

The Annals of Statistics
2009, Vol. 37, No. 3, 1489–1517
DOI: 10.1214/08-AOS615
© Institute of Mathematical Statistics, 2009

CASE-CONTROL SURVIVAL ANALYSIS WITH A GENERAL SEMIPARAMETRIC SHARED FRAILTY MODEL: A PSEUDO FULL LIKELIHOOD APPROACH¹

BY MALKA GORFINE, DAVID M. ZUCKER AND LI HSU

*Technion—Israel Institute of Technology, Hebrew University of Jerusalem
and Fred Hutchinson Cancer Research Center*

In this work we deal with correlated failure time (age at onset) data arising from population-based, case-control studies, where case and control probands are selected by population-based sampling and an array of risk factor measures is collected for both cases and controls and their relatives. Parameters of interest are effects of risk factors on the failure time hazard function and within-family dependencies among failure times after adjusting for the risk factors. Due to the retrospective sampling scheme, large sample theory for existing methods has not been established. We develop a novel technique for estimating the parameters of interest under a general semiparametric shared frailty model. We also present a simple, easily computed, and noniterative nonparametric estimator for the cumulative baseline hazard function. We provide rigorous large sample theory for the proposed method. We also present simulation results and a real data example for illustrating the utility of the proposed method.

1. Introduction. Clustered failure times arise often in medical and epidemiologic studies. Examples include disease onset times of twins (with time expressed in terms of age), multiple recurrence times of infections on an individual, or time to blindness for the two eyes within an individual. A typical case-control family study includes a random sample of independent diseased individuals (cases) and nondiseased individuals (controls), along with their family members. An array of genetic and environmental risk-factor measures is collected on these individuals. Integration of genetic and environmental data is a central problem of modern observational epidemiology. Case-control family studies are powerful because they provide an efficient way to assess the effect of risk factors on the occurrence of a rare disease, and furthermore allow researchers to dissect genetic and environmental contributions to the disease based on the familial aggregation pattern of disease clusters. Hopper [16] suggested that such study designs may be the future of epidemiology in general, not just genetic epidemiology. Hence, the need for statistical methods that can exploit such data is acute.

Received January 2008.

¹Supported in part by grants from the National Institute of Health and the United States–Israel Binational Science Foundation (BSF).

AMS 2000 subject classifications. 62N01, 62N02, 62H12.

Key words and phrases. Case-control study, correlated failure times, family study, frailty model, multivariate survival model.

In this work we focus on population-based, matched case-control family studies, where a number of case and control probands are randomly sampled from a well-defined population. The probands are the index subjects of the ascertained families. Here we use the term proband in a broad sense to refer to both cases and controls, in contrast with the traditional usage in which probands refer only to cases.

Relative to classical case-control methods, analysis of these studies is complicated in two major ways. First, comparisons are no longer solely between subjects with and without the disease under study, but rather between collections of the case probands and their relatives and the control probands and their relatives, with each collection typically including many subjects both with and without the disease under study. Second, data are clustered within families, and hence reflect intra-familial correlation due to unmeasured genetic and environmental factors.

Our work is motivated by a recent breast cancer study [23, 24]. In this study, the cases were incident breast cancer cases ascertained from the population-based Surveillance, Epidemiology, and End Results (SEER) cancer registry. The controls were selected by random digit dialing, and were matched with cases based on sex, age at diagnosis and county of residence. Female relatives of case and control probands were identified, and the risk factor and outcome information was collected on these relatives. The primary goals of the study are (a) to determine the degree of familial dependence with respect to age at breast cancer diagnosis; and (b) to assess the effects of covariates on breast cancer risk.

Two modeling approaches, marginal and conditional, are typically used for accounting for the correlation within a cluster. In the conditional model, the correlation is explicitly induced by a cluster-specific random effect, with the outcomes of the cluster members being conditionally independent given the random effect. In the survival context, this model is known as the frailty model, with the random frailty term typically assumed to act multiplicatively on the hazard. Many frailty models have been considered, including gamma, positive stable, inverse Gaussian, compound Poisson and log-normal. See [17] for a broad review. Under a frailty model, the regression coefficients are cluster-specific log-hazard ratios. In the marginal model, by contrast, the correlation is modeled through a multivariate distribution, often involving a copula function, with a specified model for the marginal hazard functions. The regression coefficients in the marginal model represent the log-hazard ratios at the population level regardless of which cluster an individual comes from. The effect is thus “population-averaged.” A comparison of the conditional and marginal modeling approaches can be found in [34].

Methods exist for case-control family studies under both modeling approaches. Shih and Chatterjee [31] proposed a semi-parametric quasi-partial-likelihood approach for estimating the regression coefficients in a bivariate copula model. Their cumulative hazard estimator requires an iterative solution and thus, so far, the properties of their estimators have been investigated only by simulation. Moreover, in the case of multiple relatives for each proband, the relatives were treated in the estimation process as if they were independent, which may lead to loss of efficiency in

the baseline hazard function estimator. By contrast, a quasi-EM algorithm method for the popular gamma frailty model is presented in [18], where the baseline hazard function estimator naturally accommodates multiple relatives in a family [19]. However, with this method as well, the properties of the proposed estimators were studied only by simulation. The Shih–Chatterjee method can be adapted to the family-specific frailty setting [27], but with the same limitation as for the marginal model: the lack of large sample theory.

In this work, we develop a new estimation technique for the general semiparametric shared frailty model, where the parameters of interest are the regression coefficients and the frailty parameters. Our method covers any frailty distribution with finite moments. The estimation procedure for the baseline hazard function leads to an estimator whose asymptotic properties can be derived and expressed in a tractable way. Thus, this paper is the first to present an estimation procedure with rigorous asymptotic theory for a frailty survival model in the case-control family study setting.

Section 2 presents our model and Section 3 describes our estimation procedure. Section 4 gives the consistency and asymptotic normality results. In Section 5, we describe an extension of our method for the case where the proband observation times are subject to a certain restriction that can arise in some studies. Section 6 presents simulation results. In Section 7 we illustrate our method with a case-control family study of early onset breast cancer. Section 8 provides a short discussion. Section 9 provides the details of the asymptotic theory. Throughout the paper, certain details have been omitted for brevity. These details are given in an expanded version of this paper, which is available at the *Front for the Mathematics ArXiv* under *Statistics*, publication number: [math.ST/0703300](https://arxiv.org/abs/math.ST/0703300).

2. Notation and model formulation. We consider a matched case-control family study where one case proband is age-matched with one control proband, and an array of risk factors is measured on the case and control probands and their relatives. Each matched set contains one case family and one control family, and there are n i.i.d. matched sets. Let T_{ij}^0 and C_{ij} denote the age at disease onset and age at censoring, respectively, for individual j of family i , $i = 1, \dots, 2n$, $j = 0, 1, \dots, m_i$, where $j = 0$ corresponds to the proband. Following [28] (page 187), we regard m_i as a random variable over $\{1, \dots, m\}$ for some m , and build up the remainder of the model conditional on m_i . Define $\delta_{ij} = I(T_{ij}^0 \leq C_{ij})$ to be the failure indicator and $T_{ij} = \min(T_{ij}^0, C_{ij})$ to be the observed follow-up time for individual ij . We assume that a p -vector of covariates is observed on all subjects, and let \mathbf{Z}_{ij} denote the value of the time-independent covariate vector for individual ij . In addition, we associate with family i an unobservable family-level covariate ω_i , the “frailty,” which induces dependence among family members. The conditional hazard function for proband i , given the family frailty ω_i , is assumed to take the form

$$(1) \quad \lambda_{i0}(t|\mathbf{Z}_{i0}, \omega_i) = \omega_i \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{i0}), \quad i = 1, \dots, 2n.$$

The conditional hazard function for relative ij , $j = 1, \dots, m_i$, given the family frailty ω_i , is assumed to take the form

$$(2) \quad \lambda_{ij}(t|\mathbf{Z}_{ij}, \omega_i) = \omega_i \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}), \quad i = 1, \dots, 2n; j = 1, \dots, m_i.$$

Here $\boldsymbol{\beta}$ is a p -vector of unknown regression coefficients, and λ_0 is a conditional baseline hazard of unspecified form. The above model implies that the proband and the relatives have a common conditional baseline hazard function λ_0 , and that all the dependence between the proband and the relatives in a given family is due to the frailty factor ω_i . This is the standard formulation of the frailty survival model; see, for example, [9, 10, 17] and [26]. Hence, the conditional hazard function of a relative, given \mathbf{Z}_{ij} , ω_i and the proband data, is a function of only \mathbf{Z}_{ij} and ω_i . The random variable ω_i is assumed to have a density $f(\omega) \equiv f(\omega; \theta)$, where θ is an unknown parameter. For simplicity, we assume that θ is a scalar, though the vector case could be developed in a similar manner.

We put $\boldsymbol{\gamma} = (\boldsymbol{\beta}^T, \theta)^T$, and let $\boldsymbol{\gamma}^\circ = (\boldsymbol{\beta}^{\circ T}, \theta^\circ)^T$ denote the true value of $\boldsymbol{\gamma}$. The objective is to estimate $\boldsymbol{\gamma}$ and the cumulative baseline hazard $\Lambda_0(t) = \int_0^t \lambda_0(u) du$. Let $\Lambda_0^\circ(t)$ denote the true value of Λ_0 . Further, let $\delta_{iR} = (\delta_{i1}, \dots, \delta_{im_i})$, $T_{iR} = (T_{i1}, \dots, T_{im_i})$ and $\mathbf{Z}_{iR} = (\mathbf{Z}_{i1}^T, \dots, \mathbf{Z}_{im_i}^T)^T$.

We assume that \mathbf{Z}_{ij} is bounded, and that the parameter $\boldsymbol{\gamma}$ lies in a compact subset \mathcal{G} of \mathbb{R}^{p+1} containing an open neighborhood of $\boldsymbol{\gamma}^\circ$. These two assumptions imply that $\boldsymbol{\beta}^T \mathbf{Z}_{ij}$ is bounded. In addition, we assume the following:

1. Conditional on $\{\mathbf{Z}_{ij}\}_{j=0}^{m_i}$ and ω_i , the censoring times are independent of the failure times and noninformative for ω_i and $(\boldsymbol{\beta}, \Lambda_0)$. In addition, the frailty ω_i is independent of $\{\mathbf{Z}_{ij}\}_{j=0}^{m_i}$.
2. The effect of the covariates on the observed time is subject-specific, that is $\Pr(T_{ij}, \delta_{ij}|\mathbf{Z}_{i0}, \mathbf{Z}_{iR}, \omega_i) = \Pr(T_{ij}, \delta_{ij}|\mathbf{Z}_{ij}, \omega_i)$. This implies $\Pr(T_{ij}, \delta_{ij}|\mathbf{Z}_{i0}, \mathbf{Z}_{iR}) = \Pr(T_{ij}, \delta_{ij}|\mathbf{Z}_{ij})$ even without frailty.

A number of additional technical assumptions are listed in Section 9.

Following [18] and [31], the likelihood function for the data can be written as

$$(3) \quad L = \prod_{i=1}^{2n} f(T_{iR}, \delta_{iR}, \mathbf{Z}_{iR}, \mathbf{Z}_{i0}|T_{i0}, \delta_{i0}) \\ = \prod_{i=1}^{2n} f(T_{iR}, \delta_{iR}|\mathbf{Z}_{iR}, \mathbf{Z}_{i0}, T_{i0}, \delta_{i0}) \times f(\mathbf{Z}_{iR}|\mathbf{Z}_{i0}) \times f(\mathbf{Z}_{i0}|T_{i0}, \delta_{i0}).$$

Since $f(\mathbf{Z}_{iR}|\mathbf{Z}_{i0})$ does not depend on the parameters of interest $(\boldsymbol{\beta}, \Lambda_0, \theta)$, this term will be ignored. In the following subsections we consider the other two terms in (3).

