time.

The data do not contain any personal information about customers, such as age, social status or education. Therefore, our analysis will be carried out only on the basis of the technical characteristics of the call. For the analysis of customer behavior, we use a notion of a "series" defined as a sequence of consecutive calls of one customer happening in chronological order. If the time elapsed between two consecutive calls is less than three days we assume that these calls belong to the same "series". Otherwise, we assume that these calls belong to two different series. This separation is based on the assumption that a customer who has not called for a long time loses his/her experience with the system.

The following analysis consists of 49,246 customers, with only one series of calls for each customer, and each customer had not called for at least two months before the beginning of the series. By this we hope to ensure that customers are not familiar with the current system at their first call. For each customer we consider up to 5 calls. Table 6 presents the distribution of the observed calls. The covariate considered is a type of customer: VIP, of medium importance or a standard customer. This information is available to the agents only and it could effect the service provided to the customer (such as priority in the system). Hence, β_1 reflects the effect of VIP vs all others, and β_2 reflects the effect of medium importance vs others.

Table 7 presents the parameter estimates under the Gamma frailty model along with their bootstrap standard errors, based on 150 customer-level bootstrap samples. The results

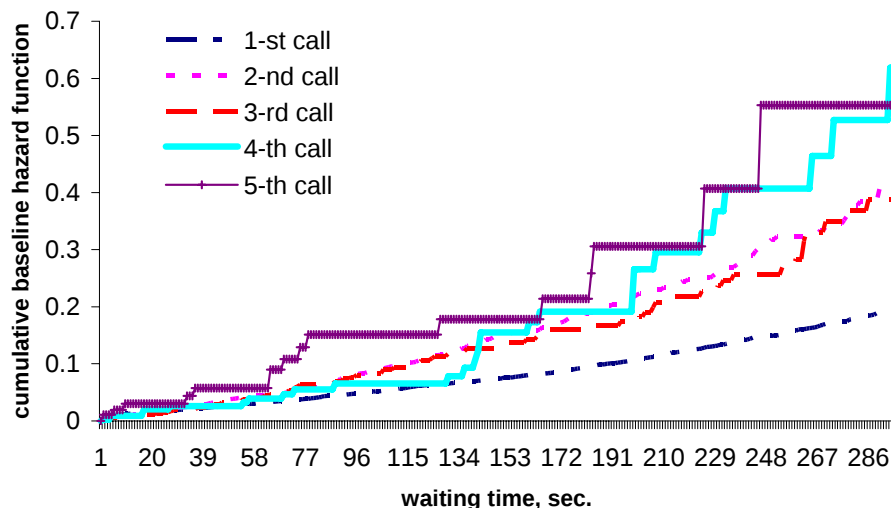


Figure 7.1: Estimates of the cumulative baseline hazard functions for the call center data.

show that the frailty parameter is close to 1 (Kendall's τ is approximately $1/3$), meaning moderate dependence between calls of the same customer. The estimates of the regression coefficients indicate that the chance to abandon before being served increases as the level of importance decreases. In Table 8 we present the estimated values of the baseline hazard functions calculated at times: 10, 50, 100, 150, 200 and 250 seconds. A graphical presentation is given in Figure 7.1. It is evident that the estimates of the first call are smaller than that of the other calls, and the estimates of the fifth call are larger than the other calls. For the other functions one could say that the differences are not so obvious. Now we would like to answer the following question: “Are the estimated baseline hazard functions really different, or are all these functions merely different estimates of the same function?”. To this end we apply our proposed test for comparing between each two cumulative baseline hazard functions. In Table 9 we present the values of the test statistic $S_n(250, \hat{\gamma})$, the estimated standard error based on (4.5), the standardized test statistic, the p -value based on the standard normal distribution and the corrected p -value based on the FDR method (Benjamini and Yekutieli, 2001) for correcting the dependent comparisons. The results show us that the baseline hazard function of the first call is significantly different from that of all other calls, even after correcting for multiple comparisons. There is also a significant difference between the baseline hazard functions of the third and the fifth calls. Differences between all the other functions are not statistically significant.

7.2. The Washington Ashkenazi Kin-Cohort breast cancer family data

In the Washington Ashkenazi Kin-Cohort Study (WAS) (Struwing et al., 1997), blood samples and questionnaire were collected from Ashkenazi Jewish men and women volunteers living in the Washington DC area. Based on blood samples, volunteers were tested for

specific mutations in BRCA1 and BRCA2 genes. The questionnaire included information on cancer and mortality history of the first-degree relatives of the volunteers.

For the current analysis we consider a subset of the data consist of female first-degree relatives of volunteers (mother, sisters and daughters). The event is the age at breast cancer diagnosis, and the covariate is the presence or absence of any BRCA1/2 mutations in the volunteer's blood sample. The data consist of 4,835 families with 1 – 8 relatives and a total of 13,030 subjects.

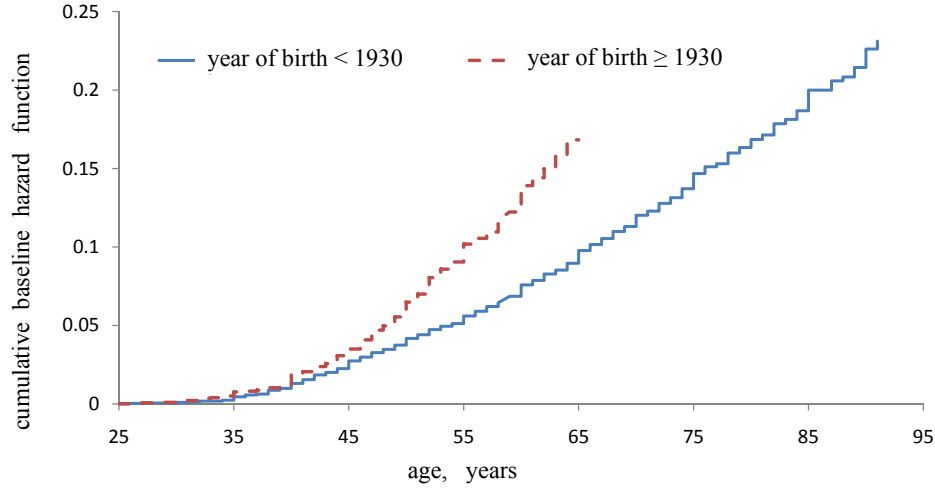


Figure 7.2: Estimates of the cumulative baseline hazard functions for the WAS data by birth year.

So far, these data were analyzed under the assumption that the baseline hazard functions are identical among all family members. However, we would like to allow for each generation to have its own baseline hazard function, where the relative's generation is defined based on her year of birth: before 1930 or otherwise. For this, we can rewrite model (2.1) as follows:

$$\lambda_{ij}(t) = \sum_{k=1}^2 \lambda_{0k}(t) w_i e^{\beta Z_{ij}} I(\tilde{Z}_{ij} = k) \quad i = 1, \dots, n \quad j = 1, \dots, m_i,$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function of generation $k = 1, 2$, Z_{ij} is an indicator for the presence of any BRCA1/2 mutations in the blood sample of the volunteer of family i , and $\tilde{Z}_{ij}$ takes the value of 1 if member j of family i was born before 1930, and value 2 otherwise. See Table 10 for sample size and number of events at each stratum.

