Semiparametric Failure Time Regression With Replicates of Mismeasured Covariates

Chengcheng Hu and D. Y. Lin

1. INTRODUCTION

Covariate measurement error often presents a serious challenge in the regression analysis of censored failure time data. Sometimes the true covariate is measured only on a fraction of the study subjects, called a validation set, whereas a surrogate covariate, that is, an error-prone version of the true covariate, is measured on every subject. More commonly, the true covariate cannot be measured precisely, and only replicate measurements of the surrogate are available.

A good example of the latter situation is the Nutritional Prevention of Cancer (NPC) trial (Clark et al. 1996), which investigated the effectiveness of the food supplement of selenium (Se) in preventing two types of skin cancers, basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). The baseline plasma Se level, an important prognostic risk factor for the cancers, can never be measured precisely due to its temporal biological fluctuation and imprecision of the instrument. Fortunately, each patient had the plasma Se measured every 6 months during the NPC study. Because this variable is quite stable over a long period of time, for each untreated patient all the measurements can be regarded as replicates of the baseline value, providing opportunities to explore the measurement error. For patients in the active treatment arm (i.e., the Se group), the serum Se level was altered by the treatment, which is Se itself. Thus, only one measurement of the baseline Se level is available for each patient in the Se group. Because each subject could potentially experience both types of skin cancers, each of which could have multiple episodes, statistical methods for handling multivariate failure time data with covariate measurement error would be required.

The problem of covariate measurement error in the analysis of univariate failure time data has been studied extensively. Prentice (1982) proposed an induced partial likelihood approach for rare events. Zhou and Pepe (1995) developed a consistent estimator for the regression parameter when the covariates are discrete and a validation set is available. The regression calibration method of Wang, Hu, Feng, and Prentice (1997) yielded approximately consistent estimators. Hu, Tsai, and Davidian (1998) described some likelihood-based methods with parametric assumptions on the covariates and/or the errors. Nakamura (1992) developed a corrected score approach for normally distributed measurement error, and the resultant estimator was shown to be consistent by Kong and Gu (1999). Buzas (1998) took a similar approach assuming that the moment generating function of the error distribution is known. For replicate measurements Huang and Wang (2000) developed a clever nonparametric correction method, which excludes subjects with only one surrogate measurement from the analysis. Hu and Lin (2002) extended the approach of Nakamura (1992) to provide a class of consistent estimators without parametric assumption on the covariate or the error. Their article focused on the situation when a validation set is available, and considered briefly the case where each subject has exactly two replicate measurements. None of the aforementioned methods can deal with the measurement error problem in the NPC trial (even if the multivariate nature of the failure time data is ignored).

There has been limited progress on covariate measurement error for multivariate failure time data. Jiang, Turnbull, and Clark (1999) developed a method for analyzing the recurrences of SCC in the NPC trial, using replicate measurements of the plasma Se level to adjust for the measurement error. This method is based on a discrete-time proportional means model (Lawless and Nadeau 1995) and imposes parametric assumptions on both the true covariate and the error, which can be checked empirically. Li and Lin (2000) proposed using the EM algorithm for handling covariate measurement error in frailty models for clustered survival data. They assumed that the true covariate, which is subject to error, is normally distributed conditional on other precisely measured covariates and that the measurement error is also normally distributed.

In this article we present a unified approach to dealing with replicates of mismeasured covariates for both univariate and multivariate failure time data. The study subjects are allowed to have varying numbers of surrogate measurements, and the...
subjects with a single measurement are naturally included in
the analysis. The regression models are semiparametric in that
the underlying failure time distribution is completely unspecified.
Furthermore, no parametric assumption is imposed on the
covariate, and the error only needs to be symmetrically distrib-
uted. We consider univariate failure time and recurrent events
data in Section 2, and study general multivariate failure time
data in Section 3. In Section 4 we use simulation studies to
compare the proposed estimator with that of Huang and Wang
(2000) for univariate failure time data as well as to assess the
operating characteristics of the proposed method for recurrent
events data. We apply the proposed method to the afore-
mentioned NPC trial in Section 5, and provide some concluding
remarks in Section 6. Most of the technical details are relegated
to the Appendix.

2. UNIVARIATE FAILURE TIME AND RECURRENT
EVENTS DATA

2.1 Preliminaries

We adopt the framework of Andersen and Gill (1982), which
is suitable for both univariate failure time data and recurrent
events data. For $i = 1, \ldots, n$ and $t \in [0, \tau]$, let $N_i(t)$ be the
number of events observed on the $i$th subject by time $t$, let $Y_i(t)$ be
the indicator of whether the $i$th subject is under observation at
time $t$, and let $Z_i(t)$ be a vector of covariates. The multiplicity
intensity model specifies that the intensity process for $N_i$
associated with $Z_i$ takes the form

$$l(t; Z_i) = Y_i(t) \exp \{ \beta_0^T Z_i(t) \} \lambda_0(t),$$

where $\beta_0$ is a vector of unknown regression parameters and
$\lambda_0(t)$ is an arbitrary positive function (Andersen and Gill 1982).

For $k = 0, 1, 2$ let $S^{(k)}(\beta, t) = n^{-1} \sum_{i=1}^n Y_i(t) \exp \{ \beta^T Z_i(t) \} \times$
$Z_i(t)^{A_k}$, where $a^{A_0} = 1$, $a^{A_1} = a$, and $a^{A_2} = a^T$ for any vec-
tor $a$. When all components of $Z_i(t)$ can be measured precisely,
a consistent estimator $\hat{\beta}$ for $\beta_0$ can be obtained from the partial
likelihood score function

$$U(\beta) = \sum_{i=1}^n \int_0^\tau \left[ Z_i(t) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right] dN_i(t).$$

The Breslow estimator for $\Lambda_0(t)$ is $\int_0^t \Lambda_0(s) \, ds$
is

$$\tilde{\Lambda}_0(t) \equiv \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{nS^{(0)}(\beta, s)}.$$  

The asymptotic properties of these estimators have been es-
lished elegantly via the counting process martingale theory
(Andersen and Gill 1982).

The preceding framework covers two important types of failure
time data: classical survival data (i.e., univariate failure time
data) and recurrent events data. For the former let $T_i$ and $C_i$ be
the survival time and censoring time for the $i$th subject, and de-
fine $I_i = I(T_i \leq C_i)$, where $a \wedge b = \min(a, b)$ and $I(\cdot)$ is the indicator function. Then $N_i(t) = \delta_i I(T_i \leq t)$ and
$Y_i(t) = I(T_i \geq t)$. Model (1) is equivalent to the Cox (1972) pro-
homogeneous model. It is assumed that $C$ is independent of
$T$ conditional on $Z$. For recurrent events data let $N^*(t)$ be the number of
events that the subject has experienced by time $t$, and let $C$ be the cen-
soring time. Then $N_i(t) = N^*(t \wedge C)$ and $Y_i(t) = I(C \geq t)$. Let

$F^*_C$ be the $\sigma$-field generated by $\{N^*(s), Z(s) : 0 \leq s \leq t\}$. Then
model (1) implies that

$$E[dN^*(t)|F^*_C] = E[dN^*(t)|Z(t)] = \exp \{ \beta_0^T Z(t) \} \, d\Lambda_0(t).$$

(4)

The first equality in (4) implies that the occurrence of pre-
cious events does not affect the risk of future recurrences unless
such dependence is parameterized through appropriate time-
dependent covariates. By removing this restriction, Pepe and
Cai (1993), Lawless and Nadeau (1995), and Lin, Wei, Yang,
and Ying (2000) studied the more general proportional rates
model

$$E[dN^*(t)|Z(t)] \equiv \exp \{ \beta_0^T Z(t) \} \, d\Lambda_0(t).$$

(5)

Independent censoring is required in that $E[dN^*(t)|Z(t), C \geq t] = E[dN^*(t)|Z(t)]$. The aforementioned maximum partial
likelihood estimator $\hat{\beta}$ and the Breslow estimator $\hat{\Lambda}_0(t)$
remain consistent under model (5). However, the variance esti-
imators need to be adjusted to account for the potential depen-
dence of the recurrent event times within the same subject.
Furthermore, the theoretical development is more delicate be-
cause the martingale theory can no longer be applied.