2.1. *The likelihood for the proband data.* To account for the matching of age of onset, as in [18] and [31], the likelihood function of the proband data, $\prod_{i=1}^{2n} f(\mathbf{Z}_{i0}|T_{i0}, \delta_{i0})$, is constructed based on the retrospective likelihood for the standard matched case-control study [29]. We express this likelihood in terms of the marginal survival function $S_{i0}(t) = \Pr(T_{i0} > t|\mathbf{Z}_{i0}) = \int \Pr(T_{i0} > t|\mathbf{Z}_{i0}, \omega) f(\omega) d\omega$. In our setting we have n one-to-one matched sets. Based on the marginal survivor function, the marginal hazard function can be written as

$$\lambda_{i0}(t|\mathbf{Z}_{i0}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{i0}) \frac{\mu_{1i}(t; \boldsymbol{\gamma}, \Lambda_0)}{\mu_{0i}(t; \boldsymbol{\gamma}, \Lambda_0)},$$

where

$$\mu_{ki}(t; \boldsymbol{\gamma}, \Lambda_0) = \int \omega^k \exp\{-\omega H_{i0}(t)\} f(\omega) d\omega, \quad k = 0, 1, 2,$$

and $H_{i0}(t) = \Lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{i0})$. We arrange the notation so that the first n families are the case families and the r th case family, $r = 1, \dots, n$, is matched with the $(n + r)$ th control family. The likelihood for the proband data is then replaced by the following conditional likelihood:

$$(4) \quad L^{(1)} = \prod_{r=1}^n \frac{\exp(\boldsymbol{\beta}^T \mathbf{Z}_{r0}) \xi_{10r}(T_{r0}; \boldsymbol{\gamma}, \Lambda_0)}{\sum_{j \in \{r, n+r\}} \exp(\boldsymbol{\beta}^T \mathbf{Z}_{j0}) \xi_{10j}(T_{j0}; \boldsymbol{\gamma}, \Lambda_0)},$$

where $\xi_{kk'i}(t; \boldsymbol{\gamma}, \Lambda_0) = \mu_{ki}(t; \boldsymbol{\gamma}, \Lambda_0) / \mu_{k'i}(t; \boldsymbol{\gamma}, \Lambda_0)$ for $k, k' = 0, 1, 2$. Given (4), the likelihood score functions $U_l^{(1)}(\boldsymbol{\gamma}, \Lambda_0) = \partial \log L^{(1)} / \partial \beta_l, l = 1, \dots, p$, and $U_{p+1}^{(1)}(\boldsymbol{\gamma}, \Lambda_0) = \partial \log L^{(1)} / \partial \theta$ can be obtained by straightforward differentiation. The detailed formulas are presented in the expanded paper.

Under the gamma frailty model, we have $\mu_{1i}(t; \boldsymbol{\gamma}, \Lambda_0) / \mu_{0i}(t; \boldsymbol{\gamma}, \Lambda_0) = \{\theta \times H_{i0}(t) + 1\}^{-1}$, and so the likelihood function (4) corresponds to that presented in [18] in the case of one-to-one matching. Extension to matching of multiple cases or multiple controls are straightforward, see, for example, [4].

2.2. *The likelihood for the data from the relatives.* Let $N_{ij}(t) = \delta_{ij} I(T_{ij} \leq t)$, $j = 1, \dots, m_i$, $N_i.(t) = \sum_{j=1}^{m_i} N_{ij}(t)$, $H_{ij}(t) = \Lambda_0(T_{ij} \wedge t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})$, $a \wedge b = \min\{a, b\}$, $j = 1, \dots, m_i$ and $H_i.(t) = \sum_{j=1}^{m_i} H_{ij}(t)$ and let τ be the maximum follow-up time. The likelihood of the data from the relatives then can be written as

$$(5) \quad L^{(2)} = \prod_{i=1}^{2n} \int \prod_{j=1}^{m_i} \{\lambda_{ij}(T_{ij}|T_{i0}, \delta_{i0}, \mathbf{Z}_{i0}, \mathbf{Z}_{ij}, \omega)\}^{\delta_{ij}} S_{ij}(T_{ij}|T_{i0}, \delta_{i0}, \mathbf{Z}_{i0}, \mathbf{Z}_{ij}, \omega) \times f(\omega|T_{i0}, \delta_{i0}, \mathbf{Z}_{i0}) d\omega$$

$$= \prod_{i=1}^{2n} \prod_{j=1}^{m_i} \{\lambda_0(T_{ij}) e^{\boldsymbol{\beta}^T \mathbf{Z}_{ij}}\}^{\delta_{ij}} \prod_{i=1}^{2n} \int \omega^{N_i.(\tau)} e^{-\omega H_i.(\tau)} f(\omega|T_{i0}, \delta_{i0}, \mathbf{Z}_{i0}) d\omega.$$

Here, by a Bayes theorem argument,

$$(6) \quad f(\omega|T_{i0}, \delta_{i0}, \mathbf{Z}_{i0}) = \frac{\omega^{\delta_{i0}} \exp(-\omega \Lambda_0(T_{i0}) e^{\beta^T \mathbf{Z}_{i0}}) f(\omega)}{\int \tilde{\omega}^{\delta_{i0}} \exp(-\tilde{\omega} \Lambda_0(T_{i0}) e^{\beta^T \mathbf{Z}_{i0}}) f(\tilde{\omega}) d\tilde{\omega}}.$$

Given (5), the likelihood score functions $U_l^{(2)}(\boldsymbol{\gamma}, \Lambda_0) = \partial \log L^{(2)} / \partial \beta_l, l = 1, \dots, p$ and $U_{p+1}^{(2)}(\boldsymbol{\gamma}, \Lambda_0) = \partial \log L^{(2)} / \partial \theta$ can be obtained by straightforward differentiation. The detailed formulas are presented in the expanded paper.

3. The proposed approach. We focus first on estimating the baseline cumulative hazard function $\Lambda_0(t)$. Let $Y_{ij}(t) = I(T_{ij} \geq t)$ and let \mathcal{F}_t denote the σ -algebra generated by $(T_{i0}, \delta_{i0}, \mathbf{Z}_{i0})$ plus the entire observed history of the relatives up to time t :

$$\mathcal{F}_t = \sigma(T_{i0}, \delta_{i0}, \mathbf{Z}_{i0}, N_{ij}(u), Y_{ij}(u), \mathbf{Z}_{ij}; i = 1, \dots, 2n; j = 1, \dots, m_i; 0 \leq u \leq t).$$

It is worth noting that the observational times T_{i0} for probands can be greater than time t and thus the filtration \mathcal{F}_t may include probands' failure times or censoring times that are beyond t , a feature that is unique for case-control family data. In regard to the relatives, however, \mathcal{F}_t includes only information up to time t .

By the innovation theorem ([3], Theorem 3.4), the stochastic intensity process for $N_{ij}(t), i = 1, \dots, 2n, j = 1, \dots, m_i$, with respect to \mathcal{F}_t is given as follows [11] and [28]:

$$(7) \quad \lambda_0(t) \exp(\beta^T \mathbf{Z}_{ij}) Y_{ij}(t) \psi_i(t-, \boldsymbol{\gamma}, \Lambda_0),$$

where, using (6),

$$\begin{aligned} \psi_i(t, \boldsymbol{\gamma}, \Lambda_0) &= E[\omega_i | \mathcal{F}_t] = \frac{\int \omega^{N_{i\cdot}(t)+1} \exp(-\omega H_{i\cdot}(t)) f(\omega|T_{i0}, \delta_{i0}, \mathbf{Z}_{i0}) d\omega}{\int \omega^{N_{i\cdot}(t)} \exp(-\omega H_{i\cdot}(t)) f(\omega|T_{i0}, \delta_{i0}, \mathbf{Z}_{i0}) d\omega} \\ &= \frac{\int \omega^{N_{i\cdot}(t)+1+\delta_{i0}} \exp(-\omega\{H_{i\cdot}(t) + H_{i0}(T_{i0})\}) f(\omega) d\omega}{\int \omega^{N_{i\cdot}(t)+\delta_{i0}} \exp(-\omega\{H_{i\cdot}(t) + H_{i0}(T_{i0})\}) f(\omega) d\omega}. \end{aligned}$$

Define (for $0 \leq r \leq m + L$ and $h \geq 0$)

$$(8) \quad \psi^*(r, h) = \frac{\int w^{r+1} e^{-hw} f(w) dw}{\int w^r e^{-hw} f(w) dw}.$$

Some salient properties of $\psi^*(r, h)$ are noted in Section 9.1. With this definition, we have $\psi_i(t, \boldsymbol{\gamma}, \Lambda_0) = \psi^*(N_{i\cdot}(t) + \delta_{i0}, H_{i\cdot}(t) + H_{i0}(T_{i0}))$.

Let $\tau_g, g = 1, \dots, G$, denote the g th ordered failure time of the relatives and assume that d_g failures were observed at time τ_g . In theory, since we are dealing with continuous survival distributions, $d_g = 1$ for all g , but we express the following estimators in a form that allows for a modest level of ties. A Breslow-type estimator of the cumulative baseline hazard function, with a jump at each observed

failure time among the relatives, can be formulated in a natural way similarly to Shih and Chatterjee [31], with the g th jump given by

$$(9) \quad \frac{d_g}{\sum_{i=1}^{2n} \psi_i(\tau_{g-1}, \boldsymbol{\gamma}, \Lambda_0) \sum_{j=1}^{m_i} Y_{ij}(\tau_g) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})}$$

However, $\psi_i(t, \boldsymbol{\gamma}, \Lambda_0)$ could be a function of $\Lambda_0(T_{i0})$ and T_{i0} could be greater than t . Consequently, the above Breslow formula for the jump in the baseline hazard estimator at time t will often involve values of Λ_0 for times beyond time t . For example, under the gamma frailty model with expectation 1 and variance θ , $\psi_i(t, \boldsymbol{\gamma}, \Lambda_0) = \{\theta^{-1} + N_{i\cdot}(t) + \delta_{i0}\} \{\theta^{-1} + H_{i\cdot}(t) + H_{i0}(T_{i0})\}^{-1}$. An iterative procedure is thus required to compute the estimator. In addition, because of this estimator's complicated structure, its asymptotic properties have not been established.

We propose instead to estimate the baseline hazard function using a non-iterative two-stage procedure. Let Λ_{\max} be some known (possibly large) upper bound for $\Lambda_0^\circ(t)$. Define $\bar{\psi}(r, h) = \psi^*(r, h \wedge h_{\max})$, with $h_{\max} = me^\nu \Lambda_{\max}$, where ν is an upper bound on $|\boldsymbol{\beta}^T \mathbf{Z}_{ij}|$. Further, define $\bar{\psi}_i(t, \boldsymbol{\gamma}, \Lambda) = \bar{\psi}(N_{i\cdot}(t), H_{i\cdot}(t, \boldsymbol{\gamma}, \Lambda))$. The first-stage estimator is then defined as a step function whose g th jump is given by

$$(10) \quad \Delta \tilde{\Lambda}_0(\tau_g) = \frac{d_g^*}{\sum_{i=1}^{2n} I(T_{i0} < \tau_g) \bar{\psi}_i(\tau_{g-1}, \boldsymbol{\gamma}, \tilde{\Lambda}_0) \sum_{j=1}^{m_i} Y_{ij}(\tau_g) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})},$$

with

$$d_g^* = \sum_{i=1}^{2n} I(T_{i0} < \tau_g) \sum_{j=1}^{m_i} dN_{ij}(\tau_g).$$

The formula (10) is of the same form as (9), with the following changes: (a) ψ_i is replaced by $\bar{\psi}_i$ for technical reasons, (b) more substantively, in computing the jump at each failure time τ_g , we include only relatives whose proband's observation time is less than τ_g . We thereby avoid the problem with (9) that was described above, and hence avoid the need for an iterative optimization process. Since (10) excludes some of the available data, these benefits are attained at the expense of a loss in efficiency. We therefore follow up with a second stage in order to recoup efficiency.

The second-stage estimator is defined as a step function whose g th jump is given by

$$(11) \quad \Delta \hat{\Lambda}_0(\tau_g) = \frac{d_g}{\sum_{i=1}^{2n} \tilde{\psi}_i(\tau_{g-1}, \boldsymbol{\gamma}) \sum_{j=1}^{m_i} Y_{ij}(\tau_g) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})},$$

where $\tilde{\psi}_i(t, \boldsymbol{\gamma})$ is defined analogously to $\bar{\psi}_i(t, \boldsymbol{\gamma}, \Lambda_0)$, with $\Lambda_0(T_{i0})$ replaced by $\tilde{\Lambda}_0(T_{i0})$ if $T_{i0} \geq t$ and by $\hat{\Lambda}_0(T_{i0})$ otherwise. The large-sample properties of $\hat{\Lambda}_0(t)$ will be determined by those of $\tilde{\Lambda}_0(t)$. The estimator $\tilde{\Lambda}_0(t, \boldsymbol{\gamma})$ is not necessarily

bounded by Λ_{\max} , but if desired, we can replace it by $\min\{\tilde{\Lambda}_0(t, \boldsymbol{\gamma}), \Lambda_{\max}\}$ without affecting the asymptotics.