We start with reporting on the point estimates and their bootstrap standard errors, based on 100 bootstrap samples. The estimated frailty parameter under the gamma frailty model equals $\hat{\theta} = 1.862$ (SE=0.210), the estimated regression coefficient equals $\hat{\beta} = 1.396$ (SE=0.170) and the estimates of the cumulative baseline hazard functions are presented in Figure 7.2. The estimated parameter of the frailty distribution indicates high dependence among family members (Kendall's τ is about 0.5). Also, it is evident that the cumulative

baseline hazard function of the older generation is always lower than that of the younger generation. Our test for comparing the two hazards provided $S_n(60, \hat{\gamma})/\hat{\sigma}_{II}(60) = 18.903$ ($\hat{\sigma}_{II}(60) = 0.058$) and $p\text{-value} < 0.001$. Thus, we verify our visual inspection and observe a significant difference between the two functions. Such a finding supports other publications reporting that cancer rates have risen in the last years (Center et al., 2009), and it is likely that such a tendency is mainly as a result of the increase usage of screening programs which detect breast cancer in earlier stages (Paltiel et al., 2007 among others). We also show that the presence of specific mutations in BRCA1/2 genes of one family member, significantly increases the risk of breast cancer for other family members.

8. Summary

We provided a test for comparing two or more cumulative baseline hazard functions for correlated survival data. The asymptotic distribution of the proposed test statistic was presented along with simulation study results. The results show that our proposed method works well and as expected, gives better results in compare to the naive approaches that ignores intra-cluster dependence. A sample-size formula was derived based on the limiting distribution of our test statistic under local alternatives. Our simulation study shows that under the proposed formula, the empirical power is reasonably close to the nominal value. A sample-size formula for testing equality of m hazard functions can be derived based on our proposed test statistic $\mathbf{S}_n^o(t, \hat{\gamma})^T [\hat{\mathbf{V}}^o(t)]^{-1} \mathbf{S}_n^o(t, \hat{\gamma})$ analogously to our analysis in Section 5 and that of Ahnn and Anderson (1995).

It should also be noted that the above theory for testing the null hypothesis of equality of baseline hazard functions can also be adopted for testing contrasts defined on the baseline hazard functions.

For simplicity we assumed $\beta_j = \beta$ $j = 1, \dots, m$. However, the estimation procedure and the proposed test statistic, along with its asymptotic distribution, can be trivially extended to the case of different regression coefficients.

Acknowledgment

The authors are grateful to Professor David Zucker, Dr. Valery Trofimov and Professor Avishai Mandelbaum for their helpful comments. Dr. Khudyakov would like to thank the Service Enterprise Engineering (SEE) Center, at the Technion, Israel, for the financial support and the cooperation in providing the data used in this work. The authors also thank Nilanjan Chatterjee for facilitating the access of the Washington Ashkenazi Jewish data.

References

- Acar, E. F., Craiu, R. V. and Yao, F. (2010). Dependence Calibration in Conditional Copulas: A Nonparametric Approach. *Biometrics*, **67**, 445-453.
- Ahnn, S. and Anderson, S. J. (1995). Sample Size Determination for Comparing More Than Two Survival Distributions. *Statistics in Medicine*, **14**, 2273-2282.

- Andersen, P. K. and Gill, D. R. (1982). Cox's Regression Model for Counting Processes: a Large Sample Study. *Ann. Statist.*, **10**, 1100-1120.
- Andersen, P. K., Borgan, Ø., Gill, D. R. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.
- Benjamini, Y. and Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing Under Dependency. *The Annals of Statistics*, **29**, 1165-1188.
- Breslow, N. (1974). Covariance Analysis of Censored Survival Data. *Biometrics*, **53**, 1475-1484.
- Center, M., Jemal, A. and Ward, E. (2009). International Trends in Colorectal Cancer Incidence Rates. *Cancer Epidemiol. Biomarker Prev.*, **18**, 1688-1694.
- Cook, R. J., Lawless, J. F. and Nadeau, C. (1996). Robust Tests for Treatment Comparisons Based on Recurrent Event Responses. *Biometrics*, **52**, 557-571.
- Cox, D. R. (1972). Regression Models and Life Tables (with discussion). *J. Roy. Statist. Soc.*, **34**, 187-220.
- Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. Springer, New York, NY.
- Eng, K. H. and Kosorok, M. R. (2005). Sample size formula for the supremum log rank statistic. *Biometrics*, **61**, 86-91.
- Fleming, T. R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
- Gangnon, R. E. and Kosorok, M. R. (2004). Sample-size formula for clustered survival data using weighted log-rank statistics. *Biometrika*, **91**, 2, 263-275.
- Gill, R. D. (1980). Censoring and Stochastic Integrals. Tract 124, Amsterdam: The Mathematical Center.
- Gorfine, M., Zucker, D. M. and Hsu, L. (2006). Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach. *Biometrika*, **93**, 3, 735-741.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag New York, NY.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The statistical analysis of failure time data*. Wiley, New York, NY.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81-93.
- Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, **48**, 795-806.

- Kosorok, M. R. and Lin, C. Y. (1999). The Versality of Function-Indexed Weighted Log-Rank Statistics. *Journal of the American Statistical Association* **94**, 320-332.
- Lawless, J. F. and Nadeau, C. (1995). Some Simple Robust Methods for the Analysis of Recurrent Events. *Technometrics*, **37**, 158-168.
- Murphy, S. A. (1994). Consistency in a Proportional Hazards Model Incorporating a Random Effect. *Ann. Statist.*, **23**, 182-98.
- Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sorensen, T. I. (1992). A counting process approach to maximum likelihood estimation of frailty models. *Scand. J. Statist.*, **19**, 25-43.
- Paltiel, O., Friedlander, Y., Deutsch, L., Yanetz, R., Calderon-Margalit, R., Tiram, E., Hochner, H., Barchana, M., Harlap, S. and Manor, O. (2007). The interval between cancer diagnosis among mothers and offspring in a population-based cohort. *Familial Cancer*, **6**, 121-129.
- Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.*, **26**, 183-214.
- Peto, R. and Peto, J. (1972). Asymptotically Efficient Rank Invariant Test Procedures. *J. Roy. Statist. Soc.*, **135**, 185-207.
- Prentice, R.L. (1978). Linear rank tests with right censored data. *Biometrika*, **65**, 167-179.
- Ripatti, S. and Palmgren, J. (1978). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **65**, 153-158.
- Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics* **39**, 499-503.
- Song, R., Kosorok, M. R. and Cai, J. (2008). Robust Covariate-Adjusted Log-Rank Statistics and Corresponding Sample Size Formula for Recurrent Events Data. *Biometrics*, **64**, 741-750.
- Struwing, J. P., Hartge, P., Wacholder, S., Baker, S. M., Berlin, M., McAdams, M., Timmerman, M. M., Brody, L. C. and Tucker, M.A. (1997). "The Risk of Cancer Associated with Specific Mutations of BRCA1 and BRCA2 among Ashkenazi Jews". *N Engl J Med*, **336**, 1401-1408.
- Zeng, D. and Lin, D. Y. (2007). Maximum Likelihood Estimation in Semiparametric Regression Models with Censored Data (with discussion). *J. Roy. Statist. Soc. Ser. B*, **69**, 507-564.
- Zucker, D. M., Gorfine M. and Hsu L. (2008). Pseudo-full likelihood estimation for prospective survival analysis with a general semiparametric shared frailty model: Asymptotic theory. *J. Statist. Plann. Inference*, **138**, 1998-2016.