Remark 1. The function $\Lambda_0(t)$ corresponds to the base-
line cumulative hazard function under the proportional hazards
model and to the baseline mean function of recurrent events un-
der the proportional rates model.

2.2 Inference With Covariate Measurement Error

We develop a unified approach to dealing with covariate mea-
surement error that is applicable to both the proportional haz-
rards model for univariate failure time data and the propor-
tional rates model for recurrent events. Write $Z_i(t) = (X_i^T, V_i(t)^T)^T$, where
the $p$ vector of time-independent covariates $X_i$ is mea-
sured with error, whereas the $q$ vector of possibly time-
dependent covariates $V_i(t)$ is measured precisely. According-
ly, let $\tilde{\beta}_0 = (\beta_0^T, \gamma_0^T)^T$, with $\alpha_0$ and $\gamma_0$ pertaining to $X_i$ and $V_i(t)$, re-
spectively. Suppose that $X_i$ is measured $n_i$ times ($n_i \geq 1$) by
the surrogates $W_{ir} \equiv X_i + \epsilon_{ir}$ ($r = 1, \ldots, n_i$), where the error
terms $\epsilon_{ir}$ ($i = 1, \ldots, n; r = 1, \ldots, n_i$) are iid with mean 0 and
are independent of $(X_i, Y_i(t), Z_i(t)$, and $n_i$. We make an addi-
tional assumption that the error distribution is symmetric; that
is, $\epsilon_{111} = -\epsilon_{111}$ have the same distribution. The number
of replicates $n_i$ is allowed to depend on $N_i(t)$, $Y_i(t)$, and $Z_i(t)$.

Write $\tilde{Z}_{ir}(t) = (W_{ir}^T, V_i(t)^T)^T$ ($r = 1, \ldots, n_i$). Although there are
no systematic differences between $W_{ir}$ and $X_i$, the replace-
ments of the $Z_i$'s involved in $U(\beta)$ with the $\tilde{Z}_{ir}$'s would not
yield a consistent estimator of $\beta_0$. The reason is that $S^{(0)}$
and $S^{(1)}$ involve the terms $e^{\beta^T Z_i}$ and $e^{\beta^T Z_i}$, which are nonlinear
in $Z_i$. We need to make appropriate corrections for $S^{(0)}$ and
$S^{(1)}$. For any $(p + q) \times \beta$ vector $\beta$, write $\beta = (\alpha^T, \gamma^T)^T$, where
$\alpha$ is of dimension $p$ and $\gamma$ of dimension $q$. Define $\eta_0(\alpha) =
E[\exp(\alpha^T \epsilon_{11}) \epsilon_{11}]$ ($k = 0, 1$) and $J = (\mathbf{1}_p \exp(\mathbf{0}_p \epsilon_{11}))^T$. It follows from
the independence of the error and the true covariate that

$$E[\exp \{ \beta^T \tilde{Z}_{ir}(t) \} | Z_i(t)] = \eta_0(\alpha) \exp \{ \beta^T Z_i(t) \},$$

$$E[\exp \{ \beta^T \tilde{Z}_{ir}(t) \} \tilde{Z}_{ir}(t) | Z_i(t)] = \eta_0(\alpha) \exp \{ \beta^T Z_i(t) | Z_i(t) \} + \exp \{ \beta^T Z_i(t) \} \mathbf{1}_q.$$
For $i = 1, \ldots, n$, let
\[ R_i^{(0)}(\beta, t) = n_i^{-1} \eta_0^{-1}(\alpha) \sum_{r=1}^{n_i} \exp\{\beta^T \tilde{Z}_r(t)\}. \]
\[ R_i^{(1)}(\beta, t) = n_i^{-1} \eta_0^{-1}(\alpha) \sum_{r=1}^{n_i} \exp\{\beta^T \tilde{Z}_r(t)\} \times \left[ \tilde{Z}_r(t) - \eta_0^{-1}(\alpha) \eta_1(\alpha) \right]. \]
Clearly, $E[R_i^{(0)}(\beta, t) | Z_i(t)] = \exp(\beta^T Z_i(t))$ and $E[R_i^{(1)}(\beta, t) | Z_i(t)] = \exp(\beta^T Z_i(t))Z_i(t)$. Thus, when $X_i$ is not available, we can use $R_i^{(0)}(\beta, t)$ and $R_i^{(1)}(\beta, t)$ as the respective substitutes for $\exp(\beta^T Z_i(t))$ and $\exp(\beta^T Z_i(t))Z_i(t)$.

We need to estimate $\eta_1(\alpha)$ ($k = 0, 1$). Because $W_1 - W_2 = \epsilon_1 - \epsilon_2$, and the errors are symmetric, we have
\[ E[\exp(\alpha^T (W_1 - W_2))] = n_0(\alpha), \]
\[ E[\exp(\alpha^T (W_1 - W_2)) | (W_1, W_2)] = 2n_0(\alpha) \eta_1(\alpha). \]
Let $\xi_l = I(n_l > 1)$ and $n_0 = \sum_{l=1}^{n_i} \xi_l$. Given $\alpha$, we can estimate $\eta_0(\alpha)$ ($k = 0, 1$) by
\[ \hat{\eta}_0(\alpha) \equiv \left[ n_0^{-1} \sum_{l=1}^{n_0} \frac{\xi_l}{n_l(n_l - 1)} \sum_{r \neq l} \exp(\alpha^T (W_{l} - W_{l})) \right]^{1/2}, \]
\[ \hat{\eta}_1(\alpha) \equiv \left[ 2n_0 \hat{\eta}_0(\alpha) \right]^{-1} \sum_{l=1}^{n_0} \frac{\xi_l}{n_l(n_l - 1)} \sum_{r \neq l} \exp(\alpha^T (W_{l} - W_{l})). \]
where, for each $i$ with $n_i > 1$, $(r, s)$ runs through all possible pairs of numbers in $\{1, \ldots, n_i\}$. When $\xi_l = 0$, $n_l - 1$ is also 0, and we define the fraction $\xi_l/(n_l - 1)$ to be 0 as well. It is straightforward to show that $\hat{\eta}_0(\alpha) = \partial \hat{\eta}_0(\alpha) / \partial \alpha$. Although in estimating $\eta_0(\alpha)$ and $\eta_1(\alpha)$ we only use subjects with at least two measurements of the surrogate $W$, the estimators are consistent as long as the errors are identically distributed.

Now for $i = 1, \ldots, n$, let
\[ \hat{R}_i^{(0)}(\beta, t) = n_i^{-1} \hat{\eta}_0^{-1}(\alpha) \sum_{r=1}^{n_i} \exp(\beta^T \tilde{Z}_r(t)). \]
\[ \hat{R}_i^{(1)}(\beta, t) = n_i^{-1} \hat{\eta}_0^{-1}(\alpha) \sum_{r=1}^{n_i} \exp(\beta^T \tilde{Z}_r(t)) \times \left[ \tilde{Z}_r(t) - \hat{\eta}_0^{-1}(\alpha) \hat{\eta}_1(\alpha) \right]. \]
Define
\[ S_i^{(k)}(\beta, t) = n_i^{-1} \sum_{r=1}^{n_i} Y_r(t) \hat{R}_i^{(k)}(\beta, t) \quad (k = 0, 1) \]
and
\[ E_C(\beta, t) = S_i^{(1)}(\beta, t) / S_i^{(0)}(\beta, t). \]

To describe the asymptotic properties of our estimators, we define
\[ v_i = \int_0^t \left[ \tilde{Z}_r(t) - \eta_0(\alpha) \right] dN_i(t), \]
\[ r_i = \left[ 2\rho \hat{\eta}_0^{-1}(\alpha) \right]^{-1} \eta_0(\alpha) \eta_1(\alpha) \eta_0(\alpha) g_i(0), \]
where $\rho = \lim n_0/n$, $g_i(0) = \int_0^t \exp(\beta^T (W_{i} - W_{l})) dN_i(t)$, $\eta_0(\alpha) = \sum_{r \neq l} \exp(\alpha^T (W_{l} - W_{l}))$, and $g_i(0) = \sum_{r \neq l} \exp(\alpha^T (W_{l} - W_{l}))$.