For estimating $(\boldsymbol{\beta}, \theta)$ we use a pseudo-likelihood approach: in the score functions based on $L^{(1)}$ and $L^{(2)}$, we replace the unknown Λ_0 by $\hat{\Lambda}_0$. Thus, the score function corresponding to β_l (for $l = 1, \dots, p$) is given by $U_l(\boldsymbol{\gamma}, \hat{\Lambda}_0) = n^{-1}\{U_l^{(1)}(\boldsymbol{\gamma}, \hat{\Lambda}_0) + U_l^{(2)}(\boldsymbol{\gamma}, \hat{\Lambda}_0)\}$, and the estimating function for θ is given by $U_{p+1}(\boldsymbol{\gamma}, \hat{\Lambda}_0) = n^{-1}\{U_{p+1}^{(1)}(\boldsymbol{\gamma}, \hat{\Lambda}_0) + U_{p+1}^{(2)}(\boldsymbol{\gamma}, \hat{\Lambda}_0)\}$. To summarize, our proposed estimation procedure is as follows. (1) Provide an initial value for $\boldsymbol{\gamma}$. (2) For the given values of $\boldsymbol{\gamma}$, estimate Λ_0 using (10) and (11). (3) For the given value of Λ_0 , estimate $\boldsymbol{\gamma}$. (4) Repeat Steps 2 and 3 until convergence is reached with respect to $\hat{\Lambda}_0$ and $\hat{\boldsymbol{\gamma}}$. Hence, instead of having two iterative processes, within each iteration between $\hat{\boldsymbol{\gamma}}$ and $\hat{\Lambda}_0$ to iterate also for estimating Λ_0 itself (as in [31]), we propose one iterative process between $\hat{\boldsymbol{\gamma}}$ and $\hat{\Lambda}_0$. By eliminating the iterative process for estimating the baseline hazard function we are able to provide the asymptotic theory of our estimators, in contrast to [18] and [31].

4. Asymptotic properties. We show that $\hat{\boldsymbol{\gamma}}$ is a consistent estimator of $\boldsymbol{\gamma}^\circ$ and that $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)$ is asymptotically mean-zero multivariate normal. In this section, we present a broad outline sketch of the argument. In Section 9, we provide the details of the proofs, including a detailed list of the technical conditions assumed. The arguments are patterned after those of [14] and [36], but with considerable expansion, as will be elaborated in Section 9.

Consistency is shown through the following steps.

CLAIM A1. $\tilde{\Lambda}_0(t, \boldsymbol{\gamma})$ converges in pr. to some function $\Lambda_0^*(t, \boldsymbol{\gamma})$ uniformly in t and $\boldsymbol{\gamma}$. The function $\Lambda_0^*(t, \boldsymbol{\gamma})$ satisfies $\Lambda_0^*(t, \boldsymbol{\gamma}^\circ) = \Lambda_0^\circ(t)$.

CLAIM A2. $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$ converges in pr. to some function $\Lambda_0(t, \boldsymbol{\gamma})$ uniformly in t and $\boldsymbol{\gamma}$. The function $\Lambda_0(t, \boldsymbol{\gamma})$ satisfies $\Lambda_0(t, \boldsymbol{\gamma}^\circ) = \Lambda_0^\circ(t)$.

CLAIM B. $\mathbf{U}(\boldsymbol{\gamma}, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}))$ converges in pr. uniformly in t and $\boldsymbol{\gamma}$ to a limit $\mathbf{u}(\boldsymbol{\gamma}, \Lambda_0(\cdot, \boldsymbol{\gamma}))$.

CLAIM C. There exists a unique consistent (in pr.) root to $\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) = \mathbf{0}$.

It should be emphasized that in Claims A1 and A2 the limits coincide at the true parameter value $\boldsymbol{\gamma}^\circ$. The proofs of Claims A1, A2 and B involve empirical process and function-space compactness arguments, while Claim C is shown using Foutz’s theorem [8] on consistency of maximum likelihood type estimators.

Asymptotic normality is shown by decomposing $\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) = \mathbf{0}$ as

$$\begin{aligned} &\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) + [\mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)] \\ &\quad + [\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ))] = \mathbf{0}. \end{aligned}$$

In Section 9 we analyze each of the above three terms and prove that $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)$ has an asymptotic mean-zero multivariate normal distribution. Although it is possible to develop a consistent closed-form sandwich estimator for the asymptotic covariance matrix, we do not present this estimator because it is too complicated to be practically useful. Instead, as discussed in Section 6, we recommend bootstrap standard-error estimates.

In Section 9 we also show the uniform consistency and weak convergence of the cumulative baseline hazard function estimator $\hat{\Lambda}_0(t, \hat{\boldsymbol{\gamma}})$. Such results were not presented in [14] or [36].

5. Extension to restricted sampling of probands. A key assumption in our procedure for estimating Λ_0 is that the support of the probands' observation times and that of relatives' observation times have the same lower limit, which is designated (without loss of generality) as time zero. In some applications, however, the probands' observed times are restricted to some range $[s_0, s_1]$ with $s_0 > 0$. For example, a multi-center case-control breast cancer study, where ages of cases and controls are restricted between ages 35–64, is presented in [25]. In a design of this form, where the probands' observed times are left-restricted by s_0 and the relatives' failure times are unrestricted, Λ_0 will be underestimated by our two-stage procedure. But this bias can be easily corrected by first estimating $\Lambda_0(s_0)$.

We present here the resulting three-stage estimator for the left-restricted design. Let $\Delta\tilde{\Lambda}_0\{\tau_g, \Lambda_0(s_0)\}$, and let $\Delta\hat{\Lambda}_0\{\tau_g, \Lambda_0(s_0)\}$ be defined analogously to $\Delta\tilde{\Lambda}_0(\tau_g)$ and $\Delta\hat{\Lambda}_0(\tau_g)$, with $\Lambda_0(T_{i0}) = \Lambda_0(s_0) + \sum_{\tau_g \in [s_0, T_{i0}]} \Delta_0(\tau_g)$. The estimator $\hat{\Lambda}_0(s_0)$ is defined as the root of

$$(12) \quad \sum_{\tau_g \in [0, s_0]} \Delta\hat{\Lambda}_0\{\tau_g, \Lambda_0(s_0)\} - \Lambda_0(s_0) = 0.$$

The root can be found by simple univariate Newton–Raphson iteration. This completes the first stage. The second stage involves calculating $\Delta\tilde{\Lambda}_0\{\tau_g, \hat{\Lambda}_0(s_0)\}$, $g = 1, \dots, G$, using the formula (10). In the third stage, we use the results of the second stage and the formula (11) to calculate the final estimate $\Delta\hat{\Lambda}_0(\tau_g)$, $g = 1, \dots, G$. In applying (11), we replace $\Lambda_0(T_{i0})$ by $\tilde{\Lambda}_0\{T_{i0}, \hat{\Lambda}_0(s_0)\}$ if $T_{i0} \geq \tau_g$ and by $\hat{\Lambda}_0(T_{i0})$ otherwise.

In Section 6 below, we present simulation results for this estimator. In theory, the asymptotic properties of the three-stage procedure could be worked out via an extension of the arguments for the two-stage procedure, but the algebra becomes very complicated. We hope to develop asymptotic theory for the left-restricted design in future work.

6. Simulation results—Gamma frailty. We have performed a simulation study to evaluate the finite sample performance of the proposed method and compare it with existing methods. One of the most extensively used frailty distributions is the gamma distribution: customarily, the gamma distribution with expectation 1 and variance θ . Under this model, the variance parameter θ quantifies

TABLE 1

Simulation results: 500 control proband matched with 500 case probands; one relative for each proband; $\beta = 0.693$, $\Lambda_0(t) = t$, $\theta = 2.0$, 500 samples

	Proposed method		Hsu et al.		Shih and Chatterjee	
	Mean	Empirical standard error	Mean	Empirical standard error	Mean	Empirical standard error
$\hat{\beta}$	0.706	0.197	0.697	0.201	0.698	0.182
$\hat{\theta}$	2.003	0.312	1.986	0.302	1.992	0.303
$\hat{\Lambda}_0(0.2)$	0.201	0.034	0.204	0.030	0.202	0.029
$\hat{\Lambda}_0(0.4)$	0.402	0.063	0.407	0.058	0.403	0.054
$\hat{\Lambda}_0(0.6)$	0.603	0.095	0.612	0.090	0.605	0.084
$\hat{\Lambda}_0(0.8)$	0.809	0.136	0.820	0.131	0.811	0.122

the heterogeneity of risk among families, with larger values of θ corresponding to stronger within-family dependence. In addition, the gamma frailty model can be re-expressed in terms of the Clayton–Oakes copula-type model [6, 27]. Moreover, the cross-ratio, introduced by Oakes [27] as a local measure of association between survival times, is constant on the support of the failure time distribution and equals $1 + \theta$. Finally, the gamma frailty model is convenient mathematically, because it admits a closed-form representation of the marginal survival distributions. These features make the gamma frailty model very popular. We therefore chose the gamma frailty model as the framework for our simulation study.

Simulation results are based on 500 control probands matched to 500 case probands, with one relative sampled for each proband. We considered a single $U[0, 1]$ distributed covariate with $\beta = \ln(2)$, $\Lambda_0(t) = t$, $\theta = 2$ and a $U[0, 1]$ censoring variable, yielding a censoring rate among the relatives of approximately 60%. In Table 1 we compare the following three estimates: the proposed estimate with the two-stage procedure for Λ_0 , the estimate of [18] and a modified version of [31] estimate, with their method adapted to the gamma frailty model. Results are based on 500 simulated data sets. The efficiency difference between our two-stage estimator and that of Shih and Chatterjee is very small.

For our method, we also performed simulations for two additional settings. In both settings, we took $\beta = 0$ and $\theta = 3$. The first setting involved a censoring distribution of $U[0, 4]$ and a covariate with a $U[0, 4]$ distribution; the second involved a censoring distribution of $U[0, 0.1]$ and a covariate with a $U[0, 1]$ distribution. The respective overall censoring rates in these two settings were approximately 30% and 90%. To construct confidence intervals, we used a bootstrap approach. In the setting of censored survival data, the usual nonparametric bootstrap is problematic because it leads to a substantial proportion of tied survival times. We therefore used the weighted bootstrap approach of [22]. For the weighted bootstrap,

TABLE 2
Simulation results for the proposed estimators: 500 control proband matched with 500 case probands; one relative for each proband; $\Lambda_0(t) = t$; 2000 samples

θ	Estimator	$\beta = 0.0$			$\beta = 0.693$		
		Mean	Empirical standard error	Coverage rate	Mean	Empirical standard error	Coverage rate
90% censoring rate							
2.0	$\hat{\beta}$	-0.013	0.217	93.5	0.694	0.200	96.0
	$\hat{\theta}$	2.127	0.872	96.0	2.082	0.667	94.8
	$\hat{\Lambda}_0(0.02)$	0.020	0.006	94.2	0.020	0.005	95.2
	$\hat{\Lambda}_0(0.04)$	0.041	0.010	94.7	0.040	0.010	95.2
	$\hat{\Lambda}_0(0.06)$	0.061	0.015	94.8	0.060	0.014	96.1
	$\hat{\Lambda}_0(0.08)$	0.081	0.020	95.0	0.080	0.019	96.1
3.0	$\hat{\beta}$	-0.025	0.226	91.7	0.689	0.206	95.4
	$\hat{\theta}$	3.126	1.142	94.2	3.172	0.964	95.7
	$\hat{\Lambda}_0(0.02)$	0.020	0.005	95.7	0.020	0.005	94.8
	$\hat{\Lambda}_0(0.04)$	0.041	0.012	95.8	0.040	0.010	95.9
	$\hat{\Lambda}_0(0.06)$	0.062	0.016	96.1	0.060	0.014	96.5
	$\hat{\Lambda}_0(0.08)$	0.082	0.021	95.9	0.080	0.019	95.7
30% censoring rate							
2.0	$\hat{\beta}$	0.007	0.047	95.5	0.703	0.063	96.5
	$\hat{\theta}$	2.013	0.247	95.3	1.993	0.196	95.5
	$\hat{\Lambda}_0(0.2)$	0.200	0.037	95.5	0.197	0.045	94.5
	$\hat{\Lambda}_0(0.4)$	0.397	0.073	95.0	0.394	0.085	94.0
	$\hat{\Lambda}_0(0.6)$	0.596	0.110	95.1	0.591	0.125	94.0
	$\hat{\Lambda}_0(0.8)$	0.794	0.147	95.5	0.788	0.166	94.1
3.0	$\hat{\beta}$	0.006	0.048	97.3	0.703	0.061	97.2
	$\hat{\theta}$	3.009	0.370	95.3	2.999	0.314	96.0
	$\hat{\Lambda}_0(0.2)$	0.200	0.040	94.0	0.197	0.047	94.4
	$\hat{\Lambda}_0(0.4)$	0.399	0.078	94.1	0.392	0.091	94.0
	$\hat{\Lambda}_0(0.6)$	0.597	0.116	95.0	0.586	0.133	94.9
	$\hat{\Lambda}_0(0.8)$	0.796	0.155	95.6	0.792	0.176	95.0

a sample of $2n$ independent and identically distributed weights from the unit exponential distribution was generated for each bootstrap sample. Let ξ_1, \dots, ξ_{2n} be the standardized weights after dividing each weight by the average weight. Then, in the estimating functions, for any given function h the empirical mean $n^{-1} \sum_{i=1}^{2n} h(T_i, \delta_i, \mathbf{Z}_i)$ is replaced by its corresponding weighted empirical mean $n^{-1} \sum_{i=1}^{2n} \xi_i h(T_i, \delta_i, \mathbf{Z}_i)$. This weighted bootstrap procedure gives valid inference for all parameters under right-censored univariate failure times [22].