Department of Biostatistics, Harvard School of Public Health, Boston, MA, U.S.A.

E-mail: (stpok@channing.harvard.edu)

Faculty of Industrial Engineering and Management, Technion, Technion City, Haifa 32000, Israel

E-mail: (gorfinm@ie.technion.ac.il, paulf@ie.technion.ac.il)

1 Appendix

The following are the main steps of the proofs and asymptotic variance calculations. For more details, the reader is referred to the Ph.D. thesis of Dr. Khudyakov available at <http://iew3.technion.ac.il/serveng/References/references.html>.

1.1 Proof of Theorem 4.1

Let

$$A_n(t, \hat{\gamma}) = \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \left\{ \frac{\bar{Y}_2(s, \hat{\gamma})}{\bar{Y}_\cdot(s, \hat{\gamma})} dM_1(s) - \frac{\bar{Y}_1(s, \hat{\gamma})}{\bar{Y}_\cdot(s, \hat{\gamma})} dM_2(s) \right\}$$

and write $S_n(t, \hat{\gamma}) = A_n(t, \hat{\gamma}) + S_n^{**}(t)$. The first order Taylor expansion of $A_n(t, \hat{\gamma})$ about γ^o gives

$$A_n(t, \hat{\gamma}) \approx \tilde{S}_n(t, \gamma^o) + \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \{ \mathbf{Q}_2^T(s, \gamma^o) dM_1(s) - \mathbf{Q}_1^T(s, \gamma^o) dM_2(s) \} (\hat{\gamma} - \gamma^o). \quad (1)$$

Given Assumptions A.9-A.10, it is easy to show that the conditional distribution of

$$B_n(t, \gamma^o) = \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \{ \mathbf{Q}_2^T(s, \gamma^o) dM_1(s) - \mathbf{Q}_1^T(s, \gamma^o) dM_2(s) \},$$

conditioning on w ., convergence to a zero-mean multivariate normally distributed random variable with finite entries of the covariance matrix that are free of the frailties. Hence, this is also the unconditional asymptotic distribution of $B_n(t, \gamma^o)$. Then, given Claim 3.3, the second term of (1) goes to zero as $n \rightarrow \infty$, by Slutsky's theorem.

1.2 Proof of Theorem 4.2

Statement (1): Given w ., treat the frailties as additional covariates. Hence, standard martingale arguments can be used to show that $\tilde{S}_n(t, \gamma^o)$ converges to a zero-mean normally distributed random variable with variance $\sigma_{\tilde{S}}^2(t)$ that is free of the frailties w .. Therefore, $\tilde{S}_n(t, \gamma^o)$ also unconditionally converges to a normally distributed random variable with the same parameters.

Statement (2): Note that $S_n^*(t, \gamma^o)$ can be rewritten in the following form

$$S_n^*(t, \gamma^o) = \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \left[\frac{\bar{Y}_2(s, \gamma^o)}{\bar{Y}_\cdot(s, \gamma^o)} \{ d\bar{M}_1(s) - d\bar{M}_1^*(s) \} - \frac{\bar{Y}_1(s, \gamma^o)}{\bar{Y}_\cdot(s, \gamma^o)} \{ d\bar{M}_2(s) - d\bar{M}_2^*(s) \} \right],$$

where $\bar{M}_j^* = \sum_{i=1}^n M_{ij}^*(s)$ is a mean-zero martingale of the process $\bar{N}_j(s)$ with respect to the stochastic intensity process (3.1). Then, by the martingale central limit theorem, we obtain that $S_n^*(t, \gamma^o)$ is asymptotically normally distributed with mean zero.

Let $X_j^*(s) = n^{-1/2} \sum_{i=1}^n Y_{ij}(s) e^{\beta^T Z_{ij}} \{w_i - \psi_i\}$,

$$g_j(s) = \hat{W}_n(s) \lambda_0(s) \frac{\bar{Y}_j(s, \gamma^o)}{\bar{Y}(s, \gamma^o)} \quad \text{and} \quad G_j(t) = \int_0^t g_j(s) X_j^*(s) ds$$

Then, $\text{Var}\{S_n^*(t, \gamma^o)\} = \text{Var}\{G_1(t)\} + \text{Var}\{G_2(t)\} - 2\text{Cov}\{G_1(t), G_2(t)\}$. Since $X_j^*(s)$ $j = 1, 2$ are with mean zero

$$\text{Var}\{G_j(t)\} = \frac{1}{n} E \left\{ \int_0^t \int_0^t g_j(s) g_j(u) \sum_{i=1}^n Y_{ij}(s \vee u) e^{2\beta^T Z_{ij}} \text{Var}(w_i \mid \mathcal{F}_{s \vee u-}) \right\} ds du,$$

and by denoting $Z_{i\cdot} = Z_{i1} + Z_{i2}$ we get

$$\text{Cov}\{G_1(t), G_2(t)\} = \frac{1}{n} E \left\{ \int_0^t \int_0^t g_1(s) g_2(u) \sum_{i=1}^n Y_{i1}(s) Y_{i2}(u) e^{\beta^T Z_{i\cdot}} \text{Var}(w_i \mid \mathcal{F}_{s \vee u-}) \right\} ds du.$$

$$\begin{aligned} \text{Var}\{S_n^*(t, \gamma^o)\} &= \frac{1}{n} \left(\sum_{j=1}^m E \left\{ \int_0^t \int_0^t g_j(s) g_j(u) \sum_{i=1}^n Y_{ij}(s \vee u) e^{2\beta^T Z_{ij}} \text{Var}(w_i \mid \mathcal{F}_{s \vee u-}) \right\} ds du \right. \\ &\quad \left. - 2E \left\{ \int_0^t \int_0^t g_1(s) g_2(u) \sum_{i=1}^n Y_{i1}(s) Y_{i2}(u) e^{\beta^T Z_{i\cdot}} \text{Var}(w_i \mid \mathcal{F}_{s \vee u-}) \right\} ds du \right). \end{aligned} \quad (2)$$

Hence, it is easy to verify that $\text{Var}\{S_n^*(t, \gamma^o)\} < \infty$.