Theorem 1. Under conditions 1–8 of the Appendix, $\beta_C$ exists and is unique in a neighborhood of $\beta_0$ with probability converging to 1 as $n \to \infty$, and $\beta_C \to \beta_0$. In addition,
\[ n^{1/2}(\beta_C - \beta_0) \xrightarrow{\mathcal{L}} N(0, \Gamma_\beta). \]

Theorem 2. Under conditions 1–8, $n^{1/2}(\hat{\lambda}_C(t) - \Lambda_0(t))$ converges weakly to a zero-mean Gaussian process with a covariance function defined in the Appendix. Furthermore,
\[ \sup_{t \in [0, T]} |\hat{\lambda}_C(t) - \Lambda_0(t)| \xrightarrow{P} 0. \]

The proofs of these two theorems are provided in the Appendix. To estimate $\Gamma_\beta$, we define
\[ \hat{v}_i = \int_0^t \left[ \tilde{Z}_r(t) - E_C(\hat{\beta}_C, t) \right] dN_i(t), \]
\[ \hat{r}_i = \left[ 2\hat{\eta}_0^{-1}(\alpha) \right]^{-1} \eta_0(\alpha) \eta_1(\alpha) \hat{g}_i(0), \]

where $\hat{\eta}_0(\alpha) = \sum_{r \neq l} \exp(\alpha^T (W_{l} - W_{l}))$, $\hat{\eta}_0(\alpha) = \sum_{r \neq l} \exp(\alpha^T (W_{l} - W_{l}))$, and $\hat{\eta}_0(\alpha) = \sum_{r \neq l} \exp(\alpha^T (W_{l} - W_{l}))$. Also, denote $\hat{\Gamma}_C = n^{-1} \sum_{r \neq l} (\hat{v}_i + \hat{r}_i)^2$ and $\hat{\Gamma}_C = n^{-1} \sum_{r \neq l} (S_i^{(2)}(\hat{\beta}_C, t) / S_i^{(0)}(\hat{\beta}_C, t) - E_C(\hat{\beta}_C, t))^2 dN_i(t)$, where $S_i^{(2)}(\beta, t) = \hat{\alpha} S_i^{(1)}(\beta, t) / \beta$. It is proved in the Appendix that $\Gamma_\beta$ is consistently estimated by $\hat{\Gamma}_\beta = \hat{\Gamma}_C$. Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
3. MULTIVARIATE FAILURE TIME DATA

In this section we consider the measurement error problem for general multivariate failure time data. Such data arise when each study subject can experience several events or when there exists natural or artificial clustering of study subjects such that the failure times of the subjects within the same cluster are correlated. The multiple events experienced by a subject may be of the same nature (e.g., recurrent events) or of different natures (e.g., different forms of cancers). Examples of clusters include families and communities. This kind of data has been previously considered by many authors, including Pren- tice, Williams, and Peterson (1981). Wei, Lin, and Weissfeld (1989) formulated the marginal distributions of the multivariate failure times through Cox proportional hazards models while leaving the nature of dependence among related failure times unspecified. A general formulation of this approach was provided by Spiekerman and Lin (1998): There are $n$ independent clusters (or subjects), each of which contains $L$ exchangeable failure times for each of $M$ different types. Here we consider a setup that is even more general than that of Spiekerman and Lin (1998) so as to cover clustered failure time data as well as multiple sequences of recurrent events.

For $i = 1, \ldots, n$, $m = 1, \ldots, M$, and $t = 1, \ldots, L$, let $X_{ilm}(t)$ be the $mt$th exchangeable counting process recording the $m$th type of failure or recurrent events in the $i$th cluster; let $Y_{ilm}(t)$ and $Z_{ilm}(t)$ be the corresponding at-risk process and covariate vector. This framework would be useful, for example, in modeling simultaneously the recurrences of BCC and SCC for the NPC trial, in which case $L = 1$ and $M = 2$, and each counting process can take multiple jumps. In another model where the first four recurrences of BCC and SCC of the same subject are considered and each recurrence is treated as a failure, we set $L = 1$ and $M = 8$. This model will be presented in Section 5.2. The marginal hazard function or rate function for $N_{ilm}(t)$ is assumed to take the form

$$
\lambda(t; Z_{ilm}) = \exp\{\beta_0^T Z_{ilm}(t)\} \lambda_{0m}(t),
$$

where $\beta_0$ is a vector of unknown regression parameters and $\lambda_{0m}(\cdot)$ is an arbitrary baseline hazard/rate function for the $mt$th type of failure/recurrence. Write $\Lambda_{0m}(t) = \int_0^t \lambda_{0m}(s) \, ds$ and

$$
S_m^{(k)}(\beta, t) = n^{-1} \sum_{i=1}^n \sum_{l=1}^L \lambda_{ilm}(t) e^{\beta^T Z_{ilm}(t)} Z_{ilm}(t)^{\delta_{ilm}} (k = 0, 1, 2).
$$

When all components of $Z_{ilm}$ are precisely measured, we may estimate $\beta_0$ by $\hat{\beta}$, the root of

$$
U(\beta) = \sum_{i=1}^n \sum_{m=1}^M \sum_{l=1}^L \int_0^t \left( Z_{ilm}(t) - \frac{S_m^{(1)}(\beta, t)}{S_m^{(0)}(\beta, t)} \right) dN_{ilm}(t).
$$

The Breslow-type estimators for $\Lambda_{0m}(t)$ are

$$
\tilde{\Lambda}_{0m}(t) = \sum_{i=1}^n \sum_{l=1}^L \int_0^t \frac{dN_{ilm}(s)}{n S_m^{(0)}(\beta, s)}, \quad m = 1, \ldots, M.
$$

By combining the arguments of Spiekerman and Lin (1998) with those of Lin et al. (2000), we can show that these estimators are consistent and asymptotically normal.

We shall develop a method to correct for the bias in parameter estimation when some components of the covariate vector are measured with error. Write $Z_{ilm}(t) = (X_{ilm}(t), Y_{ilm}(t)^T)^T$, where $X_{ilm}(t)$ is measured with error and $Y_{ilm}(t)$ is measured precisely. Suppose that $X_{ilm}(t)$ is measured $n_{ilm}$ times ($n_{ilm} \geq 1$) by the surrogates $W_{ilm} = X_{ilm} + e_{ilm}$ ($r = 1, \ldots, n_{ilm}$), where for each $m$ the error terms $e_{ilm} = i = 1, \ldots, n_i; l = 1, \ldots, L; r = 1, \ldots, n_{ilm}$ are i.i.d and symmetric with mean 0, and are independent of $N_{ilm}(\cdot), Y_{ilm}(\cdot), Z_{ilm}(\cdot)$, and $n_{ilm}$. The number of replicates $n_{ilm}$ is allowed to depend on $N_{ilm}(\cdot), Y_{ilm}(\cdot)$, and $Z_{ilm}(\cdot)$. Define $\tilde{Z}_{ilm}(t) = (W_{ilm}, Y_{ilm}(t)^T)^T$. Because $N_{ilm}$ is exchangeable for the same $m$, it is reasonable to assume that for the same $m$ all other random variables are also identically distributed across different $t$’s.

Remark 2. Model (6) is very flexible and can accommodate common or separate regression parameters for the $M$ types of failure or recurrence. In addition, the decomposition of $Z$ into $X$ and $V$ does not need to be the same among the $M$ failure types. Thus, our framework allows every type of event to have its own set of covariates with its unique error structure.