Results for the two-stage procedure for Λ_0 are presented in Table 2 for various levels of censoring. We present the mean, the empirical standard error, and the

TABLE 3

Simulation results of left-restricted data: 500 control proband matched with 500 case probands; one relative for each proband; $s_0 = 0.1$, $\beta = 0.693$, $\Lambda_0(t) = t$, $\theta = 2.0$, 500 samples

	Proposed method		Hsu et al.		Shih and Chatterjee	
	Mean	Empirical standard error	Mean	Empirical standard error	Mean	Empirical standard error
$\hat{\beta}$	0.735	0.214	0.698	0.234	0.694	0.170
$\hat{\theta}$	2.040	0.336	2.080	0.338	2.080	0.337
$\hat{\Lambda}_0(0.2)$	0.195	0.049	0.198	0.034	0.198	0.031
$\hat{\Lambda}_0(0.4)$	0.392	0.090	0.402	0.068	0.401	0.062
$\hat{\Lambda}_0(0.6)$	0.589	0.129	0.604	0.102	0.603	0.092
$\hat{\Lambda}_0(0.8)$	0.786	0.172	0.813	0.143	0.810	0.128
$\hat{\Lambda}_0(s_0)$	0.098	0.025	—	—	—	—

coverage rate of the 95% weighted bootstrap confidence interval. The results are based on 50 bootstrap samples for each of the 2000 simulated data sets of each configuration. Our estimates perform well in terms of bias and coverage probability.

To study the case of left-restricted data, we considered a configuration similar to that of Table 1, but with the probands observation times restricted to be > 0.1 . In Table 3, we present results for our three-stage estimator as well as for the methods of [18] and [31]. We see that estimating $\Lambda_0(s_0)$ leads to modest efficiency loss in $\hat{\Lambda}_0$ relative to the other two methods.

7. Example. We apply our method to the breast-cancer study mentioned in the Introduction. Various risk factors were measured on probands and their relatives. For illustrative purposes we consider age at first full-term pregnancy with the relatives of the probands being the mothers. The following analysis is based on 437 breast-cancer case probands matched with 437 control probands and a total of 874 mothers. The number of mothers who had breast cancer was 70 among the case families and 35 among the control families. The number of women whose first live birth occurred before age 20 was 142 among the probands and 181 among the mothers. We use the gamma frailty model with expectation 1 and variance θ . Three estimation procedures are considered: our proposed method, the estimate of the Hsu et al. [18] method and a modified version of the Shih and Chatterjee [31] estimate. For our proposed method, the two-stage procedure for Λ_0 is used since the age range of the mothers with breast cancer was 20–76 and the age range of the probands was 22–44. Table 4 presents the regression coefficient parameter estimate $\hat{\beta}$, the dependency parameter estimate, $\hat{\theta}$, and $\hat{\Lambda}_0$ at ages 40, 50, 60 and 70 years, along with their respective bootstrap standard errors. The proposed approach and that of Shih and Chatterjee yielded similar dependency estimates

TABLE 4
Analysis of a case-control family study of breast cancer

	Proposed method		Hsu et al.		Shih and Chatterjee	
	Mean	Bootstrap standard error	Mean	Bootstrap standard error	Mean	Bootstrap standard error
$\hat{\beta}$	-0.440	0.158	-0.484	0.216	-0.476	0.168
$\hat{\theta}$	0.952	0.443	0.889	0.443	0.944	0.460
$\hat{\Lambda}_0(40)$	0.005	0.002	0.005	0.002	0.005	0.002
$\hat{\Lambda}_0(50)$	0.022	0.006	0.023	0.006	0.023	0.006
$\hat{\Lambda}_0(60)$	0.048	0.010	0.051	0.010	0.049	0.010
$\hat{\Lambda}_0(70)$	0.091	0.016	0.095	0.016	0.092	0.016

with the proposed approach being moderately more efficient. Hsu et al.’s approach gave a slightly lower dependence estimate. The regression coefficient estimates of Hsu et al. and that of Shih and Chatterjee are similar, with the latter being slightly more efficient. The proposed approach yielded a slightly lower covariate effect. The cumulative baseline hazard estimates are similar under the three estimation techniques. The results of all the three methods imply that women who had their first full-term pregnancy before age 20 have a reduced risk of developing breast cancer, supporting the observation that breast cancer risk is reduced by early first full-term pregnancy (e.g., [7], among others). The estimates of the dependency parameter imply that after adjusting for the first full-term pregnancy, there remains a significant dependency between the ages of onset for mothers and daughters with cross ratio $(1 + \theta)$ close to 2.

8. Discussion. In this work we have presented a new estimator for matched case-control family study survival data under a frailty model, allowing an arbitrary frailty distribution with finite moments. Rigorous large sample theory has been provided. Simulation results under the popular gamma frailty model indicate that the proposed procedure provides estimates with minimal bias and confidence intervals with the appropriate coverage rate. Moreover, our estimators were seen to be essentially identical in efficiency to estimators based on the more complex approach of Shih and Chatterjee [31].

Rigorous large sample theory has been provided for age-unrestricted sampling of cases and controls. For age-restricted sampling, the asymptotic theory could be worked out largely following the arguments for the two-stage estimator but the algebra becomes very complicated. This development is a potential topic for future work.

Having suggested a model with an arbitrary frailty distribution with finite moments, we offer some remarks on how to choose the frailty distribution and the effect of this choice on the parameter estimates. Hougaard [17] provides a

comprehensive discussion of the theoretical properties and the fit of the following distributions: gamma, positive stable, power variance function (PVF), inverse Gaussian and lognormal. Hsu et al. [20] show by simulation that the biases in the marginal regression estimates and the marginal hazard function are generally 10% or lower under the assumed gamma distribution and misspecification of the frailty distribution. This suggests that the gamma frailty model can be a practical choice if the marginal parameters are of primary interest. However, when the dependence function is also of interest, a correct specification of the frailty distribution is crucial. A general diagnostic approach to check the bivariate association structure of clustered failure times is given in [13]. Additional tests and graphical procedures for checking the dependence structure of clustered failure-time data can be found in [5, 12, 30, 32, 33]. These procedures, however, are not directly applicable to the retrospective setting we deal with here; extension of the procedures to this setting will be needed.

9. Asymptotic theory: Conditions and proofs. This section presents the technical conditions we assume for the asymptotic results and the proofs of these results. The pattern of the argument is as in [14, 35] and [36], but considerable extension is required, mainly because of the two-stage cumulative baseline hazard function estimator for the case-control family data. A pseudo full likelihood estimation procedure for prospective survival data with a general semiparametric shared frailty model is given by [14] and [36]. A pseudo partial likelihood method for semiparametric survival model with errors in covariates is provide by [35]. We focus here on the added arguments needed for the present setting, and refer back to [35] and [36] for the other derivations. Briefly, the main extensions required in this work are as follows: (i) In showing consistency of $\tilde{\Lambda}_0(\cdot, \boldsymbol{\gamma})$, the proof of Claim A in [36] cannot be applied directly since the quantity

$$\sum_{i=1}^{2n} I(T_{i0} < t) \bar{\psi}_i(t-, \boldsymbol{\gamma}, \Lambda) \sum_{j=1}^{m_i} Y_{ij}(t) \exp \boldsymbol{\beta}^T \mathbf{Z}_{ij},$$

which is the denominator of the term for the jumps in $\tilde{\Lambda}_0$, tends to 0 as $t \rightarrow 0$. (ii) For the asymptotic normality of $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)$, we need a workable representation of the baseline hazard function estimators (10) and (11). The approach of [36] cannot be applied directly since $\tilde{\Lambda}_0(t) - \Lambda_0^\circ(t)$ involves the “vanishing-denominator” problem mentioned above. For $\hat{\Lambda}_0(t) - \Lambda_0^\circ(t)$ we use the representation of $\tilde{\Lambda}_0(t) - \Lambda_0^\circ(t)$ along with a recursive solution only for the relatives’ failure times.

9.1. *Assumptions, background and preliminaries.* For the asymptotic theory, we make a number of assumptions. Several of these assumptions have already been listed in the main text. Below we list the additional assumptions.

1. There is a maximum follow-up time $\tau \in (0, \infty)$, with $E[\sum_{j=1}^{m_i} Y_{ij}(\tau)] = y^* > 0 \forall i$.

2. The frailty random variable ω_i has finite moments up to order $(m + 2)$.
3. The baseline hazard function $\lambda_0^\circ(t)$ is bounded over $[0, \tau]$ by some fixed (but not necessarily known) constant λ_{\max} .
4. The function $f'(w; \theta) = (d/d\theta)f(w; \theta)$ is absolutely integrable.
5. For any given family, there is a positive probability of at least two failures.
6. Defining $\pi(s) = E[I(T_{i0} < s) \sum_{j=1}^{m_i} Y_{ij}(s)]$, we have

$$(13) \quad \xi_r(u) \equiv \int_0^u \frac{\lambda_0^\circ(s)}{\pi(s)^r} ds < \infty \quad \text{for all } u \in [0, \tau] \text{ and } r = 1, 2, 3.$$

This assumption is needed in the analysis of the first-stage estimator. For $r = 1$, it parallels Assumption 5.4 of [21].

7. The matrix $[(\partial/\partial \boldsymbol{\gamma})\mathbf{U}(\boldsymbol{\gamma}, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}))]|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^\circ}$ is invertible with probability going to 1 as $n \rightarrow \infty$. It should be noted that a general proof of invertibility is intractable, but given the data, one can easily check that numerically the matrix is invertible.

With $\psi^*(r, h)$ as in (8), we define $\psi_{\min}^*(h) = \min_{0 \leq r \leq m} \psi^*(r, h)$ and $\psi_{\max}^*(h) \times \max_{0 \leq r \leq m} \psi^*(r, h)$. It is easily seen that $\psi_{\max}^*(h)$ is finite and $\psi_{\min}^*(h)$ is strictly positive. The two lemmas below correspond to Lemmas 1 and 3 of [36].

LEMMA 1. *The function $\psi^*(r, h)$ is decreasing in h . Hence for all $\boldsymbol{\gamma} \in \mathcal{G}$ and all t , $\psi_i(\boldsymbol{\gamma}, \Lambda, t) \leq \psi_{\max}^*(0)$ and $\psi_i(\boldsymbol{\gamma}, \Lambda, t) \geq \psi_{\min}^*(me^\nu \Lambda(t))$.*

LEMMA 2. *For any $\varepsilon > 0$, we have $\sup_{s \in [\varepsilon, \tau]} |\tilde{\Lambda}_0(s, \boldsymbol{\gamma}^\circ) - \tilde{\Lambda}_0(s-, \boldsymbol{\gamma}^\circ)| \rightarrow 0$ as $n \rightarrow \infty$.*

9.2. *Consistency.* As indicated in Section 4, the consistency proof proceeds in several stages.