Statement (3): Note that under the null hypothesis

$$\begin{aligned} \text{Cov}(\tilde{S}_n(t, \gamma^o), S_n^*(t, \gamma^o)) &= \int_0^t \int_0^t E \left[D_n(s, \gamma^o) D_n(u, \gamma^o) \left\{ \frac{d\bar{M}_1(s)}{\bar{Y}_1(s, \gamma^o)} \right. \right. \\ &\quad \left. \left. - \frac{d\bar{M}_2(s)}{\bar{Y}_2(s, \gamma^o)} \right\} \left\{ \frac{\tilde{Y}_1(u, \gamma^o)}{\bar{Y}_1(u, \gamma^o)} - \frac{\tilde{Y}_2(u, \gamma^o)}{\bar{Y}_2(u, \gamma^o)} \right\} d\Lambda_0(u) \right]. \end{aligned}$$

For $s \geq u$ we get

$$\begin{aligned} &E \left[E \left(D_n(s, \gamma^o) D_n(u, \gamma^o) \left\{ \frac{d\bar{M}_1(s)}{\bar{Y}_1(s, \gamma^o)} - \frac{d\bar{M}_2(s)}{\bar{Y}_2(s, \gamma^o)} \right\} \left\{ \frac{\tilde{Y}_1(u, \gamma^o)}{\bar{Y}_1(u, \gamma^o)} - \frac{\tilde{Y}_2(u, \gamma^o)}{\bar{Y}_2(u, \gamma^o)} \right\} d\Lambda_0(u) \mid \mathcal{F}_{s-} \right) \right] \\ &= E \left[D_n(s, \gamma^o) D_n(u, \gamma^o) \left\{ \frac{\tilde{Y}_1(u, \gamma^o)}{\bar{Y}_1(u, \gamma^o)} - \frac{\tilde{Y}_2(u, \gamma^o)}{\bar{Y}_2(u, \gamma^o)} \right\} E \left(\left\{ \frac{d\bar{M}_1(s)}{\bar{Y}_1(s, \gamma^o)} - \frac{d\bar{M}_2(s)}{\bar{Y}_2(s, \gamma^o)} \right\} \mid \mathcal{F}_{s-} \right) d\Lambda_0(u) \right] = 0, \end{aligned}$$

and for $u > s$, we get

$$E \left[D_n(s, \gamma^o) D_n(u, \gamma^o) \left\{ \frac{d\bar{M}_1(s)}{\bar{Y}_1(s, \gamma^o)} - \frac{d\bar{M}_2(s)}{\bar{Y}_2(s, \gamma^o)} \right\} E \left(\left\{ \frac{\tilde{Y}_1(u, \gamma^o)}{\bar{Y}_1(u, \gamma^o)} - \frac{\tilde{Y}_2(u, \gamma^o)}{\bar{Y}_2(u, \gamma^o)} \right\} \mid \mathcal{F}_{u-} \right) d\Lambda_0(u) \right] = 0.$$

1.3 An estimator of the variance of $S_n(t, \hat{\gamma})$

In the following, we generate our variance estimators of $Var\{\tilde{S}_n(t, \gamma^o)\}$ and $Var\{S_n^*(t, \gamma^o)\}$. Let us start with an estimator of $Var\{\tilde{S}_n(t, \gamma^o)\}$. By using the law of total variance we get $Var\{\tilde{S}_n(t, \gamma^o)\} = E\left[Var\{\tilde{S}_n(t, \gamma^o) \mid w.\}\right] + Var\left[E\{\tilde{S}_n(t, \gamma^o) \mid w.\}\right]$ and under the null hypothesis, the second term equals 0. Since calls of customer i are conditionally independent given w_i , the predictable variation process of $\tilde{S}_n(t, \gamma^o)$, given $w.$, is given by

$$\langle \tilde{S}_n \mid w. \rangle (t, \gamma^o) = \int_0^t D_n^2(s, \gamma^o) \left(Var\left\{ \frac{d\bar{M}_1(s)}{\bar{Y}_1(s, \gamma^o)} \mid w., \mathcal{F}_{s-} \right\} + Var\left\{ \frac{d\bar{M}_2(s)}{\bar{Y}_2(s, \gamma^o)} \mid w., \mathcal{F}_{s-} \right\} \right).$$

Since $Var\{dM_{ij}(s) \mid w., \mathcal{F}_{s-}\} = Y_{ij}(s)e^{\beta^T Z_{ij}} w_i \lambda_{0j}(s) ds$ we get

$$\langle \tilde{S}_n \mid w. \rangle (t, \gamma^o) = \sum_{i=1}^n w_i \int_0^t D_n^2(s, \gamma^o) \left\{ \frac{e^{\beta^T Z_{i1}} Y_{i1}(s)}{\bar{Y}_1^2(s, \gamma^o)} d\Lambda_{01}(s) + \frac{e^{\beta^T Z_{i2}} Y_{i2}(s)}{\bar{Y}_2^2(s, \gamma^o)} d\Lambda_{02}(s) \right\}.$$

Then, the expectation with respect to the unknown frailties gives

$$\sum_{i=1}^n E(w_i) \int_0^t D_n^2(s, \gamma^o) \left\{ \frac{e^{\beta^T Z_{i1}} Y_{i1}(s)}{\bar{Y}_1^2(s, \gamma^o)} d\Lambda_{01}(s) + \frac{e^{\beta^T Z_{i2}} Y_{i2}(s)}{\bar{Y}_2^2(s, \gamma^o)} d\Lambda_{02}(s) \right\}. \quad (3)$$

The variance of $Var\{S_n^*(t, \gamma^o)\}$ is presented in (2). Then, we replace all the unknown parameters in (2) and (3) by their estimates and get the estimators as presented in Section 4.2.

1.4 Proof of Theorem 4.3

Statement (2): Since

$$E\left\{ \frac{\tilde{Y}_j(s, \gamma^o)}{\bar{Y}_j(s, \gamma^o)} - \frac{\tilde{Y}_j(s, \gamma^o)}{\bar{Y}_j(s, \gamma^o)} \mid \mathcal{F}_{s-} \right\} = \frac{\bar{Y}_j(s, \gamma^o)}{\bar{Y}_j(s, \gamma^o)} - \frac{\bar{Y}_j(s, \gamma^o)}{\bar{Y}_j(s, \gamma^o)} = 0 \quad j = 1, \dots, m$$

it is easy to show that under the null hypothesis $E[S_{nj}^*(t, \gamma^o)] = 0$, for $j = 1, \dots, m$. Also, using again the law of total expectation by conditioning on $\mathcal{F}_{u \vee s-}$ we obtain that under the null hypothesis $Cov[S_{nj}^*(t, \gamma^o), S_{nk}^*(t, \gamma^o)] = 0$. Using similar arguments as in the proof of Theorem 4.2, one can show that each $S_{nj}(t, \gamma^o)$ is asymptotically normally distributed with a finite variance.