We partition the parameter vector $\beta$ into $\alpha$ and $\gamma$ according to the decomposition of $Z$. Define $\eta_{mk}(\alpha) = E(e^{\alpha^T E_{m}^{(k)}(\beta, t)} | e_{ilm}^{(k)})$ for $k = 0, 1, 2$ and $m = 1, \ldots, M$. Let $\xi_{ilm} = I(n_{ilm} > 1)$ and $n_{0m} = \sum_{i=1}^n \sum_{l=1}^L \xi_{ilm}$. We can estimate $\eta_{mk}(\alpha) (k = 0, 1; m = 1, \ldots, M)$ by

$$
\hat{\eta}_{mk}(\alpha) = \left( \frac{n_{0m}^{-1} \sum_{i=1}^n \sum_{l=1}^L \frac{\xi_{ilm}}{n_{ilm}(n_{ilm} - 1)} \exp[\alpha^T (W_{ilm} - W_{ilm})]} \right)^{1/2}.
$$

$$
\hat{\eta}_{m1}(\alpha) = \left\{ \frac{2n_{0m}\hat{\eta}_{mk}(\alpha)}{\hat{\eta}_{mk}(\alpha)} \right\}^{-1/2} \sum_{i=1}^n \sum_{l=1}^L \frac{\xi_{ilm}}{n_{ilm}(n_{ilm} - 1)} \times \sum_{r \neq s} (W_{ilm} - W_{ilm}) \exp[\alpha^T (W_{ilm} - W_{ilm})].
$$

Note that only the information from the $mt$th type of failure or recurrence is used in defining the estimators $\hat{\eta}_{mk}$ ($k = 0, 1$). When the errors are identically distributed for all $M$ different types, we can use all the data to calculate the estimates $\hat{\eta}_{01}$ and $\hat{\eta}_{11}$, and set $\hat{\eta}_{mk}(\alpha) = \hat{\eta}_{k}(\alpha)$.

For $k = 0, 1$, let $S_m^{(k)}(\beta, t) = n^{-1} \sum_{i=1}^n \sum_{l=1}^L \lambda_{ilm}(t) \times \tilde{R}_{ilm}^{(k)}(\beta, t)$, where

$$
\hat{R}_{ilm}^{(0)}(\beta, t) = n_{ilm}^{-1} \hat{\eta}_{01}(\alpha) \sum_{r=1}^{n_{ilm} - 1} e^{\beta^T \tilde{Z}_{ilm}(t)},
$$

$$
\hat{R}_{ilm}^{(1)}(\beta, t) = n_{ilm}^{-1} \hat{\eta}_{11}(\alpha) \sum_{r=1}^{n_{ilm} - 1} e^{\beta^T \tilde{Z}_{ilm}(t)} \times [\tilde{Z}_{ilm}(t) - \hat{\eta}_{mk}(\alpha) \tilde{V}_{ilm}(t)].
$$

We then propose the following estimating function for $\beta_0$:

$$
U_C(\beta) = \sum_{m=1}^M \sum_{l=1}^L \sum_{i=1}^n \int_0^t \left[ \tilde{Z}_{ilm}(t) - E_{ilm}(\beta, t) \right] dN_{ilm}(t).
$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
where $\hat{Z}_{int}(t) = n_{int}^{-1} \sum_{i=1}^{n_{int}} \tilde{Z}_{int}(t)$ and $\text{E}_{MC}(\beta, t) = S_{MC}^{(1)}(\beta, t) / S_{MC}^{(0)}(\beta, t)$. The resultant estimator is denoted by $\hat{\beta}_C$. We estimate $\Lambda_{int}(t)$ by

$$\hat{\Lambda}_{MC}(t) \equiv \frac{1}{n_{MC}} \sum_{i=1}^{n_{MC}} \sum_{t=1}^{T} \frac{dN_{int}(s)}{S_{MC}^{(0)}(\hat{\beta}_C, s)} \quad m = 1, \ldots, M.$$  

Let $s_n^{(k)}(\beta, t) = E S_n^{(k)}(\beta, t) \quad (k = 0, 1, 2)$ and $e_n(\beta, t) = S_n^{(1)}(\beta, t) / S_n^{(0)}(\beta, t)$. Define

$$v_{int} = \int_0^T \left[ \tilde{Z}_{int}(t) - e_n(\beta_0, t) \right] dN_{int}(t)$$

$$\int_0^T \left[ R_n^{(1)}(\beta_0(t)) - R_n^{(0)}(\beta_0(t)) e_n(\beta_0(t)) \right] d\Lambda_{int}(t),$$

where $R_n^{(k)}$ is the same as $R_n^{(k)}$ but with $\eta_{int}$ replaced by $\eta_{int}$ (k = 0, 1). Also, let

$$r_{int} = \left[ 2 \rho_m L_n \omega_{int}(\alpha_0) \right]^{-1} \eta_{int} \omega_{int}$$

$$\times J \left[ g_n^{(1)} - 2 g_n^{(0)}(\alpha_0) \eta_{int}(\alpha_0) g_n^{(0)}(\alpha_0) \right].$$

We can use the quadratic form $\hat{\theta}^T \hat{\Psi}^{-1} \hat{\theta}$ to test the global null hypothesis $H_0: \theta_1 = \theta_2 = \cdots = \theta_M = 0$. We can also use the sequential multiple testing procedure described in Wei et al. (1989) to determine which of the $M$ subhypotheses $H_m: \theta_m = 0$ ($m = 1, \ldots, M$) should be rejected. Under the condition that $\theta_1 = \theta_2 = \cdots = \theta_M = \theta$, we can estimate the common covariate effect $\theta$ by incorporating this restriction into model (6) and the corresponding estimating function $U_C(\beta)$. An alternative method, as suggested by Wei et al. (1989), is to estimate $\theta$ by the linearly combined estimator $\hat{\theta} \equiv \sum_{m=1}^M \gamma_m \hat{\theta}_m$, where the weights $\gamma_m$ satisfy $\sum_{m=1}^M \gamma_m = 1$. The variance of $\hat{\theta}$ can be easily obtained from $\Psi$. Equal weighting is achieved by setting $c_m = M^{-1}$ for all $m$. As shown by Wei et al. (1989), the choice of $(c_1, \ldots, c_M)^T = (e^T \hat{\Phi}^{-1} e)^{-1} \hat{\Phi}^{-1} e$, where $e = (1, \ldots, 1)^T$, minimizes the variance among all linear estimators. Note that the Wald statistic based on the combined estimator provides a 1-degree-of-freedom test of the global null hypothesis $H_0$, which tends to be more powerful than the $M$-degree-of-freedom test $\hat{\theta}^T \hat{\Psi}^{-1} \hat{\theta}$ if the $\theta_m$'s are similar.

4. SIMULATION STUDIES

In this section we first use simulation to compare the proposed method with the nonparametric correction method of Huang and Wang (2000) for univariate failure time data. We then evaluate the performance of the proposed method for recurrent event data.

4.1 Univariate Failure Time

We considered the proportional hazards model with unit exponential baseline hazard. Two covariates, $X$ and $V$, were generated from the bivariate normal distribution with $\text{Var}(X) = \text{Var}(V) = 1$ and $\text{Cov}(X, V) = 0$ or .5. The covariate $X$ was measured with error, whereas $V$ was measured precisely. The surrogate $W$ of $X$ was generated from the classical error model $W = X + \epsilon$, where $\epsilon$ is mean-zero normal with standard deviation $\sigma = .5$. We set $\beta_0 = (0.0, 0.0)^T = (.5, .5)^T$. Censoring times were generated from the uniform distribution on $[0, 0.5]$, where $C_M$ was set to be .5 or .30, inducing a censoring rate of 76% or 34%, respectively. Four different settings were simulated to explore the influence of the numbers of replicates. In the first setting every subject had three replicates of the surrogate ($n = 3$). In the second setting $n_t$ was set to 1 or 3 with equal probability and independently of other variables. For the last two settings, $n_t$ was set to 1 or 3 with probabilities dependent on the survival time $T_t$. In the third setting $P(n_t = 1) = .75$ if $T_t$ is less than the median survival time and $P(n_t = 1) = .25$ otherwise. Such a situation may arise if a covariate is stable over a long period of time and replicates are obtained periodically, so that the subjects who live longer tend to have more measurements. A good example of this sort is the NPC trial. The fourth setting is the opposite of the third, with $P(n_t = 1) = .25$ if $T_t$ is less than the median survival time and $P(n_t = 1) = .75$ otherwise. This scenario may occur if sicker patients tend to be more closely monitored. For each combination of the simulation parameters, 10,000 datasets of size 500 were generated. Four estimators of $\beta_0 = (0.0, 0.0)^T$ were evaluated: the full-data estimator, which uses the true value of $X$ for all subjects; the
naive estimator, which replaces $X$ with the average of all surrogates; the nonparametric correction estimator of Huang and Wang (hereafter referred to as HW; and the new estimator described in Section 2.