CLAIM A1. *$\tilde{\Lambda}_0(t, \boldsymbol{\gamma})$ converges in probability to some function $\Lambda_0^*(t, \boldsymbol{\gamma})$ uniformly in t and $\boldsymbol{\gamma}$. The function $\Lambda_0^*(t, \boldsymbol{\gamma})$ satisfies $\Lambda_0^*(t, \boldsymbol{\gamma}^\circ) = \Lambda_0^\circ(t)$.*

PROOF. We can write $\tilde{\Lambda}_0(t, \boldsymbol{\gamma})$ as

$$(14) \quad \tilde{\Lambda}_0(t, \boldsymbol{\gamma}) = \int_0^t \frac{n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \sum_{j=1}^{m_i} dN_{ij}(s)}{n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \bar{\psi}_i(s-, \boldsymbol{\gamma}, \tilde{\Lambda}) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})}.$$

The proof builds here on that of the corresponding Claim A in [36]. The main point needing attention here is the fact that, because of the indicators $I(T_{i0} < s)$, the denominator of (14) tends to 0 as $s \rightarrow 0$.

Define, in parallel with [36],

$$\Xi_n(t, \boldsymbol{\gamma}, \Lambda, \varepsilon) = \int_\varepsilon^t \frac{n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \sum_{j=1}^{m_i} dN_{ij}(s)}{n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \bar{\psi}_i(s-, \boldsymbol{\gamma}, \Lambda) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})}$$

and

$$\begin{aligned} \Xi(t, \boldsymbol{\gamma}, \Lambda, \varepsilon) &= \int_\varepsilon^t \frac{E[I(T_{i0} < s) \bar{\psi}_i(s-, \boldsymbol{\gamma}, \Lambda_0^\circ) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{ij})]}{E[I(T_{i0} < s) \bar{\psi}_i(s-, \boldsymbol{\gamma}, \Lambda) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})]} \lambda_0^\circ(s) ds. \end{aligned}$$

We write $\Xi_n(t, \boldsymbol{\gamma}, \Lambda)$ for $\Xi_n(t, \boldsymbol{\gamma}, \Lambda, 0)$ and $\Xi(t, \boldsymbol{\gamma}, \Lambda)$ for $\Xi(t, \boldsymbol{\gamma}, \Lambda, 0)$. By definition, $\tilde{\Lambda}_0(t, \boldsymbol{\gamma})$ satisfies the equation $\tilde{\Lambda}_0(t, \boldsymbol{\gamma}) = \Xi_n(t, \boldsymbol{\gamma}, \tilde{\Lambda}_0(\cdot, \boldsymbol{\gamma}))$.

REMARK. In [36], we had $\Xi_n(t, \boldsymbol{\gamma}, \Lambda) \rightarrow \Xi(t, \boldsymbol{\gamma}, \Lambda)$ a.s. as $n \rightarrow \infty$, uniformly over $t \in [0, \tau]$, $\boldsymbol{\gamma} \in \mathcal{G}$, and Λ in a certain set. We could not obtain such a result here; the argument of [2] fails in the neighborhood of zero because of the “vanishing denominator” problem. This is why we give only an in pr. consistency result rather than an a.s. result.

Next, define

$$q_{\boldsymbol{\gamma}}(s, \Lambda) = \frac{E[I(T_{i0} < s) \bar{\psi}_i(s-, \boldsymbol{\gamma}, \Lambda_0^\circ) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{ij})]}{E[I(T_{i0} < s) \bar{\psi}_i(s-, \boldsymbol{\gamma}, \Lambda) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})]} \lambda_0^\circ(s).$$

This function $q_{\boldsymbol{\gamma}}(s, \Lambda)$ has the same properties as its counterpart in [36]: these properties are not interfered with by the insertion of the indicator function $I(T_{i0} < s)$. In particular, from Lemma 1 we have

$$[\bar{\psi}_i(s-, \boldsymbol{\gamma}, \Lambda_0^\circ) / \bar{\psi}_i(s-, \boldsymbol{\gamma}, \Lambda)] \leq [\psi_{\max}^*(0) / \psi_{\min}^*(h_{\max})].$$

This leads to a bound on $q_{\boldsymbol{\gamma}}(s, \Lambda)$. In addition, the function $q_{\boldsymbol{\gamma}}(s, \Lambda)$ has the following Lipschitz-like property:

$$|q_{\boldsymbol{\gamma}}(s, \Lambda_1) - q_{\boldsymbol{\gamma}}(s, \Lambda_2)| \leq K \sup_{0 \leq u \leq s} |\Lambda_1(u) - \Lambda_2(u)|.$$

Hence, by [15] (Theorem 1.1), the equation $\Lambda(t) = \Xi(t, \boldsymbol{\gamma}, \Lambda)$ has a unique solution, which we denote by $\Lambda_0^*(t, \boldsymbol{\gamma})$. The claim then is that $\tilde{\Lambda}_0(t, \boldsymbol{\gamma})$ converges in pr. (uniformly in t and $\boldsymbol{\gamma}$) to $\Lambda_0^*(t, \boldsymbol{\gamma})$.

We now define $\tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon)$ to be the solution of $\tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon) = \Xi_n(t, \boldsymbol{\gamma}, \tilde{\Lambda}_0(\cdot, \boldsymbol{\gamma}), \varepsilon)$, starting from $\tilde{\Lambda}_0(\varepsilon, \boldsymbol{\gamma}, \varepsilon) = 0$. For $t < \varepsilon$ we set $\tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon) = 0$. Similarly, we define $\Lambda_0^*(t, \boldsymbol{\gamma}, \varepsilon)$ to be the solution of $\Lambda_0(t, \boldsymbol{\gamma}, \varepsilon) = \Xi(t, \boldsymbol{\gamma}, \Lambda_0(\cdot, \boldsymbol{\gamma}), \varepsilon)$, starting from $\Lambda_0(\varepsilon, \boldsymbol{\gamma}, \varepsilon) = 0$, and set $\Lambda_0^*(t, \boldsymbol{\gamma}, \varepsilon) = 0$ for $t < \varepsilon$. An induction argument similar to that in the proof of [15] (Theorem 1.1) shows that

$|\Lambda_0^*(t, \boldsymbol{\gamma}, \varepsilon) - \Lambda_0^*(t, \boldsymbol{\gamma})| \leq e^K \Lambda_0(\varepsilon, \boldsymbol{\gamma})$, where K is the Lipschitz constant for $q_{\boldsymbol{\gamma}}(s, \Lambda)$. We thus have

$$(15) \quad \sup_{\boldsymbol{\gamma} \in \mathcal{G}, t \in [0, \tau]} |\Lambda_0^*(t, \boldsymbol{\gamma}, \varepsilon) - \Lambda_0^*(t, \boldsymbol{\gamma})| \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Now, for any given $\varepsilon > 0$, there is no “vanishing denominator” problem on the interval $[\varepsilon, \tau]$. Hence, the argument in [36] goes through as is, and we get the following result: for any $\varepsilon > 0$,

$$(16) \quad \sup_{\boldsymbol{\gamma} \in \mathcal{G}, t \in [\varepsilon, \tau]} |\tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon) - \Lambda_0^*(t, \boldsymbol{\gamma}, \varepsilon)| \rightarrow 0 \quad \text{as } n \rightarrow \infty \text{ a.s.}$$

(and hence in probability). In fact, in the supremum above, we can replace $[\varepsilon, \tau]$ by $[0, \tau]$, since $\tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon) = \Lambda_0(t, \boldsymbol{\gamma}, \varepsilon) = 0$ for $t < \varepsilon$.

Our aim now is to show that $\sup_{\boldsymbol{\gamma} \in \mathcal{G}, t \in [0, \tau]} |\tilde{\Lambda}_0(t, \boldsymbol{\gamma}) - \Lambda_0^*(t, \boldsymbol{\gamma})| \rightarrow 0$ in pr. as $n \rightarrow \infty$. Now

$$(17) \quad \begin{aligned} |\tilde{\Lambda}_0(t, \boldsymbol{\gamma}) - \Lambda_0^*(t, \boldsymbol{\gamma})| &\leq |\tilde{\Lambda}_0(t, \boldsymbol{\gamma}) - \tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon)| \\ &\quad + |\tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon) - \Lambda_0^*(t, \boldsymbol{\gamma}, \varepsilon)| \\ &\quad + |\Lambda_0^*(t, \boldsymbol{\gamma}, \varepsilon) - \Lambda_0^*(t, \boldsymbol{\gamma})|. \end{aligned}$$

The second and third terms are easily dealt with using (15) and (16). It remains to deal with the first term.

Define

$$\begin{aligned} C(s, \boldsymbol{\gamma}, \varepsilon) &= \frac{1}{n} \sum_{i=1}^{2n} I(T_{i0} < s) \bar{\psi}_i(s-, \boldsymbol{\gamma}, \tilde{\Lambda}(\cdot, \boldsymbol{\gamma}, \varepsilon)) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}), \\ C^*(s, \boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{2n} I(T_{i0} < s) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}). \end{aligned}$$

We can then write

$$(18) \quad \tilde{\Lambda}_0(t, \boldsymbol{\gamma}) - \tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon) = \tilde{\Lambda}_0(t \wedge \varepsilon, \boldsymbol{\gamma}) + A(t, \varepsilon),$$

where

$$A(t, \varepsilon) = \int_{t \wedge \varepsilon}^t [C(s, \boldsymbol{\gamma}, 0)^{-1} - C(s, \boldsymbol{\gamma}, \varepsilon)^{-1}] \left[\frac{1}{n} \sum_{i=1}^{2n} I(T_{i0} < s) \sum_{j=1}^{m_i} dN_{ij}(s) \right].$$

We deal with the two terms on the right side of (18) in turn. In what follows, we let R denote a “generic” constant which may vary from one appearance to another, but does not depend on the unknown parameters or ε .

Denote $\Pi(s) = C^*(s, \mathbf{0})$. It is clear that $\tilde{\Lambda}(t, \boldsymbol{\gamma}) \leq R\Upsilon(t, \boldsymbol{\gamma})$, where

$$\Upsilon(t, \boldsymbol{\gamma}) = \int_0^t \Pi(s)^{-1} \left[\frac{1}{n} \sum_{i=1}^{2n} I(T_{i0} < s) \sum_{j=1}^{m_i} dN_{ij}(s) \right].$$

We can write

$$\Upsilon(t, \boldsymbol{\gamma}) = \int_0^t \left[\frac{C^*(s, \boldsymbol{\beta}^\circ)}{\Pi(s)} \right] \lambda_0^\circ(s) ds + \int_0^t \Pi(s)^{-1} \left[\frac{1}{n} \sum_{i=1}^{2n} I(T_{i0} < s) \sum_{j=1}^{m_i} dM_{ij}(s) \right],$$

where M_{ij} is the martingale process corresponding to N_{ij} :

$$(19) \quad M_{ij}(t) = N_{ij}(t) - \int_0^t \lambda_0(u) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{ij}) Y_{ij}(u) \psi_i(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, u-) du.$$

The first term in $\Upsilon(t, \boldsymbol{\gamma})$ is clearly bounded by $R\Lambda_0^\circ(t)$. Thus, denoting the second term by $M^*(t)$, we have

$$(20) \quad \tilde{\Lambda}(t, \boldsymbol{\gamma}) \leq R \left[\Lambda_0^\circ(t) + \sup_{u \in [0, \tau]} |M^*(u)| \right].$$

We next examine $A(t, \varepsilon)$. We can restrict to $t \geq \varepsilon$, since $A(t, \varepsilon) = 0$ for $t < \varepsilon$. Denote $\Delta(t) = \tilde{\Lambda}_0(t, \boldsymbol{\gamma}) - \tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon)$. Bearing in mind the Lipschitz property of $\tilde{\psi}$, we find that

$$|A(t, \varepsilon)| \leq R \int_\varepsilon^t |\Delta(s-)| d\Upsilon(s).$$

Note that, for $t \geq \varepsilon$, $dA(t, \varepsilon) = d\Delta(t)$. Thus, a simple induction and some additional simple manipulations lead to the following, where we employ the symbol \mathcal{P} to denote product integral and use the fact that $\Delta(\varepsilon) = \tilde{\Lambda}(\varepsilon, \boldsymbol{\gamma})$:

$$\begin{aligned} |A(t, \varepsilon)| &\leq |\Delta(\varepsilon)| \mathcal{P}_\varepsilon^t (1 + Rd\Upsilon(s)) \leq |\Delta(\varepsilon)| \exp(R[\Upsilon(t) - \Upsilon(\varepsilon)]) \\ &\leq |\tilde{\Lambda}(\varepsilon, \boldsymbol{\gamma})| \exp(R\Upsilon(\tau)). \end{aligned}$$

In view of the analysis above of $\Upsilon(t)$, we get

$$(21) \quad |A(t, \varepsilon)| \leq |\tilde{\Lambda}(\varepsilon, \boldsymbol{\gamma})| \exp \left(R \left[\Lambda_0^\circ(\tau) + \sup_{u \in [0, \tau]} |M^*(u)| \right] \right).$$

Putting (18), (20) and (21) together, we get

$$(22) \quad \begin{aligned} &|\tilde{\Lambda}_0(t, \boldsymbol{\gamma}) - \tilde{\Lambda}_0(t, \boldsymbol{\gamma}, \varepsilon)| \\ &\leq R_1 \left[\Lambda_0^\circ(\varepsilon) + \sup_{u \in [0, \tau]} |M^*(u)| \right] \left(1 + \exp \left(R_2 \left[1 + \sup_{u \in [0, \tau]} |M^*(u)| \right] \right) \right) \end{aligned}$$

for suitable absolute constants R_1 and R_2 .