Statement (3): Note that under the null hypothesis, the covariance between $\tilde{S}_{nj}(t, \gamma^o)$ and $S_{nk}^*(t, \gamma^o)$ for all $j, k = 1, \dots, m$ can be written as

$$\begin{aligned} Cov\left\{ \tilde{S}_{nj}(t, \gamma^o), S_{nk}^*(t, \gamma^o) \right\} &= \int_0^t \int_0^t E\left[\frac{\hat{W}_n(s)}{\sqrt{n}} \bar{Y}_j(s, \gamma^o) \frac{\hat{W}_n(u)}{\sqrt{n}} \bar{Y}_k(u, \gamma^o) \left\{ \frac{d\bar{M}_j(s)}{\bar{Y}_j(s, \gamma^o)} \right. \right. \\ &\quad \left. \left. - \frac{d\bar{M}_j(s)}{\bar{Y}_j(s, \gamma^o)} \right\} \left\{ \frac{\tilde{Y}_k(u, \gamma^o)}{\bar{Y}_k(u, \gamma^o)} - \frac{\tilde{Y}_k(u, \gamma^o)}{\bar{Y}_k(u, \gamma^o)} \right\} \right] d\Lambda_0(u). \end{aligned} \quad (4)$$

Now, analogously to the proof of statement (3) of Theorem 4.3 one can show that

$$Cov\left\{ \tilde{S}_{nj}(t, \gamma^o), S_{nk}^*(t, \gamma^o) \right\} = 0.$$

1.5 The estimation of $V(t)$

The predictable variation process of $\tilde{S}_n(t, \gamma^o)$ given $w.$ equals

$$\langle \tilde{S}_{nj} \mid w. \rangle (t, \gamma^o) = V_{jj}^{(n)}(t) \quad \text{and} \quad \langle \tilde{S}_{nk}, \tilde{S}_{nj} \mid w. \rangle (t, \gamma^o) = V_{jk}^{(n)}(t)$$

for $j, k = 1, \dots, m$ since $\text{Var}\{d\bar{M}_j(s) \mid w., \mathcal{F}_{t-}\} = \tilde{Y}_j(s, \gamma)\lambda_{0j}(s)ds$. Then, by taking the expectation with respect to $w.$ and by replacing all the unknown parameters by their estimates we obtain the estimators (4.4) and (4.5).

1.6 Proof of Theorem 5.1

Write

$$\begin{aligned} \tilde{S}_n^*(t, \gamma^o) &= \frac{1}{\sqrt{n}} \int_0^t \hat{W}_n(s) \mathcal{G}(s, \gamma^o) \left\{ \frac{d\bar{M}_1(s)}{\bar{Y}_1(s, \gamma^o)} - \frac{d\bar{M}_2(s)}{\bar{Y}_2(s, \gamma^o)} \right\} \\ &\quad + \frac{1}{n} \int_0^t \hat{W}_n(s) \mathcal{G}(s, \gamma^o) \frac{\tilde{Y}_2(s, \gamma^o)}{\bar{Y}_2(s, \gamma^o)} d\Lambda_{02}^n(s) \sqrt{n} \left\{ \frac{\tilde{Y}_1(s, \gamma^o) \bar{Y}_2(s, \gamma^o) d\Lambda_{01}^n(s)}{\bar{Y}_2(s, \gamma^o) \bar{Y}_1(s, \gamma^o) d\Lambda_{02}^n(s)} - 1 \right\}. \end{aligned} \quad (5)$$

The first term of the right-hand side of (5) converges to a normal random variable with mean zero and the proof is similar to that of Theorem 4.2. By Assumptions A.8-A.10 we obtain that

$$\sup_{s \in [0, \tau]} \left| \sqrt{n} \left\{ \frac{\tilde{Y}_1(s, \gamma^o) \bar{Y}_2(s, \gamma^o) d\Lambda_{01}^n(s)}{\bar{Y}_2(s, \gamma^o) \bar{Y}_1(s, \gamma^o) d\Lambda_{02}^n(s)} - 1 \right\} - \varphi(s) \right| \rightarrow 0.$$

Therefore, the second term of the right-hand side of (5) converges to $\mu_1(t)$ in probability, as $n \rightarrow \infty$.

Table 1: *Summary of parameter estimates $\{\hat{\theta}, \hat{\beta}, \hat{\Lambda}_{0j}\}$ under exponential hazards.*

	independence			mild dependence			strong dependence		
	true value	mean	SD	true value	mean	SD	true value	mean	SD
<i>exponential (I)</i>									
<i>n=250</i>									
θ	0.01005	0.084	0.150	1	0.929	0.240	4	3.806	0.812
β_1	1	1.023	0.328	1	0.969	0.249	1	1.009	0.355
β_2	2	2.059	0.318	2	1.935	0.264	2	1.986	0.368
$\Lambda_{01}(t_1)$	0.005	0.005	0.003	0.005	0.005	0.003	0.005	0.005	0.003
$\Lambda_{01}(t_2)$	0.01	0.010	0.004	0.01	0.011	0.004	0.01	0.010	0.005
$\Lambda_{01}(t_3)$	0.05	0.049	0.016	0.05	0.053	0.015	0.05	0.050	0.018
$\Lambda_{01}(t_4)$	0.1	0.096	0.031	0.1	0.104	0.029	0.1	0.098	0.034
$\Lambda_{01}(t_1)$	0.005	0.005	0.003	0.005	0.006	0.003	0.005	0.005	0.003
$\Lambda_{01}(t_2)$	0.01	0.010	0.005	0.01	0.011	0.005	0.01	0.010	0.005
$\Lambda_{01}(t_3)$	0.05	0.049	0.017	0.05	0.054	0.015	0.05	0.051	0.017
$\Lambda_{01}(t_4)$	0.1	0.097	0.033	0.1	0.108	0.029	0.1	0.101	0.031
<i>n=500</i>									
θ	0.01005	0.064	0.098	1	1.025	0.175	4	3.925	0.596
β_1	1	1.013	0.219	1	1.008	0.198	1	1.007	0.242
β_2	2	2.021	0.211	2	2.003	0.201	2	1.999	0.262
$\Lambda_{01}(t_1)$	0.005	0.005	0.002	0.005	0.005	0.002	0.005	0.005	0.002
$\Lambda_{01}(t_2)$	0.01	0.010	0.003	0.01	0.010	0.003	0.01	0.010	0.004
$\Lambda_{01}(t_3)$	0.05	0.049	0.011	0.05	0.050	0.010	0.05	0.050	0.013
$\Lambda_{01}(t_4)$	0.1	0.098	0.021	0.1	0.101	0.019	0.1	0.100	0.025
$\Lambda_{01}(t_1)$	0.005	0.005	0.002	0.005	0.005	0.002	0.005	0.005	0.002
$\Lambda_{01}(t_2)$	0.01	0.010	0.003	0.01	0.010	0.003	0.01	0.010	0.004
$\Lambda_{01}(t_3)$	0.05	0.049	0.011	0.05	0.050	0.010	0.05	0.051	0.014
$\Lambda_{01}(t_4)$	0.1	0.098	0.021	0.1	0.101	0.019	0.1	0.100	0.025
<i>exponential (II)</i>									
<i>n=250</i>									
θ	0.01005	0.049	0.132	1	0.953	0.336	4	3.824	0.784
β_1	1	1.078	0.208	1	1.011	0.265	1	0.994	0.304
β_2	2	2.228	0.251	2	2.008	0.317	2	1.989	0.400
$\Lambda_{01}(t_1)$	0.005	0.004	0.002	0.005	0.005	0.003	0.005	0.005	0.003
$\Lambda_{01}(t_2)$	0.01	0.008	0.003	0.01	0.010	0.005	0.01	0.010	0.005
$\Lambda_{01}(t_3)$	0.05	0.046	0.012	0.05	0.050	0.018	0.05	0.050	0.017
$\Lambda_{01}(t_4)$	0.1	0.097	0.022	0.1	0.100	0.033	0.1	0.099	0.032
$\Lambda_{01}(t_1)$	0.005	0.005	0.003	0.005	0.005	0.003	0.005	0.005	0.003
$\Lambda_{01}(t_2)$	0.01	0.009	0.005	0.01	0.010	0.005	0.01	0.010	0.005
$\Lambda_{01}(t_3)$	0.05	0.048	0.013	0.05	0.048	0.017	0.05	0.050	0.017
$\Lambda_{01}(t_4)$	0.1	0.099	0.023	0.1	0.096	0.031	0.1	0.099	0.029
<i>n=500</i>									
θ	0.01005	0.042	0.076	1	1.020	0.214	4	3.888	0.562
β_1	1	1.011	0.148	1	1.012	0.194	1	0.984	0.214
β_2	2	2.018	0.175	2	2.009	0.233	2	1.964	0.284
$\Lambda_{01}(t_1)$	0.005	0.005	0.002	0.005	0.005	0.002	0.005	0.005	0.002
$\Lambda_{01}(t_2)$	0.01	0.010	0.003	0.01	0.010	0.003	0.01	0.011	0.004
$\Lambda_{01}(t_3)$	0.05	0.049	0.009	0.05	0.050	0.011	0.05	0.051	0.013
$\Lambda_{01}(t_4)$	0.1	0.097	0.016	0.1	0.101	0.019	0.1	0.102	0.023
$\Lambda_{01}(t_1)$	0.005	0.005	0.002	0.005	0.005	0.002	0.005	0.005	0.003
$\Lambda_{01}(t_2)$	0.01	0.010	0.003	0.01	0.010	0.004	0.01	0.010	0.004
$\Lambda_{01}(t_3)$	0.05	0.049	0.009	0.05	0.050	0.010	0.05	0.051	0.012
$\Lambda_{01}(t_4)$	0.1	0.097	0.015	0.1	0.1	0.018	0.1	0.101	0.021