Table 1 summarizes the results under $Cov(X, V) = 0$. The naive estimator for $\alpha_0$ is always biased toward 0, whereas there is no bias for the naive estimator of $\gamma_0$. The proposed estimator performs consistently well in all settings. In the first two settings, the HW method corrects the bias well, but in the latter two settings where $P(n_1 = 1)$ depends on the failure time, it shows bias in both $\hat{\alpha}$ and $\hat{\gamma}$. In the first setting the proposed and the HW estimators have similar standard errors. In the second setting, however, the HW estimator shows more variability than the proposed estimator, because it can only use the subjects with multiple measurements of the surrogate, whereas the proposed method uses all subjects in the estimation. Although half of the subjects have only a single surrogate measurement in settings II-IV, the standard error of the proposed estimator is only slightly inflated relative to setting I. Simulation under $Cov(X, V) = .5$ (results not shown) revealed similar patterns except that the naive estimator is also biased for $\gamma_0$.

The proposed method requires symmetric error, whereas the HW method does not. Simulation studies were carried out to assess the bias of the proposed estimator when this assumption is violated. First, the error was generated from the standardized Beta$(2, 4)$ distribution and scaled to have standard deviation of .5. Here the symmetry condition is moderately violated. The error was then generated from the standardized Beta$(4, 2)$ and similarly scaled. This distribution is symmetric to the first one. We also considered more extreme distributions, the standardized $\chi^2_1$ distribution truncated at 5 and divided by 2, along with its symmetrical counterpart. For all settings we set $n = 500$ and $n_1 = 3$. The results are displayed in Table 2. When the error follows the Beta distribution, the bias of the proposed estimator is minimal; when the error follows the truncated chi-squared distributions, the bias becomes larger.

4.2 Recurrent Events

Extensive simulation studies were carried out to investigate the finite-sample properties of the proposed estimator for recurrent event data. The covariate vector $Z = (X, V)^T$ was generated in the same fashion as in Section 4.1. The surrogate $W$ was generated from the classical error model $Y = X + \epsilon$, where $\epsilon$ is normally distributed with mean 0 and standard deviation $\sigma_\epsilon = .2, .5, or 1.0$. For each subject $X$ was measured five times. We set $\beta_0 = (.5, .5)^T$. We generated recurrent event times from the random-effect intensity model $t(i; Z, \xi) = \xi \exp(\beta_0^T Z) \lambda_0(t) Y(t)$, where $\lambda_0 = 1$ and $\xi$ is a gamma variable with mean 1 and variance 0 (no random effect), .25, .5, or 1. The censoring times were generated from the uniform $[0, C_M]$ distribution, where $C_M = 1$ or 3. For each setting 10,000 datasets of size 500 were generated. Three estimators were evaluated: the full-data estimator, which uses the true value of $X$ for all subjects; the naive estimator, which replaces $X$ with the average of the five replicate measurements; and the proposed estimator.

Table 3 shows the results for $\alpha$ under $Cov(X, V) = 0$. On average, there are approximately .64 and 1.92 events per subject under $C_M = 1$ and 3, respectively. The naive estimator of $\alpha_0$ is biased, and the bias increases as the magnitude of the error increases. The bias of the naive estimator of $\gamma_0$ is small (results not shown). The proposed estimator corrects the bias very well. The proposed variance estimator reflects accurately the true variability, and the confidence intervals have satisfactory coverage probabilities. When the number of replicate measurements is correlated with the intensity of recurrent events,
Table 2. Simulation Results for Asymmetrical Error Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>34% censored</th>
<th></th>
<th></th>
<th>76% censored</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\alpha}$</td>
<td>SE</td>
<td>$\hat{\gamma}$</td>
<td>SE</td>
<td>$\hat{\alpha}$</td>
<td>SE</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>.060</td>
<td>Mean</td>
<td>.060</td>
<td>Mean</td>
<td>096</td>
</tr>
<tr>
<td>Full data</td>
<td>.503</td>
<td>.060</td>
<td>.503</td>
<td>.060</td>
<td>.503</td>
<td>.096</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>.459</td>
<td>.057</td>
<td>.498</td>
<td>.060</td>
<td>.461</td>
<td>.092</td>
</tr>
<tr>
<td>HW</td>
<td>.504</td>
<td>.065</td>
<td>.503</td>
<td>.061</td>
<td>.504</td>
<td>.102</td>
</tr>
<tr>
<td>Proposed</td>
<td>.496</td>
<td>.063</td>
<td>.502</td>
<td>.061</td>
<td>.496</td>
<td>.099</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error is $.5 \times $ Standardized Beta(2, 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>.460</td>
<td>.057</td>
<td>.498</td>
<td>.060</td>
<td>.464</td>
<td>.092</td>
</tr>
<tr>
<td>HW</td>
<td>.504</td>
<td>.064</td>
<td>.503</td>
<td>.061</td>
<td>.505</td>
<td>.102</td>
</tr>
<tr>
<td>Proposed</td>
<td>.513</td>
<td>.066</td>
<td>.504</td>
<td>.061</td>
<td>.513</td>
<td>.105</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error is $.5 \times $ Standardized Beta(4, 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>.465</td>
<td>.058</td>
<td>.499</td>
<td>.060</td>
<td>.467</td>
<td>.092</td>
</tr>
<tr>
<td>HW</td>
<td>.505</td>
<td>.064</td>
<td>.504</td>
<td>.062</td>
<td>.505</td>
<td>.101</td>
</tr>
<tr>
<td>Proposed</td>
<td>.478</td>
<td>.060</td>
<td>.500</td>
<td>.061</td>
<td>.480</td>
<td>.093</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error is $.5 \times $ Standardized $\chi^2$ Truncated at 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>.469</td>
<td>.058</td>
<td>.499</td>
<td>.060</td>
<td>.471</td>
<td>.094</td>
</tr>
<tr>
<td>HW</td>
<td>.503</td>
<td>.063</td>
<td>.503</td>
<td>.061</td>
<td>.504</td>
<td>.101</td>
</tr>
<tr>
<td>Proposed</td>
<td>.539</td>
<td>.074</td>
<td>.508</td>
<td>.061</td>
<td>.538</td>
<td>.116</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Error is $.5 \times $ Standardized $\chi^2$ Truncated at 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>.469</td>
<td>.058</td>
<td>.499</td>
<td>.060</td>
<td>.471</td>
<td>.094</td>
</tr>
<tr>
<td>HW</td>
<td>.503</td>
<td>.063</td>
<td>.503</td>
<td>.061</td>
<td>.504</td>
<td>.101</td>
</tr>
<tr>
<td>Proposed</td>
<td>.539</td>
<td>.074</td>
<td>.508</td>
<td>.061</td>
<td>.538</td>
<td>.116</td>
</tr>
</tbody>
</table>

NOTE: Mean and SE stand for the mean and standard error of the estimator.

Table 3. Simulation Results for Proportional Rates Models With Covariate Measurement Error