The last main step is to deal with the martingale process

$$M^*(u) = \int_0^u \Pi(s)^{-1} \left[\frac{1}{n} \sum_{i=1}^{2n} I(T_{i0} < s) \sum_{j=1}^{m_i} dM_{ij}(s) \right].$$

By an argument using Lenglart’s and Markov’s inequalities, as in [21] (page 595), we obtain (with ξ_1 as in Assumption 6)

$$(23) \quad \Pr\left(\sup_{u \in [0, c]} |M^*(u)| > \kappa n^{-1/2}\right) \leq \eta + R\kappa^2\eta^{-1}\xi_1(c).$$

Given (22) and (23), we have control over the first term of (17), and the proof is thus complete. \square

CLAIM A2. $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$ converges in probability to some function $\Lambda_0(t, \boldsymbol{\gamma})$ uniformly in t and $\boldsymbol{\gamma}$. The function $\Lambda_0(t, \boldsymbol{\gamma})$ satisfies $\Lambda_0(t, \boldsymbol{\gamma}^\circ) = \Lambda_0^\circ(t)$.

PROOF. We can write $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$ as

$$\hat{\Lambda}_0(t, \boldsymbol{\gamma}) = \int_0^t \frac{n^{-1} \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dN_{ij}(s)}{n^{-1} \sum_{i=1}^{2n} \tilde{\psi}_i(s-, \boldsymbol{\gamma}) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})}.$$

In view of Claim A1 above, up to a uniform error of $o_P(1)$ we can replace all instances of $\tilde{\Lambda}_0(u, \boldsymbol{\gamma})$ in the definition of $\tilde{\psi}_i(s-, \boldsymbol{\gamma})$ by $\Lambda_0^*(u, \boldsymbol{\gamma})$. The desired result then can be obtained using the argument used to prove Claim A of [36]. \square

CLAIM A3. We have

$$\begin{aligned} \sup_{s \in [0, \tau], \boldsymbol{\gamma} \in \mathcal{G}} |\tilde{\Lambda}_0(s, \boldsymbol{\gamma}) - \tilde{\Lambda}_0(s-, \boldsymbol{\gamma})| &\xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty, \\ \sup_{s \in [0, \tau], \boldsymbol{\gamma} \in \mathcal{G}} |\hat{\Lambda}_0(s, \boldsymbol{\gamma}) - \hat{\Lambda}_0(s-, \boldsymbol{\gamma})| &\xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

PROOF. By appeal to Claims A1 and A2, and to continuity of $\Lambda_0^*(t, \boldsymbol{\gamma})$ and $\Lambda_0(t, \boldsymbol{\gamma})$. \square

CLAIM B. $\mathbf{U}(\boldsymbol{\gamma}, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}))$ converges in probability uniformly in t and $\boldsymbol{\gamma}$ to a limit $\mathbf{u}(\boldsymbol{\gamma}, \Lambda_0(\cdot, \boldsymbol{\gamma}))$.

PROOF. As in Claim B of [36]. \square

CLAIM C. There exists a unique consistent (in pr.) root to $\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) = \mathbf{0}$.

PROOF. By appeal to Foutz’s theorem [8], as in Claim C of [36]. \square

9.3. A workable representation of $\hat{\Lambda}_0(t) - \Lambda_0^\circ(t)$. To develop our asymptotic normality result, we need a workable representation of $\hat{\Lambda}_0(t) - \Lambda_0^\circ(t)$. The first step is to develop a suitable representation of $\tilde{\Lambda}_0(t) - \Lambda_0^\circ(t)$. Then, building on this, we develop our representation of $\hat{\Lambda}_0(t) - \Lambda_0^\circ(t)$.

9.3.1. *Representation of $\tilde{\Lambda}_0(t) - \Lambda_0^\circ(t)$.* Our starting point is the following simple lemma.

LEMMA 3. *Let $\mathcal{R}_n(t)$ and $\mathcal{S}_n(t)$ be stochastic processes, and let $A_n(t, \varepsilon)$ and $B_n(t, \varepsilon)$ be quantities that are bounded in pr. uniformly in t and ε . Define $\mathcal{R}_n(t, \varepsilon) = \mathcal{R}_n(t) - A_n(t, \varepsilon)\mathcal{R}_n(\varepsilon)$ and $\mathcal{S}_n(t, \varepsilon) = B_n(t, \varepsilon)[\mathcal{S}_n(t) - \mathcal{S}_n(\varepsilon)]$. Suppose that:*

1. $\sup_{t \in [\varepsilon, \tau]} \sqrt{n} |\mathcal{R}_n(t, \varepsilon) - \mathcal{S}_n(t, \varepsilon)| \xrightarrow{P} 0$ as $n \rightarrow \infty$ for any fixed $\varepsilon > 0$.
2. $\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} \Pr(\sup_{t \in [0, \varepsilon]} \sqrt{n} |\mathcal{R}_n(t)| > \delta) = 0$ for all $\delta > 0$.
3. $\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} \Pr(\sup_{t \in [0, \varepsilon]} \sqrt{n} |\mathcal{S}_n(t)| > \delta) = 0$ for all $\delta > 0$.
4. $\lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, \tau]} |B_n(t, \varepsilon) - B_n(t, 0)| = 0$ with probability going to 1 as $n \rightarrow \infty$.

Then $\sup_{t \in [0, \tau]} \sqrt{n} |\mathcal{R}_n(t) - B_n(t, 0)\mathcal{S}_n(t, 0)| \xrightarrow{P} 0$.

We apply this lemma with $\mathcal{R}_n(t) = \sqrt{n}[\tilde{\Lambda}_0(t) - \Lambda_0^\circ(t)]$. We have to check the four conditions enumerated in the lemma.

Condition 1. Arguments along the lines of [36] yield the result of Condition 1, with

$$(24) \quad \mathcal{S}_n(t) = \int_0^t \frac{\tilde{p}(s-, \varepsilon)}{\tilde{y}(s, \Lambda_0^\circ)} \left[\frac{1}{n} \sum_{i=1}^{2n} \sum_{j=1}^{m_i} I(T_{i0} < s) dM_{ij}(s) \right]$$

and $A_n(t, \varepsilon) = B_n(t, \varepsilon) = \tilde{p}(t, \varepsilon)^{-1}$, where

$$\tilde{p}(t, \varepsilon) = \prod_{s \in [\varepsilon, t]} \left[1 + n^{-1} \sum_{i=1}^{2n} \sum_{j=0}^{m_i} \{ \Omega_{ij}(s, t) d\tilde{N}_{ij}(s) + \Omega^*(s) I(T_{i0} < s) \delta_{ij} \} \right].$$

Here

$$\tilde{y}(s, \Lambda) = \frac{1}{n} \sum_{i=1}^{2n} I(T_{i0} < s) \psi_i(\boldsymbol{\gamma}^\circ, \Lambda, s) R_i.(s),$$

$$R_i.(s) = \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{ij}),$$

$$\Omega^*(s) = \frac{1}{n} \sum_{k=1}^{2n} \frac{R_k.(s) \eta_{1k}(s) I(T_{k0} < s)}{\{\tilde{y}(s, \Lambda_0^\circ)\}^2} \sum_{l=1}^{m_k} I(T_{kl} > s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{kl}),$$

$$\Omega_{i0}(s, t) = \frac{1}{n} \int_s^t \frac{R_i.(u) \eta_{1i}(u) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{i0})}{\{\tilde{y}(u, \Lambda_0^\circ)\}^2} \sum_{k=1}^{2n} \sum_{l=1}^{m_k} dN_{kl}(u),$$

$$\Omega_{ij}(s, t) = \frac{1}{n} \int_s^t \frac{I(T_{i0} < u) R_i(u) \eta_{1i}(u) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{ij})}{\{\tilde{\Psi}(u, \Lambda_0^\circ)\}^2} \times \sum_{k=1}^{2n} \sum_{l=1}^{m_k} I(T_{k0} < u) dN_{kl}(u), \quad j \geq 1.$$

In the above, $\eta_{1i}(s)$ is defined as

$$\eta_{1i}(s) = \frac{\phi_{3i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, s)}{\phi_{1i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, s)} - \left\{ \frac{\phi_{2i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, s)}{\phi_{1i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, s)} \right\}^2.$$

In Section 9.3.2 below, we present in detail a similar argument for $\hat{\Lambda}_0(t) - \Lambda_0^\circ(t)$.

Appealing to assumption 6 and using arguments similar to those used in the consistency proof, we find that the Ω quantities converge in probability uniformly in s and t , so that $\tilde{p}(t, \varepsilon)$ converges in probability to a deterministic limit uniformly in t and ε .

Conditions 2, 3 and 4. In regard to Condition 2, we have $\tilde{\Lambda}_0(t, \boldsymbol{\gamma}) - \Lambda_0^\circ(t) = \Delta_1(t) + \Delta_2(t)$, where

$$\Delta_1(t) = \int_0^t [\Gamma(s, \boldsymbol{\gamma}) - 1] \lambda_0^\circ(s) ds,$$

$$\Delta_2(t) = \int_0^t \frac{n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \sum_{j=1}^{m_i} dM_{ij}(s)}{n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \tilde{\Psi}_i(s-, \boldsymbol{\gamma}, \tilde{\Lambda}) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})},$$

where

$$\Gamma(s, \boldsymbol{\gamma}) = \frac{n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \psi_i(s-, \boldsymbol{\gamma}^\circ, \Lambda_0^\circ) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{ij})}{n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \tilde{\Psi}_i(s-, \boldsymbol{\gamma}, \tilde{\Lambda}) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})}$$

and $M_{ij}(t)$ is defined as in (19). We will deal with $\Delta_1(t)$ and $\Delta_2(t)$ in turn, starting with $\Delta_2(t)$. In the development below, R denotes a “generic” absolute constant.

The quadratic variation process of $\Delta_2(t)$ is given by

$$\langle \Delta_2 \rangle(t) = \int_0^t \left[\frac{n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \psi_i(s-, \boldsymbol{\gamma}^\circ, \Lambda_0^\circ) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{ij})}{[n^{-1} \sum_{i=1}^{2n} I(T_{i0} < s) \tilde{\Psi}_i(s-, \boldsymbol{\gamma}, \tilde{\Lambda}) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})]^2} \right] \times \lambda_0^\circ(s) ds.$$

By arguments similar to those of [21] (page 595), we find that $E[n \langle \Delta_2 \rangle(t)] \leq R \xi_1(t)$. An application of Lenglar’s inequality then gives

$$\Pr\left(\sqrt{n} \sup_{t \in [0, \varepsilon]} |\Delta_2(t)| > \kappa\right) \leq \eta + R \kappa^2 \eta^{-1} \xi_1(\varepsilon) \quad \forall \eta > 0.$$

Assumption 6 implies that $\xi_1(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$, and this takes care of $\Delta_2(t)$.