Table 2: *Summary of parameter estimates $\{\hat{\theta}, \hat{\beta}, \hat{\Lambda}_{0j}\}$ under Weibull hazards.*

	independence			mild dependence			strong dependence		
	true value	mean	SD	true value	mean	SD	true value	mean	SD
<i>Weibull (I)</i>									
<i>n=250</i>									
θ	0.01005	0.035	0.081	1	0.892	0.377	4	3.695	0.680
β_1	1	1.027	0.278	1	0.995	0.351	1	0.973	0.307
β_2	2	2.054	0.274	2	1.985	0.360	2	1.92	0.327
$\Lambda_{01}(t_1)$	0.01	0.010	0.004	0.01	0.010	0.006	0.01	0.011	0.005
$\Lambda_{01}(t_2)$	0.04	0.038	0.012	0.04	0.039	0.018	0.04	0.042	0.015
$\Lambda_{01}(t_3)$	0.09	0.086	0.024	0.09	0.087	0.037	0.09	0.092	0.030
$\Lambda_{01}(t_4)$	0.16	0.152	0.042	0.16	0.154	0.066	0.16	0.161	0.052
$\Lambda_{01}(t_1)$	0.01	0.010	0.004	0.01	0.010	0.005	0.01	0.011	0.005
$\Lambda_{01}(t_2)$	0.04	0.039	0.012	0.04	0.041	0.018	0.04	0.042	0.015
$\Lambda_{01}(t_3)$	0.09	0.088	0.025	0.09	0.087	0.038	0.09	0.093	0.030
$\Lambda_{01}(t_4)$	0.16	0.154	0.044	0.16	0.153	0.066	0.16	0.163	0.051
<i>n=500</i>									
θ	0.01005	0.032	0.057	1	1.027	0.295	4	3.825	0.509
β_1	1	1.017	0.182	1	1.002	0.259	1	0.991	0.236
β_2	2	2.030	0.190	2	1.985	0.270	2	1.978	0.245
$\Lambda_{01}(t_1)$	0.01	0.010	0.003	0.01	0.010	0.004	0.01	0.010	0.003
$\Lambda_{01}(t_2)$	0.04	0.039	0.008	0.04	0.040	0.010	0.04	0.041	0.010
$\Lambda_{01}(t_3)$	0.09	0.087	0.017	0.09	0.090	0.022	0.09	0.091	0.022
$\Lambda_{01}(t_4)$	0.16	0.152	0.030	0.16	0.158	0.037	0.16	0.160	0.038
$\Lambda_{01}(t_1)$	0.01	0.010	0.003	0.01	0.010	0.003	0.01	0.010	0.003
$\Lambda_{01}(t_2)$	0.04	0.039	0.008	0.04	0.040	0.011	0.04	0.040	0.010
$\Lambda_{01}(t_3)$	0.09	0.087	0.018	0.09	0.090	0.022	0.09	0.090	0.021
$\Lambda_{01}(t_4)$	0.16	0.152	0.031	0.16	0.159	0.037	0.16	0.159	0.036
<i>Weibull (II)</i>									
<i>n=250</i>									
θ	0.01005	0.038	0.083	1	0.907	0.311	4	3.614	0.751
β_1	1	1.017	0.220	1	0.999	0.267	1	0.988	0.300
β_2	2	2.050	0.267	2	1.977	0.344	2	1.939	0.410
$\Lambda_{01}(t_1)$	0.01	0.010	0.004	0.01	0.010	0.005	0.01	0.011	0.005
$\Lambda_{01}(t_2)$	0.04	0.038	0.011	0.04	0.040	0.015	0.04	0.041	0.016
$\Lambda_{01}(t_3)$	0.09	0.086	0.022	0.09	0.089	0.031	0.09	0.090	0.032
$\Lambda_{01}(t_4)$	0.16	0.152	0.039	0.16	0.157	0.054	0.16	0.158	0.056
$\Lambda_{01}(t_1)$	0.01	0.010	0.005	0.01	0.010	0.006	0.01	0.010	0.006
$\Lambda_{01}(t_2)$	0.04	0.039	0.012	0.04	0.039	0.015	0.04	0.041	0.015
$\Lambda_{01}(t_3)$	0.09	0.088	0.023	0.09	0.087	0.030	0.09	0.090	0.029
$\Lambda_{01}(t_4)$	0.16	0.156	0.037	0.16	0.155	0.049	0.16	0.158	0.048
<i>n=500</i>									
θ	0.01005	0.031	0.056	1	0.956	0.216	4	3.790	0.580
β_1	1	1.005	0.163	1	0.988	0.185	1	0.994	0.222
β_2	2	2.018	0.183	2	1.989	0.241	2	1.975	0.295
$\Lambda_{01}(t_1)$	0.01	0.010	0.003	0.01	0.010	0.003	0.01	0.010	0.004
$\Lambda_{01}(t_2)$	0.04	0.039	0.008	0.04	0.040	0.010	0.04	0.041	0.011
$\Lambda_{01}(t_3)$	0.09	0.088	0.016	0.09	0.090	0.020	0.09	0.091	0.024
$\Lambda_{01}(t_4)$	0.16	0.154	0.029	0.16	0.160	0.034	0.16	0.159	0.040
$\Lambda_{01}(t_1)$	0.01	0.010	0.004	0.01	0.010	0.004	0.01	0.010	0.004
$\Lambda_{01}(t_2)$	0.04	0.039	0.009	0.04	0.040	0.009	0.04	0.040	0.010
$\Lambda_{01}(t_3)$	0.09	0.088	0.016	0.09	0.089	0.019	0.09	0.089	0.020
$\Lambda_{01}(t_4)$	0.16	0.156	0.027	0.16	0.158	0.031	0.16	0.157	0.033