<table>
<thead>
<tr>
<th>Var($\zeta$)</th>
<th>$\sigma_i$</th>
<th>Method</th>
<th>$C_M = 1$</th>
<th></th>
<th>$C_M = 3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>SEE</td>
<td>CP</td>
<td>Mean</td>
</tr>
<tr>
<td>0</td>
<td>*</td>
<td>Full data</td>
<td>499</td>
<td>.057</td>
<td>.056</td>
<td>.940</td>
</tr>
<tr>
<td></td>
<td>.2</td>
<td>Naive</td>
<td>495</td>
<td>.057</td>
<td>.055</td>
<td>.939</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Proposed</td>
<td>500</td>
<td>.058</td>
<td>.056</td>
<td>.940</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Naive</td>
<td>476</td>
<td>.056</td>
<td>.055</td>
<td>.916</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>Proposed</td>
<td>500</td>
<td>.056</td>
<td>.058</td>
<td>.940</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Naive</td>
<td>416</td>
<td>.054</td>
<td>.052</td>
<td>.830</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>Proposed</td>
<td>504</td>
<td>.071</td>
<td>.069</td>
<td>.945</td>
</tr>
<tr>
<td>.25</td>
<td>*</td>
<td>Full data</td>
<td>500</td>
<td>.067</td>
<td>.065</td>
<td>.940</td>
</tr>
<tr>
<td></td>
<td>.2</td>
<td>Naive</td>
<td>496</td>
<td>.067</td>
<td>.065</td>
<td>.939</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Proposed</td>
<td>500</td>
<td>.067</td>
<td>.065</td>
<td>.940</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Naive</td>
<td>476</td>
<td>.066</td>
<td>.064</td>
<td>.915</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>Proposed</td>
<td>501</td>
<td>.069</td>
<td>.067</td>
<td>.940</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Naive</td>
<td>417</td>
<td>.062</td>
<td>.060</td>
<td>.695</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>Proposed</td>
<td>506</td>
<td>.081</td>
<td>.079</td>
<td>.947</td>
</tr>
<tr>
<td>.5</td>
<td>*</td>
<td>Full data</td>
<td>500</td>
<td>.077</td>
<td>.072</td>
<td>.930</td>
</tr>
<tr>
<td></td>
<td>.2</td>
<td>Naive</td>
<td>496</td>
<td>.076</td>
<td>.072</td>
<td>.928</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Proposed</td>
<td>500</td>
<td>.077</td>
<td>.073</td>
<td>.929</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Naive</td>
<td>476</td>
<td>.075</td>
<td>.071</td>
<td>.914</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>Proposed</td>
<td>501</td>
<td>.080</td>
<td>.075</td>
<td>.932</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Naive</td>
<td>416</td>
<td>.071</td>
<td>.067</td>
<td>.723</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>Proposed</td>
<td>505</td>
<td>.092</td>
<td>.087</td>
<td>.941</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>Full data</td>
<td>499</td>
<td>.090</td>
<td>.085</td>
<td>.929</td>
</tr>
<tr>
<td></td>
<td>.2</td>
<td>Naive</td>
<td>495</td>
<td>.089</td>
<td>.084</td>
<td>.926</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Proposed</td>
<td>499</td>
<td>.091</td>
<td>.085</td>
<td>.928</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Naive</td>
<td>475</td>
<td>.088</td>
<td>.083</td>
<td>.910</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>Proposed</td>
<td>500</td>
<td>.093</td>
<td>.087</td>
<td>.929</td>
</tr>
<tr>
<td></td>
<td>.5</td>
<td>Naive</td>
<td>416</td>
<td>.083</td>
<td>.078</td>
<td>.756</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>Proposed</td>
<td>505</td>
<td>.116</td>
<td>.100</td>
<td>.939</td>
</tr>
</tbody>
</table>

NOTE: Mean and SE are the mean and standard error of the estimator. SEE is the mean of the standard error estimator, and CP is the coverage probability of the 95% confidence interval.

* Results do not depend on this parameter.
the coefficient estimator, the variance estimator, and the coverage probability of the proposed method are all satisfactory (results not shown). Additionally, under $\text{cov}(X, V) = .5$ (results not shown), the naive estimators are biased for both $a_0$ and $x_0$, whereas the proposed method continues to perform well.

5. THE NUTRITIONAL PREVENTION OF CANCER TRIAL

The Nutritional Prevention of Cancer (NPC) trial is a double-blind, placebo-controlled clinical trial to evaluate the long-term effects of a nutritional supplement of selenium (Se) in preventing skin cancers (Clark et al. 1996). A total of 1,312 patients with previous histories of squamous cell carcinoma (SCC) or basal cell carcinoma (BCC) were randomized into two groups of approximately equal sizes. Each patient in the active treatment (i.e., Se) group was supposed to take a pill containing 200 $\mu$g of Se every day, and the patients in the other group were only given placebo pills. The primary outcome measures were new cases of SCC and BCC. The patients were followed up for up to 12 years.

Clark et al. (1996) reported adverse but statistically non-significant treatment effects on SCC and BCC. Jiang et al. (1999) studied the SCC recurrences by using a discrete-time proportional means model, which includes as covariates the treatment indicator and the baseline plasma Se level. As mentioned in Section 1, the latter covariate cannot be measured precisely due to its temporal biological fluctuation and the measurement error of the instrument. For all patients this covariate was measured at each routine clinic visit scheduled every 6 months after the randomization. Jiang et al. (1999) showed the plasma Se levels for the control patients were quite stable over time. Thus, they suggested using these measurements as replicates of the baseline value. For patients in the Se group, however, the plasma Se levels were altered by the treatment. Thus, each Se patient only had one measurement of the baseline plasma Se value. Jiang et al. assumed that the true covariate $X$ is normally distributed given the surrogate $W = X + \epsilon$ and the conditional mean $E(X|W)$ is a linear function of $W$. In addition, the error was assumed to be normally distributed.

We apply the proposed methods to the NPC trial. We confine our attention to the 1,286 patients with baseline Se measurements, of whom 647 belonged to the placebo group and 639 to the Se group. We first apply the (continuous-time) proportional rate model of Section 2 to the SCC recurrent events and then use the method of Section 3 to make simultaneous inference on the multiple episodes of BCC and SCC. Note that the nonparametric correction method of Huang and Wang (2000) is not applicable here (even if one considers only a single event) because it can only make use of the subjects with multiple replicates and in this case would exclude all the Se patients.

5.1 Recurrences of SCC

About one third (432) of the patients had at least one SCC after randomization, with 193 in the placebo group and 239 in the Se group. The control patients had as many as 27 Se measurements during the study, and all these readings are considered as replicates of the baseline value. Every patient in the Se group has only one valid measurement of this covariate. Following Jiang et al. (1999), we let $X$ be the logarithm of the true Se level and the surrogate $W$ be the logarithm of the observed Se level. The classical error model $W = X + \epsilon$ is assumed. To check the validity of this model, we adopt the idea of Jiang et al. (1999) to focus on the 285 patients in the placebo group with 10 or more Se measurements. Suppose that the $i$th patient belongs to this group. The average of all his/her logarithmic Se readings, $\bar{X}_i = n_i^{-1} \sum_{r=1}^{n_i} W_{ir}$, is a good approximation of the long-term average of that patient, which is $X_i$ if the classical error model holds. Thus, $\bar{W}_i = W_{ir} - X_i$ ($r = 1, \ldots, n_i$) are close to the true errors $\epsilon_{ir}$. Plots of the approximate errors $\bar{W}_i$ over time and against $X_i$ (not shown here) revealed no systematic trends. Thus, the error model is reasonable. Figure 1 presents the estimated density functions of the approximate errors $\bar{W}_i$ obtained by kernel smoothing. We show the error distributions at randomization, in the middle of the trial (between the 36th and 54th months), late in the trial (after the 84th month), and during the full course of the trial. It can be seen the error distributions are quite similar at different stages of the trial, so that the later Se measurements can be considered as replicates of the baseline value. The densities of $-\epsilon$ are also displayed in Figure 1, indicating that the symmetric error assumption is reasonable.

Having verified the classical additive error model and the symmetry condition on the error, we apply the method of Section 2 to the NPC data. The model includes the logarithm of the baseline plasma Se level and the treatment indicator as the covariates. All plasma Se readings of the control patients throughout the trial were used to produce the first set of results shown in Table 4. The naive estimate of the regression parameter of log(Se) is substantially smaller than the proposed estimate, reflecting the attenuation nature of the naive method.

Despite the foregoing exploratory data analysis, the question may still remain as to whether the readings obtained over 10 years after randomization can serve as replicates of the baseline value. To address this question, we perform some sensitivity analysis, using Se measurements in different time periods to fit the model. Table 4 provides the results based on two different subsets of the replicate data. The results based only on the measurements of the first 4 years are almost identical to those based on all Se readings, and the results based only on the first year's measurements are also close. The results show that subjects with higher baseline Se levels had significantly less SCC recurrences, whereas the negative effect of Se treatment was not statistically significant. Incidentally, based on a slightly different dataset, Jiang et al. (1999) reported the parameter estimate for log(Se) as $-2.076$ with an estimated standard error of .963. This estimate is much larger than our estimate. Jiang et al.'s estimate of the treatment effect is similar to ours.