We now turn to $\Delta_1(t)$. Denote $J(s) = I(\Pi(s) > 0)$. We can write $\Delta_1(t) = \Delta_{1a}(t) + \Delta_{1b}(t)$, with

$$\Delta_{1a}(t) = \int_0^t [\Gamma(s, \boldsymbol{\gamma}) - 1]J(s)\lambda_0^\circ(s) ds$$

and

$$\Delta_{1b}(t) = \int_0^t [J(s) - 1]\lambda_0^\circ(s) ds.$$

The term $\Delta_{1b}(t)$ can be shown to be uniformly $O_p(n^{-1/2})$ by the argument in the middle of page 595 in [21]. As for $\Delta_{1a}(t)$, we have

$$\begin{aligned} \Delta_{1a}(t) &\leq Rt|\tilde{\Lambda}_0(t, \boldsymbol{\gamma}) - \Lambda_0^\circ(t)| \leq Rt|\Delta_1(t)| + Rt|\Delta_2(t)| \\ &\leq Rt|\Delta_{1a}(t)| + Rt|\Delta_{1b}(t)| + Rt|\Delta_2(t)|. \end{aligned}$$

Thus, for t small, $|\Delta_{1a}(t)| \leq [Rt/(1 - Rt)][\Delta_{1b}(t) + \Delta_2(t)]$, and the terms on the right hand side have already been taken care of.

The proof of Condition 3 is similar to that given above for $\Delta_2(t)$. Condition 4 follows easily from the uniform convergence of the Ω quantities.

9.3.2. Representation of $\hat{\Lambda}_0(t) - \Lambda_0^\circ(t)$. Define

$$\mathcal{Y}(s, \{\tilde{\Lambda}_0, \hat{\Lambda}_0\}) = \frac{1}{n} \sum_{i=1}^{2n} \tilde{\psi}_i(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0, s)R_i.(s)$$

and

$$\mathcal{Y}(s, \Lambda) = \frac{1}{n} \sum_{i=1}^{2n} \psi_i(\boldsymbol{\gamma}^\circ, \Lambda, s)R_i.(s),$$

where in $\tilde{\psi}_i(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0, s)$ we take $\tilde{\Lambda}_0(T_{i0})$ if $T_{i0} \geq s$ and $\hat{\Lambda}_0(T_{ij})$ if $T_{ij} < s, j \geq 0$. By Claim A3, we have $\sup_{s \in [0, \tau]} |\hat{\Lambda}_0(s, \boldsymbol{\gamma}^\circ) - \hat{\Lambda}_0(s-, \boldsymbol{\gamma}^\circ)|$ converges to zero. Thus, we obtain the following approximation, uniformly over $t \in [0, \tau]$:

$$\begin{aligned} &\hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t) \\ &\approx \frac{1}{n} \int_0^t \frac{\sum_{i=1}^{2n} \sum_{j=1}^{m_i} dM_{ij}(s)}{\mathcal{Y}(s, \Lambda_0^\circ)} \\ &\quad + \frac{1}{n} \int_0^t [\{\mathcal{Y}(s, \{\tilde{\Lambda}_0, \hat{\Lambda}_0\})\}^{-1} - \{\mathcal{Y}(s, \Lambda_0^\circ)\}^{-1}] \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dN_{ij}(s). \end{aligned}$$

Now let $\mathcal{X}(s, r) = \{\mathcal{Y}(s, \Lambda_0^\circ + r\Delta^*)\}^{-1}$ with $\Delta^* = \hat{\Lambda}_0 - \Lambda_0^\circ$ or $\tilde{\Lambda}_0 - \Lambda_0^\circ$, according to the estimator being used. Define $\dot{\mathcal{X}}$ and $\ddot{\mathcal{X}}$ as the first and second derivative of \mathcal{X} with respect to r , respectively. Then, by a first-order Taylor expansion of $\mathcal{X}(s, r)$

we get

$$\begin{aligned} & \hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t) \\ & \approx n^{-1} \int_0^t \{\boldsymbol{y}(s, \Lambda_0^\circ)\}^{-1} \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dM_{ij}(s) \\ & - n^{-2} \int_0^t \sum_{k=1}^{2n} \frac{R_{k\cdot}(s)\eta_{1k}(s)}{\{\boldsymbol{y}(s, \Lambda_0^\circ)\}^2} \sum_{l=1}^{m_k} I(T_{kl} > s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{kl}) \{\hat{\Lambda}_0(s) - \Lambda_0^\circ(s)\} \\ & \quad \times \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dN_{ij}(s) \\ & - n^{-2} \int_0^t \sum_{k=1}^{2n} \frac{R_{k\cdot}(s)\eta_{1k}(s)}{\{\boldsymbol{y}(s, \Lambda_0^\circ)\}^2} \sum_{l=1}^{m_k} I(T_{kl} \leq s) \\ & \quad \times \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{kl}) \{\hat{\Lambda}_0(T_{kl}) - \Lambda_0^\circ(T_{kl})\} \\ & \quad \times \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dN_{ij}(s) \\ & - n^{-2} \int_0^t \sum_{k=1}^{2n} \frac{R_{k\cdot}(s)\eta_{1k}(s)}{\{\boldsymbol{y}(s, \Lambda_0^\circ)\}^2} I(T_{k0} \geq s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{k0}) \{\tilde{\Lambda}_0(T_{k0}) - \Lambda_0^\circ(T_{k0})\} \\ & \quad \times \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dN_{ij}(s) \\ & - n^{-2} \int_0^t \sum_{k=1}^{2n} \frac{R_{k\cdot}(s)\eta_{1k}(s)}{\{\boldsymbol{y}(s, \Lambda_0^\circ)\}^2} I(T_{k0} < s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{k0}) \{\hat{\Lambda}_0(T_{k0}) - \Lambda_0^\circ(T_{k0})\} \\ & \quad \times \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dN_{ij}(s). \end{aligned}$$

The justification for ignoring the remainder term in the Taylor expansion is as in the parallel argument in [36]. Note that in the above approximation probands' data are involved since the derivative involves an estimator of Λ_0 for the probands (either $\hat{\Lambda}_0$ or $\tilde{\Lambda}_0$) and not only an estimator for the relatives.

The second, third and fifth terms of the above equation can be written, by interchanging the order of integration, as

$$-n^{-1} \int_0^t \{\hat{\Lambda}_0(s) - \Lambda_0^\circ(s)\} \sum_{i=1}^{2n} \sum_{j=0}^{m_i} \Upsilon_{ij}(s, t) d\tilde{N}_{ij}(s),$$

where $\tilde{N}_{ij}(s) = I(T_{ij} \leq t)$,

$$\Upsilon_{i0}(s, t) = n^{-1} \int_s^t \frac{R_{i\cdot}(u)\eta_{1i}(u) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{i0})}{\{\boldsymbol{y}(u, \Lambda_0^\circ)\}^2} \sum_{k=1}^{2n} \sum_{l=1}^{m_k} dN_{kl}(u)$$

and for $j \geq 1$

$$\begin{aligned} \Upsilon_{ij}(s, t) &= n^{-1} \int_s^t \frac{R_{i\cdot}(u)\eta_{1i}(u) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{ij})}{\{\mathcal{Y}(u, \Lambda_0^\circ)\}^2} \sum_{k=1}^{2n} \sum_{l=1}^{m_k} dN_{kl}(u) \\ &\quad + n^{-1} \sum_{k=1}^{2n} \frac{R_{k\cdot}(s)\eta_{1k}(s)}{\{\mathcal{Y}(s, \Lambda_0^\circ)\}^2} \sum_{l=1}^{m_k} I(T_{kl} > s) \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{kl}) \delta_{kl}. \end{aligned}$$

The fourth term can be written, by plugging in the representation for $\tilde{\Lambda}_0 - \Lambda_0^\circ$, as

$$-n^{-1} \int_0^\tau \frac{A(s, t) \tilde{p}(s-)}{\tilde{\mathcal{Y}}(s, \Lambda_0^\circ)} \sum_{i=1}^{2n} \sum_{j=1}^{m_i} I(T_{i0} < s) dM_{ij}(s),$$

where

$$\begin{aligned} A(s, t) &= n^{-2} \int_0^t \sum_{k=1}^{2n} \frac{R_{k\cdot}(s)\eta_{1k}(s)}{\{\mathcal{Y}(s, \Lambda_0^\circ)\}^2} \exp(\boldsymbol{\beta}^{\circ T} \mathbf{Z}_{k0}) \\ &\quad \times \left[\int_s^\tau \{\tilde{p}(v)\}^{-1} dN_{k0}^*(v) \right] \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dN_{ij}(s) \end{aligned}$$

and $N_{k0}^*(t) = I(T_{k0} \leq t)$. Given all the above, we get

$$\begin{aligned} \hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t) &\approx n^{-1} \int_0^t \{\mathcal{Y}(s, \Lambda_0^\circ)\}^{-1} \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dM_{ij}(s) \\ &\quad - n^{-1} \int_0^\tau \frac{A(s, t) \tilde{p}(s-)}{\tilde{\mathcal{Y}}(s, \Lambda_0^\circ)} \sum_{i=1}^{2n} \sum_{j=1}^{m_i} I(T_{i0} < s) dM_{ij}(s) \\ &\quad - n^{-1} \int_0^t \{\hat{\Lambda}_0(s) - \Lambda_0^\circ(s)\} \sum_{i=1}^{2n} \sum_{j=0}^{m_i} \Upsilon_{ij}(s, t) d\tilde{N}_{ij}(s). \end{aligned}$$

By solving the above approximation recursively, for the relatives' failure times, we get

$$\begin{aligned} &\hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t) \\ &\approx \frac{1}{n \hat{p}(t)} \int_0^t \frac{\hat{p}(s-)}{\mathcal{Y}(s, \Lambda_0^\circ)} \sum_{i=1}^{2n} \sum_{j=1}^{m_i} dM_{ij}(s) \\ (25) \quad &+ \frac{1}{n \hat{p}(t)} \int_0^\tau B(s, t) \sum_{i=1}^{2n} \sum_{j=1}^{m_i} \frac{\tilde{p}(s-)}{\tilde{\mathcal{Y}}(s, \Lambda_0^\circ)} I(T_{i0} < s) dM_{ij}(s) \\ &- \frac{\hat{p}(t-) dN(t)}{n^2 \hat{p}(t)} \int_0^\tau A(s, t) \sum_{i=1}^{2n} \sum_{j=1}^{m_i} \frac{\tilde{p}(s-)}{\tilde{\mathcal{Y}}(s, \Lambda_0^\circ)} I(T_{i0} < s) dM_{ij}(s), \end{aligned}$$

where $N(t) = \sum_{i=1}^{2n} \sum_{j=1}^{m_i} N_{ij}(t)$,

$$B(s, t) = n^{-1} \int_0^{t-} A(s, u) \hat{p}(u-) \sum_{i=1}^{2n} \sum_{j=0}^{m_i} \Upsilon_{ij}(u, t-) dN_{ij}(u)$$

and

$$\hat{p}(t) = \prod_{s \leq t} \left[1 + \frac{1}{n} \sum_{i=1}^{2n} \sum_{j=0}^{m_i} \Upsilon_{ij}(s, t) d\tilde{N}_{ij}(s) \right].$$

9.4. *Asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)$.* To show that $\hat{\boldsymbol{\gamma}}$ is asymptotically normally distributed, we expand $\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) = \mathbf{0}$ as

$$\begin{aligned} &\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) + [\mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)] \\ &\quad + [\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ))] = \mathbf{0}. \end{aligned}$$

We examine in turn each of the terms on the left-hand side of the above equation.

Step I. We can write $\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) = n^{-1}(\sum_{i=1}^n \boldsymbol{\xi}_i^{(1)} + \sum_{i=1}^{2n} \boldsymbol{\xi}_i^{(2)})$, where $\boldsymbol{\xi}_i^{(1)}$, $i = 1, \dots, n$, are i.i.d. mean-zero random $(p + 1)$ -vectors stemming from the likelihood of the proband data, while $\boldsymbol{\xi}_i^{(2)}$, $i = 1, \dots, 2n$, are i.i.d. mean-zero random $(p + 1)$ -vectors stemming from the likelihood of the relatives' data. It follows immediately from the classical central limit theorem that $n^{-1/2}\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)$ is asymptotically mean-zero multivariate normal.