Table 3: Comparison of $\hat{\sigma}_I^2(t)$ and $\hat{\sigma}_{II}^2(t)$ for $n = 250$ and 500 under exponential hazards

θ	minimum	1-st quartile	median	mean	3-rd quartile	maximum
<i>exponential (I)</i>						
<i>n=250</i>						
0.01005	$\hat{\sigma}_I^2(t)$	0.064	0.086	0.092	0.093	0.133
	$\hat{\sigma}_{II}^2(t)$	0.063	0.083	0.089	0.090	0.125
1	$\hat{\sigma}_I^2(t)$	0.077	0.118	0.130	0.131	0.234
	$\hat{\sigma}_{II}^2(t)$	0.072	0.108	0.117	0.118	0.190
4	$\hat{\sigma}_I^2(t)$	0.080	0.132	0.149	0.154	0.341
	$\hat{\sigma}_{II}^2(t)$	0.071	0.115	0.130	0.134	0.305
<i>n=500</i>						
0.01005	$\hat{\sigma}_I^2(t)$	0.075	0.091	0.096	0.096	0.100
	$\hat{\sigma}_{II}^2(t)$	0.075	0.090	0.095	0.095	0.115
1	$\hat{\sigma}_I^2(t)$	0.107	0.129	0.138	0.138	0.180
	$\hat{\sigma}_{II}^2(t)$	0.098	0.117	0.124	0.124	0.162
4	$\hat{\sigma}_I^2(t)$	0.089	0.123	0.137	0.138	0.199
	$\hat{\sigma}_{II}^2(t)$	0.078	0.109	0.119	0.120	0.169
<i>exponential (II)</i>						
<i>n=250</i>						
0.01005	$\hat{\sigma}_I^2(t)$	0.047	0.066	0.072	0.072	0.116
	$\hat{\sigma}_{II}^2(t)$	0.047	0.064	0.070	0.071	0.104
1	$\hat{\sigma}_I^2(t)$	0.051	0.086	0.095	0.097	0.166
	$\hat{\sigma}_{II}^2(t)$	0.049	0.080	0.088	0.089	0.149
4	$\hat{\sigma}_I^2(t)$	0.063	0.101	0.116	0.119	0.235
	$\hat{\sigma}_{II}^2(t)$	0.056	0.090	0.103	0.106	0.208
<i>n=500</i>						
0.01005	$\hat{\sigma}_I^2(t)$	0.067	0.083	0.087	0.087	0.110
	$\hat{\sigma}_{II}^2(t)$	0.067	0.082	0.087	0.087	0.107
1	$\hat{\sigma}_I^2(t)$	0.076	0.094	0.101	0.102	0.140
	$\hat{\sigma}_{II}^2(t)$	0.073	0.087	0.093	0.093	0.123
4	$\hat{\sigma}_I^2(t)$	0.074	0.108	0.119	0.120	0.207
	$\hat{\sigma}_{II}^2(t)$	0.068	0.096	0.105	0.107	0.185
<i>Weibull (I)</i>						
<i>n=250</i>						
0.01005	$\hat{\sigma}_I^2(t)$	0.058	0.084	0.089	0.089	0.127
	$\hat{\sigma}_{II}^2(t)$	0.058	0.083	0.088	0.089	0.117
1	$\hat{\sigma}_I^2(t)$	0.043	0.078	0.086	0.088	0.155
	$\hat{\sigma}_{II}^2(t)$	0.041	0.070	0.077	0.078	0.132
4	$\hat{\sigma}_I^2(t)$	0.045	0.112	0.129	0.132	0.264
	$\hat{\sigma}_{II}^2(t)$	0.042	0.098	0.112	0.114	0.235
<i>n=500</i>						
0.01005	$\hat{\sigma}_I^2(t)$	0.062	0.075	0.079	0.079	0.100
	$\hat{\sigma}_{II}^2(t)$	0.062	0.075	0.079	0.079	0.098
1	$\hat{\sigma}_I^2(t)$	0.062	0.080	0.086	0.087	0.138
	$\hat{\sigma}_{II}^2(t)$	0.056	0.072	0.077	0.077	0.116
4	$\hat{\sigma}_I^2(t)$	0.082	0.121	0.134	0.135	0.213
	$\hat{\sigma}_{II}^2(t)$	0.073	0.104	0.115	0.117	0.181
<i>Weibull (II)</i>						
<i>n=250</i>						
0.01005	$\hat{\sigma}_I^2(t)$	0.036	0.057	0.062	0.062	0.097
	$\hat{\sigma}_{II}^2(t)$	0.035	0.056	0.062	0.062	0.097
1	$\hat{\sigma}_I^2(t)$	0.045	0.067	0.075	0.076	0.122
	$\hat{\sigma}_{II}^2(t)$	0.042	0.062	0.069	0.070	0.109
4	$\hat{\sigma}_I^2(t)$	0.045	0.068	0.076	0.077	0.124
	$\hat{\sigma}_{II}^2(t)$	0.042	0.063	0.070	0.070	0.111
<i>n=500</i>						
0.01005	$\hat{\sigma}_I^2(t)$	0.045	0.058	0.062	0.062	0.086
	$\hat{\sigma}_{II}^2(t)$	0.045	0.058	0.062	0.062	0.085
1	$\hat{\sigma}_I^2(t)$	0.051	0.069	0.075	0.076	0.112
	$\hat{\sigma}_{II}^2(t)$	0.048	0.064	0.069	0.069	0.098
4	$\hat{\sigma}_I^2(t)$	0.053	0.070	0.076	0.077	0.112
	$\hat{\sigma}_{II}^2(t)$	0.049	0.064	0.069	0.070	0.099