5.2 Multivariate Analysis of BCC and SCC

We now use the method of Section 3 to examine the effects of the logarithmic baseline Se level ($X$) and treatment indicator ($V$) on the first four recurrences of BCC and SCC. We fit model (6) with $M = 8$ and $L = 1$. The main results are shown in Table 5. By sequential multiple testing (Wei et al. 1989), neither the baseline Se level nor the treatment has a significant effect on any of the four recurrences of BCC at the .05 level; the treatment has a significantly adverse effect on the time to the first SCC ($p$ value = .044), and the baseline Se level has a significantly negative effect on the time to each of the first four recurrences of SCC. The lower panel of Table 5 shows the estimates.
Figure 1. Kernel Smoothing Densities of Error Distributions During Different Time Periods. In each plot the solid curve is the density of the error, and the dotted curve is the density of the negative error.

for the common covariate effects based on the three methods mentioned at the end of Section 3. The first two methods suggest that the treatment effects on the first four recurrences of SCC are not significant. The results based on the last method are very close to the results based only on the first recurrence. All three methods show significantly negative effects of the baseline Se level on the first four recurrences of SCC.

<table>
<thead>
<tr>
<th>Method</th>
<th>log(Se) Estimate</th>
<th>(SE)</th>
<th>Treatment Estimate</th>
<th>(SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive</td>
<td>-0.728</td>
<td>(.302)</td>
<td>.108</td>
<td>(.118)</td>
</tr>
<tr>
<td>Proposed</td>
<td>-1.276</td>
<td>(.502)</td>
<td>.100</td>
<td>(.119)</td>
</tr>
<tr>
<td>Measurements in the first 48 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>-0.695</td>
<td>(.303)</td>
<td>.110</td>
<td>(.118)</td>
</tr>
<tr>
<td>Proposed</td>
<td>-1.236</td>
<td>(.511)</td>
<td>.103</td>
<td>(.119)</td>
</tr>
<tr>
<td>Measurements in the first 12 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td>-0.578</td>
<td>(.314)</td>
<td>.118</td>
<td>(.117)</td>
</tr>
<tr>
<td>Proposed</td>
<td>-1.010</td>
<td>(.518)</td>
<td>.115</td>
<td>(.118)</td>
</tr>
</tbody>
</table>

*Binary covariate, with placebo group as 0 and Se group as 1.

6. DISCUSSION

In this article we provide a general approach to dealing with covariate measurement error in the analysis of univariate and multivariate failure time data. The existing methods of Jiang et al. (1999) and Li and Lin (2000) can only handle special forms of multivariate failure time data and require that the true covariate and the measurement error follow specific parametric distributions. The proposed approach encompasses all types of multivariate failure time data and does not impose stringent parametric assumptions. The only assumption is that the error is symmetrically distributed.

For univariate failure time data, the method of Huang and Wang (2000) does not require symmetric error, and is thus recommended if the symmetry condition does not hold or is difficult to verify. When this assumption is reasonable, however, the proposed estimator has several advantages. In particular, if a subject has only one measurement of the surrogate, the proposed estimator can still incorporate the information of that subject into the estimation, whereas the Huang and Wang estimator has to exclude that subject. Thus, when a large proportion of the study subjects have only one surrogate measurement, the proposed estimator is more efficient. If multiple replicates are
not available at all levels of a discrete covariate, as in the case of the NPC trial, the effect of that covariate cannot be estimated by the Huang–Wang method. If the number of replicates for the subject depends on the failure time, then the Huang–Wang estimator can be biased, whereas the proposed estimator remains consistent.

As demonstrated in the simulation studies of Section 4, a slight to moderate deviation from the symmetry condition does not incur much bias in the proposed estimator. If the symmetry assumption cannot be verified by internal data, external data can be used instead, provided that the mechanisms generating the measurement error are similar between the internal and external data.

Although this article focuses on the case when replicates of the surrogate are available, the proposed approach can be adapted to accommodate validation data. In fact, our approach consists of two separate stages: first the estimation of the functions \( \eta_k(\alpha) \) (\( k = 0, 1 \)) and then the estimation of the parameters \( \beta_0 \) and \( \Delta_0(\cdot) \). How we estimate \( \eta_k(\alpha) \) does not interfere with the estimation of \( \beta_0 \) and \( \Delta_0(\cdot) \). Thus, our approach is quite flexible. We can estimate \( \eta_k(\alpha) \) using any information we have, either replicates or validation data, or a combination of them, or even external data. Finally, it is worthwhile to point out that one can extend the method of Huang and Wang (2000) to the setting of multivariate failure time data by following the steps outlined in Sections 2 and 3.

APPENDIX: TECHNICAL DETAILS

A.1 Regularity Conditions

The following conditions are required for Section 2:

1. \( \{N_i(t), Y_i(t), Z_i(t) \mid t = 1, \ldots, n \} \) are independent and identically distributed.

2. \( \alpha(t) \) (\( i = 1, \ldots, n \)) are bounded by a constant.

3. \( \alpha(t) \) is continuous and \( \Delta(t) < \infty \).

4. \( \Delta(1) = 1 \).

5. For all sample paths of \( Z_i(t) \) (\( i = 1, \ldots, n \)), \( |Z_i(0)| + \int_0^t |dZ_i(t)| \leq K < \infty \), where \( Z_i(t) \) is the \( j \)-th component of \( Z_i(t) \) and \( K \) is a constant.

6. The errors \( \epsilon \) are identically and symmetrically distributed, and are independent of any other random variables.

7. \( E(\epsilon_1^2) < \infty \). In addition, there exists a compact neighborhood \( A \) of \( \alpha \) such that \( E[\sup_{\alpha \in A} |\epsilon_1^2 - E(\epsilon_1^2)|] < \infty \) and \( E[\sup_{\alpha \in A} |\epsilon_1| |\epsilon_1^2 - E(\epsilon_1^2)|] < \infty \).

8. \( \Gamma \) is positive definite.

Note that condition 8 holds if, at least for some interval of \( t \), the distribution of \( Z_i(t) \) conditional on \( Y_i(t) = 1 \) does not concentrate on a \( (p + q - 1) \)-dimensional hyperplane.

For Section 3 we impose the same conditions, but with \( N_i, Y_i, Z_i, \Delta_i \) and \( \epsilon_{11} \) replaced by \( N_{i,1}, Y_{i,1}, Z_{i,1}, \Delta_{0,1} \) and \( \epsilon_{11} \), respectively. We label these conditions \( \Delta^{*2} \).

A.2 Proofs

**Proof of Theorem 1.** We shall prove the theorem for the proportional rates model. The proof for the proportional hazards model is similar and simpler and, thus, omitted. We first establish the asymptotic normality of \( n^{-1/2} U_C(\beta_0) \). Write

\[
U_C(\beta_0) = \sum_{i=1}^n \int_0^T \left[ \tilde{Z}_i(t) - E_C(\beta_0, t) \right] dN_i(t) - \tilde{R}_i(0)(\beta_0, t)Y_i(t)d\Delta_0(t) \]

\[
+ \sum_{i=1}^n \int_0^T \left[ \tilde{Z}_i(t) - E_C(\beta_0, t) \right] \tilde{R}_i(0)(\beta_0, t)Y_i(t)d\Delta_0(t). \quad (A.1)
\]

We decompose the first term on the right side of (A.1) as \( B_1 + B_2 + B_3 + B_4 \), where

\[
B_1 = \sum_{i=1}^n \int_0^T \left[ \tilde{Z}_i(t) - E_C(\beta_0, t) \right] \left[ dN_i(t) - \tilde{R}_i(0)(\beta_0, t)Y_i(t)d\Delta_0(t) \right].
\]

\[
B_2 = \sum_{i=1}^n \int_0^T \left[ \tilde{Z}_i(t) - E_C(\beta_0, t) \right] \left[ dN_i(t) - \epsilon_{11i}(\tilde{Z}_i(t))Y_i(t)d\Delta_0(t) \right],
\]

\[
B_3 = \sum_{i=1}^n \int_0^T \left[ \epsilon_{11i}(\tilde{Z}_i(t)) \right] \left[ d\Delta_0(t) - \epsilon_{11i}(\tilde{Z}_i(t))Y_i(t)d\Delta_0(t) \right],
\]

\[
B_4 = \sum_{i=1}^n \int_0^T \left[ \tilde{Z}_i(t) - E_C(\beta_0, t) \right] \left[ d\Delta_0(t) - \epsilon_{11i}(\tilde{Z}_i(t))Y_i(t)d\Delta_0(t) \right].
\]