Step II. Let $\hat{U}_r = U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0)$, $r = 1, \dots, p$ and $\hat{U}_{p+1} = U_{p+1}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0)$ (in this segment of the proof, when we write $(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0)$ the intent is to signify $(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ))$). Further, denote $Q_{ijr}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, T_{ij}) = [\partial U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) / \partial \Lambda_0^\circ(T_{ij})]$, $i = 1, \dots, 2n$, $j = 0, \dots, m_i$, $r = 1, \dots, p + 1$. First-order Taylor expansion of \hat{U}_r about Λ_0° , $r = 1, \dots, p + 1$, then gives

$$\begin{aligned} &n^{1/2}\{U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0) - U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)\} \\ &= n^{-1/2} \sum_{i=1}^{2n} \sum_{j=0}^{m_i} Q_{ijr}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, T_{ij}) \{\hat{\Lambda}_0(T_{ij}, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(T_{ij})\} + o_p(1) \\ &= \frac{1}{n} \sum_{i=1}^{2n} \sum_{j=0}^{m_i} \int_0^\tau Q_{ijr}(\boldsymbol{\gamma}^\circ, \Lambda^\circ, s) \{\hat{\Lambda}_0(s, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(s)\} d\tilde{N}_{ij}(s) + o_p(1). \end{aligned}$$

Although probands are involved in the above stochastic integral, its integrand is predictable since, by definitions (10) and (11), $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$ depends only on data up to time $t-$. Using the representation in Section 9.3.2 for $\sqrt{n}[\hat{\Lambda}_0(s, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(s)]$ and replacing certain empirical sums by their limiting values (see the expanded

paper for details), we obtain a representation of the form

$$\begin{aligned}
 &U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0) - U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) \\
 &\approx \frac{1}{n} \sum_{k=1}^{2n} \sum_{l=1}^{m_i} \int_0^\tau [\alpha_r^{(1)}(u) + I(T_{k0} < u) \{\alpha_r^{(2)}(u) - \alpha_r^{(3)}(u)\}] dM_{kl}(u).
 \end{aligned}$$

Thus, we have represented $U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0) - U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)$ $r = 1, \dots, p + 1$ as the average of mean zero i.i.d. random variables. Hence, asymptotic normality follows from the classical central limit theorem.

Step III. First-order Taylor expansion of $\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}}))$ about $\boldsymbol{\gamma}^\circ = (\boldsymbol{\beta}^{\circ T}, \theta^\circ)^T$ gives

$$\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) = \mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) + \mathbf{D}(\boldsymbol{\gamma}^\circ)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)^T + o_p(1),$$

where $D_{ls}(\boldsymbol{\gamma}) = \partial U_l(\boldsymbol{\gamma}, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma})) / \partial \gamma_s$ for $l, s = 1, \dots, p + 1$.

Combining the results of Steps I–III above we get that $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)$ is asymptotically zero-mean normally distributed with a covariance matrix that can be consistently estimated by a sandwich-type estimator.

9.5. *Asymptotic properties of $\hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})$.* We can write

$$\begin{aligned}
 (26) \quad &\sqrt{n}\{\hat{\Lambda}_0(t, \hat{\boldsymbol{\gamma}}) - \Lambda_0^\circ(t)\} \\
 &= \sqrt{n}\{\hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t)\} + \sqrt{n}\{\hat{\Lambda}_0(t, \hat{\boldsymbol{\gamma}}) - \hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ)\}.
 \end{aligned}$$

In (25), we have a representation of the first term above in terms of integrals with respect to the martingale processes M_{ij} . Weak convergence of the first term can thus be established using the martingale central limit theorem, as in [1]. In particular, the first term is tight. In regard to the second term, Taylor expansion yields

$$\hat{\Lambda}_0(t, \hat{\boldsymbol{\gamma}}) - \hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) = W(t, \boldsymbol{\gamma}^\circ)^T (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ) + o_p(1),$$

where

$$\begin{aligned}
 W(t, \boldsymbol{\gamma}) &= \int_0^t n^{-1} \sum_{i=1}^{2n} \sum_{j=1}^{m_i} \left\{ \frac{\partial}{\partial \boldsymbol{\gamma}} \tilde{\psi}_i(u-, \boldsymbol{\gamma}) + \tilde{\psi}_i(u-, \boldsymbol{\gamma}) \mathbf{Z}_{ij} \right\} Y_{ij}(u) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) \\
 &\quad \times \left(n^{-1} \sum_{i=1}^{2n} \tilde{\psi}_i(u-, \boldsymbol{\gamma}) R_i(u) \right)^{-1} d\hat{\Lambda}_0(u, \boldsymbol{\gamma}).
 \end{aligned}$$

The limiting value of $W(t, \boldsymbol{\gamma})$ is

$$\begin{aligned}
 w(t, \boldsymbol{\gamma}) &= \int_0^t \mathbf{E} \left[n^{-1} \sum_{i=1}^{2n} \sum_{j=1}^{m_i} \left\{ \frac{\partial}{\partial \boldsymbol{\gamma}} \tilde{\psi}_i(u-, \boldsymbol{\gamma}) + \tilde{\psi}_i(u-, \boldsymbol{\gamma}) \mathbf{Z}_{ij} \right\} Y_{ij}(u) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) \right] \\
 &\quad \times \left(\mathbf{E} \left[n^{-1} \sum_{i=1}^{2n} \tilde{\psi}_i(u-, \boldsymbol{\gamma}) R_i(u) \right] \right)^{-1} \lambda_0^\circ(u) du.
 \end{aligned}$$

Since the integrand is bounded, the function $w(t, \boldsymbol{\gamma})$ is Lipschitz in t . We just showed that $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)$ converges to a mean-zero normal variate. Hence the second term in (26) is tight. Accordingly, the entire expression (26) is tight.

Now, as seen in the normality proof of Section 9.4, both terms in (26) can be represented in terms of i.i.d. sums over i of functions of the data on family i . Hence, asymptotic normality of the finite dimensional distributions of the entire expression (26) follows from the classical central limit theorem. This, together with the tightness just shown, establishes weak convergence of $\sqrt{n}\{\hat{\Lambda}_0(t, \hat{\boldsymbol{\gamma}}) - \Lambda_0^\circ(t)\}$ to a Gaussian process. A fortiori,

$$\sup_t |\hat{\Lambda}_0(t, \hat{\boldsymbol{\gamma}}) - \Lambda_0^\circ(t)| = O_p(n^{-1/2}).$$

Acknowledgment. The authors would like to thank Dr. Kathleen Malone for sharing the data from the case-control family study of breast cancer, which motivated the development of this work.

REFERENCES

- [1] ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. [MR0673646](#)
- [2] AALEN, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3** 15–27. [MR0400529](#)
- [3] AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726. [MR0491547](#)
- [4] BRESLOW, N. E. and DAY, N. E. (1980). *Statistical Methods in Cancer Research: Vol. 1: The Analysis of Case-Control Studies*. IARC Scientific Publication, Lyon, France.
- [5] CHEN, M. C. and BANDEEN-ROCHE, K. (2005). A diagnostic for association in bivariate survival models. *Lifetime Data Anal.* **11** 245–264. [MR2158784](#)
- [6] CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65** 141–151. [MR0501698](#)
- [7] CODITZ, G. A., ROSNER, B. A. and SPEIZER, F. E. (1996). Risk factors for breast cancer according to family history of breast cancer. For the Nurses' Health Study Research Group. *Journal of National Cancer Institute* **88** 365–371.
- [8] FOUTZ, R. V. (1977). On the unique consistent solution to the likelihood equation. *J. Amer. Statist. Assoc.* **72** 147–148. [MR0445686](#)
- [9] GILL, R. D. (1985). Discussion of the paper by D. Clayton and J. Cuzick. *J. Roy. Statist. Soc. Ser. A* **148** 108–109.
- [10] GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. *Scand. J. Statist.* **16** 97–128. [MR1028971](#)
- [11] GILL, R. D. (1992). Marginal partial likelihood. *Scand. J. Statist.* **19** 133–137. [MR1173595](#)
- [12] GLIDDEN, D. V. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika* **86** 381–393. [MR1705406](#)
- [13] GLIDDEN, D. V. (2007). Pairwise dependence diagnostics for clustered failure time data. *Biometrika* **94** 371–385. [MR2380566](#)
- [14] 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** 735–741. [MR2261454](#)

- [15] HARTMAN, P. (1973). *Ordinary Differential Equations*, 2nd ed. Birkhäuser, Boston, MA. [MR0658490](#)
- [16] HOPPER, J. L. (2003). Commentary: Case-control-family design: A paradigm for future epidemiology research? *International Journal of Epidemiology* **32** 48–50.
- [17] HOUGAARD, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York. [MR1777022](#)
- [18] HSU, L., CHEN, L., GORFINE, M. and MALONE, K. (2004). Semiparametric estimation of marginal hazard function from case-control family studies. *Biometrics* **60** 936–944. [MR2133546](#)
- [19] HSU, L. and GORFINE, M. (2006). Multivariate survival analysis for case-control family data. *Biostatistics* **7** 387–398.
- [20] HSU, L., GORFINE, M. and MALONE, K. (2007). On robustness of marginal regression coefficient estimates and hazard functions in multivariate survival analysis of family data when the frailty distribution is misspecified. *Stat. Med.* **26** 4657–4678.
- [21] KEIDING, N. and GILL, R. (1990). Random truncation models and Markov processes. *Ann. Statist.* **18** 582–602. [MR1056328](#)
- [22] KOSOROK, M. R., LEE, B. L. and FINE, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *Ann. Statist.* **32** 1448–1491. [MR2089130](#)
- [23] MALONE, K. E., DALING, J. R., THOMPSON, J. D., CECILIA, A. O. FRANCISCO, L. V. and OSTRANDER E. A. (1998). BRCA1 mutations and breast cancer in the general population. *Journal of the American Medical Association* **279** 922–929.
- [24] MALONE, K. E., DALING, J. R., NEAL, C., SUTER, N. M., O'BRIEN, C., CUSHING-HAUGEN, K., JONASDOTTIR, T. J., THOMPSON, J. D. and OSTRANDER E. A. (2000). Frequency of BRCA1/BRCA2 mutations in a population-based sample of young breast carcinoma cases. *Cancer* **88** 1393–1402.
- [25] MALONE, K. M., DALING, J. R., DOODY, D. R., HSU, L., BERNSTEIN, L., COATES, R. J., MARCHBANKS, P. A., SIMON, M. S., McDONALD, J. A., NORMAN, S. A., STROM, B. L., BURKMAN, R. T., URSIN, G., DEAPEN, D., WEISS, L. K., FOLGER, S., MADEOY, J. J., FRIEDRICHSEN, D. M., SUTER, N. M., HUMPHREY, M. C., SPIRTAS, R. and OSTRANDER, E. A. (2006). Prevalence and predictors of BRCA1 and BRCA2 mutations in a population-based study of breast cancer in white and black American women aged 35–64 years. *Cancer Research* **16** 8297–8308.
- [26] NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. and SØRENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation of frailty models. *Scand. J. Statist.* **19** 25–43. [MR1172965](#)
- [27] OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.* **84** 487–493. [MR1010337](#)
- [28] PARNER, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26** 183–214. [MR1611788](#)
- [29] PRENTICE, R. L. and BRESLOW, N. E. (1978). Retrospective studies and failure time models. *Biometrika* **65** 153–158.
- [30] SHIH, J. H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika* **85** 189–200. [MR1627281](#)
- [31] SHIH, J. H. and CHATTERJEE, N. (2002). Analysis of survival data from case-control family studies. *Biometrics* **58** 502–509. [MR1925547](#)
- [32] SHIH, J. H. and LOUIS, T. A. (1995). Inference on the association parameter in copula models for bivariate survival data. *Biometrics* **51** 1384–1399. [MR1381050](#)
- [33] VISWANATHAN, B. and MANATUNGA, A. K. (2001). Diagnostic plots for assessing the frailty distribution in multivariate survival data. *Lifetime Data Anal.* **7** 143–155. [MR1842324](#)
- [34] ZEGER, S., LIANG, K.-Y. and ALBERT, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44** 1049–1060. [MR0980999](#)

- [35] ZUCKER, D. M. (2005). A pseudo-partial likelihood method for semiparametric survival regression with covariate errors. *J. Amer. Statist. Assoc.* **100** 1264–1277. [MR2236440](#)
- [36] 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. [MR2406421](#)

M. GORFINE
FACULTY OF INDUSTRIAL ENGINEERING
AND MANAGEMENT
TECHNION CITY
HAIFA 32000
ISRAEL
E-MAIL: gorfinm@ie.technion.ac.il

D. M. ZUCKER
DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY, MT. SCOPUS
JERUSALEM 91905
ISRAEL
E-MAIL: mszucker@mscc.huji.ac.il

L. HSU
DIVISION OF PUBLIC HEALTH SCIENCES
FRED HUTCHINSON CANCER RESEARCH CENTER
SEATTLE, WASHINGTON 98109-1024
USA
E-MAIL: lih@fhcrc.org