Table 4: *Comparison of our proposed variance estimators with naive estimators*

θ	empirical SD of $S_n(t, \hat{\gamma})$	Naive		Song et al.		Proposed I		Proposed II	
		$\hat{\sigma}_1(t)$	empirical Type I error	$\hat{\sigma}_2(t)$	empirical Type I error	$\hat{\sigma}_I(t)$	empirical Type I error	$\hat{\sigma}_{II}(t)$	empirical Type I error
<i>exponential (I)</i>									
<i>n=250</i>									
0.01005	0.292	0.297	0.045	0.298	0.045	0.304	0.038	0.299	0.040
1	0.335	0.312	0.064	0.312	0.066	0.361	0.030	0.343	0.037
4	0.330	0.280	0.098	0.279	0.100	0.390	0.013	0.364	0.024
<i>n=500</i>									
0.01005	0.317	0.305	0.064	0.305	0.064	0.309	0.062	0.308	0.064
1	0.353	0.320	0.074	0.319	0.076	0.371	0.041	0.352	0.051
4	0.334	0.271	0.097	0.271	0.099	0.370	0.015	0.346	0.031
<i>exponential (II)</i>									
<i>n=250</i>									
0.01005	0.286	0.263	0.071	0.263	0.072	0.268	0.070	0.265	0.071
1	0.318	0.275	0.085	0.275	0.087	0.310	0.047	0.297	0.060
4	0.314	0.259	0.095	0.258	0.099	0.343	0.018	0.323	0.027
<i>n=500</i>									
0.01005	0.314	0.293	0.067	0.293	0.067	0.295	0.063	0.294	0.064
1	0.314	0.282	0.077	0.282	0.081	0.318	0.052	0.305	0.060
4	0.324	0.261	0.113	0.260	0.116	0.346	0.028	0.325	0.043
<i>Weibull (I)</i>									
<i>n=250</i>									
0.01005	0.305	0.294	0.056	0.294	0.056	0.299	0.053	0.297	0.055
1	0.279	0.256	0.068	0.256	0.067	0.295	0.031	0.278	0.043
4	0.302	0.264	0.081	0.263	0.080	0.361	0.014	0.336	0.023
<i>n=500</i>									
0.01005	0.283	0.279	0.053	0.279	0.050	0.281	0.050	0.282	0.050
1	0.278	0.256	0.065	0.256	0.067	0.295	0.033	0.278	0.041
4	0.308	0.266	0.080	0.265	0.078	0.340	0.026	0.367	0.017
<i>Weibull (II)</i>									
<i>n=250</i>									
0.01005	0.273	0.246	0.073	0.247	0.071	0.249	0.067	0.248	0.067
1	0.285	0.246	0.093	0.245	0.091	0.275	0.058	0.263	0.069
4	0.287	0.246	0.094	0.246	0.094	0.277	0.058	0.264	0.070
<i>n=500</i>									
0.01005	0.268	0.248	0.071	0.248	0.071	0.250	0.070	0.249	0.070
1	0.278	0.246	0.084	0.246	0.083	0.275	0.052	0.262	0.065
4	0.280	0.246	0.086	0.246	0.085	0.277	0.051	0.264	0.066

Table 5: *Empirical power of a two-sided test with $\alpha = 0.05$ and $\pi = 0.80$.*

ε	independence		mild dependence		strong dependence		Schoenfeld's sample size
	sample size	empirical power	sample size	empirical power	sample size	empirical power	
<i>exponential (I)</i>							
0.3	201	0.819	264	0.766	536	0.769	174
0.5	70	0.805	110	0.789	241	0.833	63
0.6	50	0.773	84	0.823	181	0.836	44
<i>exponential (II)</i>							
0.3	138	0.770	234	0.761	540	0.812	174
0.5	47	0.797	100	0.779	258	0.821	63
0.6	36	0.804	78	0.813	200	0.823	44
<i>Weibull (I)</i>							
0.3	166	0.798	291	0.801	773	0.785	174
0.5	65	0.819	110	0.778	344	0.819	63
0.6	46	0.783	89	0.820	178	0.813	44
<i>Weibull (II)</i>							
0.3	158	0.792	207	0.788	584	0.803	174
0.5	62	0.763	92	0.781	203	0.800	63
0.6	42	0.764	75	0.787	163	0.800	44

Table 6: *Summary of the call center data set.*

	1-st call	2-nd call	3-rd call	4-th call	5-th call
number of calls	49246	7759	1646	488	198
number of events	1416	360	89	32	18

Table 7: *The call center data set: parameters' estimates and bootstrap standard errors.*

	$\hat{\theta}$	$\hat{\beta}_1$	$\hat{\beta}_2$
point estimate	0.9973	-0.3006	-0.1211
bootstrap SE	0.1767	0.1046	0.1046

Table 8: *The call center data set: Estimates of the cumulative baseline hazard functions.*

	1-st call		2-nd call		3-rd call		4-th call		5-th call	
	bootstrap		bootstrap		bootstrap		bootstrap		bootstrap	
t	$\hat{\Lambda}_{01}(t)$	SE	$\hat{\Lambda}_{02}(t)$		$\hat{\Lambda}_{03}(t)$	SE	$\hat{\Lambda}_{04}(t)$	SE	$\hat{\Lambda}_{05}(t)$	
10	0.012	0.001	0.010	0.002	0.009	0.003	0.009	0.005	0.030	0.012
50	0.027	0.002	0.039	0.004	0.034	0.007	0.026	0.009	0.057	0.022
100	0.051	0.003	0.085	0.007	0.080	0.015	0.065	0.016	0.151	0.048
150	0.075	0.004	0.152	0.014	0.132	0.023	0.155	0.046	0.178	0.062
200	0.108	0.006	0.221	0.020	0.174	0.029	0.266	0.069	0.305	0.105
250	0.148	0.009	0.301	0.026	0.256	0.040	0.407	0.107	0.553	0.183

Table 9: *The call center data set: results of the paired tests.*

	calls	1 - 2	1 - 3	1-4	1-5	2 - 3
$S_n(250, \hat{\gamma})$		-0.464	-0.771	-0.051	-0.048	0.027
$\hat{\sigma}_{II}(250)$		0.039	0.199	0.018	0.016	0.027
$S_n(250, \hat{\gamma})/\hat{\sigma}_{II}(250)$		-11.915	-3.871	-2.884	-3.019	1.024
<i>p-value</i>		< 0.001	< 0.001	0.002	0.001	0.847
<i>FDR p-value</i>		< 0.001	< 0.001	0.042	0.003	1.000
	calls	2 - 4	2 - 5	3-4	3-5	4 - 5
$S_n(250, \hat{\gamma})$		0.058	-0.029	-0.014	-0.030	-0.019
$\hat{\sigma}_{II}(250)$		0.163	0.016	0.016	0.015	0.012
$S_n(250, \hat{\gamma})/\hat{\sigma}_{II}(250)$		0.355	-1.841	-0.907	-2.051	-1.493
<i>p-value</i>		0.639	0.033	0.182	0.020	0.068
<i>FDR p-value</i>		1.000	0.096	1.000	0.042	0.422

Table 10: *Summary of the WAS data set.*

	year of birth < 1930	year of birth $\geq$ 1930
number of subjects	4961	8069
number of events	709	297