Write \( B_2 = \int_0^T \left[ \epsilon_{11i}(\tilde{Z}_i(t)) \right] \left[ d\Delta_0(t) - \epsilon_{11i}(\tilde{Z}_i(t))Y_i(t)d\Delta_0(t) \right]. \) Because we assume the more general proportional rates model rather than the intensity model, \( M_i(t) \) may not be a martingale. Following Lin et al. (2000), we appeal to the modern empirical process theory. Write \( \mathcal{M}(t) = \sum_{i=1}^n M_i(t_i) \). By Example 2.11.16 of van der Vaart and Wellner (1996, p. 215), \( n^{-1/2} \mathcal{M} \) converges in \( \mathcal{L}^2 \) to a zero-mean Gaussian process \( \mathcal{W}_M \) with continuous paths. It follows from the uniform law of large-num-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
bers that, in a neighborhood $B$ of $\beta_0$,
\[
\sup_{t \in [0, \tau], \beta \in B} \left| S_t^{(k)}(\beta, t) - s_t^{(k)}(\beta, t) \right| \overset{P}{\to} 0 \quad (A.2)
\]
for $k = 0, 1$. By the strong embedding theorem (Shorack and Wellner 1986, pp. 47–48), there exists a new probability space in which $W^{(2)}_t\{S_t^{(0)}(\cdot, t), S_t^{(1)}(\cdot, t)\}$ also converges almost surely. It then follows from Lemma 1 of Lin et al. (2002) that
\[
n^{-1/2} \int_0^t S_t^{(1)}(\beta, t) \, d\tilde{M}(s) \overset{P}{\to} 0 \quad (A.3)
\]
almost surely uniformly in $t$. Likewise, $n^{-1/2} \int_0^t e_t(\beta, t) \, d\tilde{M}(s)$ goes to the same limit almost surely. Thus, $n^{-1/2} B_1 \overset{P}{\to} 0$ in the new probability space and $n^{-1/2} B_2 \overset{P}{\to} 0$ in the original space.

Similar arguments can be used to show that $n^{-1/2} B_3 \overset{P}{\to} 0$. Let $J_t(t) = \int_0^t e_t(\beta, t) \, d\tilde{M}(t)$, $I_t(t) = \sum_{\nu=1}^m J_t(t)$, and $\tilde{I}_t(t) = \sum_{\nu=1}^m J_t(t)$. Then $E[I_t(t)] = 0$ and $n^{-1/2} \tilde{I}_t(t)$ converges in $L^2$, $P$-a.s. to a zero-mean Gaussian process $\mathcal{V}_t$ with continuous sample paths. This, in a new probability space, $n^{-1/2} \tilde{I}_t(t)$ converges in $L^2$, $P$-a.s. to a zero-mean Gaussian process $\mathcal{V}_t$ with continuous sample paths.

For $B_6$ note that
\[
\begin{align*}
\hat{R}_i^{(1)}(\beta_0, \theta) - \hat{R}_i^{(1)}(\beta_0, t) &= \left[ \hat{\eta}_0^{-1}(\alpha_0) - \hat{\eta}_0^{-1}(\alpha_0) \right] \eta_i^{-1} \sum_{r=1}^{n} \epsilon_{i}^{(r)} \hat{Z}_r(\theta) Y_i(t) \overset{P}{\to} 0, \\
\hat{R}_i^{(2)}(\beta_0, \theta) - \hat{R}_i^{(2)}(\beta_0, t) &= J \left[ \hat{\eta}_0^{-2}(\alpha_0) \eta_i^{-1}(\alpha_0) - \hat{\eta}_0^{-2}(\alpha_0) \eta_i^{-1}(\alpha_0) \right] \eta_i^{-1} \sum_{r=1}^{n} \epsilon_{i}^{(r)} \hat{Z}_r(\theta).
\end{align*}
\]
We make the decomposition: $B_6 = B_{6,1} + B_{6,2} + B_{6,3}$, where
\[
\begin{align*}
B_{6,1} &= \left[ \hat{\eta}_0^{-1}(\alpha_0) - \hat{\eta}_0^{-1}(\alpha_0) \right] \\
B_{6,2} &= \left[ \hat{\eta}_0^{-1}(\alpha_0) - \hat{\eta}_0^{-1}(\alpha_0) \right] \\
B_{6,3} &= \left[ \hat{\eta}_0^{-2}(\alpha_0) \eta_i^{-1}(\alpha_0) - \hat{\eta}_0^{-2}(\alpha_0) \eta_i^{-1}(\alpha_0) \right]
\end{align*}
\]

Let $\varphi_1 = \int_0^t s_t^{(1)}(\beta_0, t) \, d\tilde{M}_t(t)$. We have
\[
\begin{align*}
B_{6,1} &= \left[ \hat{\eta}_0^{-1}(\alpha_0) \varphi_1 + \eta_0^{-1}(\alpha_0) \eta_1(\alpha_0) \eta_0 \right] \left[ \hat{\eta}_0^{-1}(\alpha_0) - \eta_0^{-1}(\alpha_0) \right] \\
&\quad + o_p(n^{1/2}), \\
B_{6,2} &= \left[ \hat{\eta}_0^{-1}(\alpha_0) \varphi_1 + \eta_0^{-1}(\alpha_0) \eta_1(\alpha_0) \eta_0 \right] \left[ \hat{\eta}_0^{-1}(\alpha_0) - \eta_0^{-1}(\alpha_0) \right] \\
&\quad + o_p(n^{1/2}), \\
B_{6,3} &= \eta_0^{-1}(\alpha_0) \eta_0 \eta_1(\alpha_0) \left[ \hat{\eta}_0^{-1}(\alpha_0) - \eta_0^{-1}(\alpha_0) \right] \\
&\quad - 2\eta_0^{-2}(\alpha_0) \eta_1(\alpha_0) \eta_0 \left[ \hat{\eta}_0^{-1}(\alpha_0) - \eta_0^{-1}(\alpha_0) \right] + o_p(n^{1/2}).
\end{align*}
\]

Hence,
\[
B_6 = -\eta_0^{-2}(\alpha_0) \eta_1(\alpha_0) \eta_0 \left[ \hat{\eta}_0^{-1}(\alpha_0) - \eta_0^{-1}(\alpha_0) \right] \\
&\quad + \eta_0^{-1}(\alpha_0) \varphi_1 \eta_1(\alpha_0) \left[ \hat{\eta}_0^{-1}(\alpha_0) - \eta_0^{-1}(\alpha_0) \right] + o_p(n^{1/2}).
\]
The left side of (A.5) can be written as
\[ \frac{\eta^*_0}{\eta_0} \eta^*_1(\alpha) - \eta_0(\alpha) \eta^*_1(\alpha) + \eta_0(\alpha) \eta^*_1(\alpha) - \eta_0(\alpha) \eta^*_1(\alpha) \]
\[ = \eta_1(\alpha) \left[ \frac{\eta_0(\alpha)}{\eta_0(\alpha)} - \eta_1(\alpha) \right] + \eta_0(\alpha) \left[ \frac{\eta_0(\alpha)}{\eta_0(\alpha)} - \eta_1(\alpha) \right] \]
\[ + o_p(n^{-1/2}). \] 

It then follows from (A.5) and (A.6) that
\[ n \left[ \hat{\eta}_1(\alpha) - \eta_1(\alpha) \right] \]
\[ = \left\{ 2\rho_0(\alpha) \frac{\eta^*_0}{\eta_0} \right\} \sum_{i=1}^n \left[ \omega_i g_i(\alpha) - \eta_0(\alpha) \eta_1(\alpha) \omega_i g_i(\alpha) \right] \]
\[ - \eta_0(\alpha) \eta_1(\alpha) + o_p(n^{-1/2}). \] (A.7)

Plugging both (A.6) and (A.7) into (A.3), we have
\[ B_n + B_n = \sum_{i=1}^n \xi_i r_i + o_p(n^{-1/2}). \]

Note that \( E(\xi_i r_i) = 0 \) for \( i = 1, \ldots, n \).

Summarizing the preceding results, we have
\[ n^{-1/2} U_C(\beta_0) = n^{-1/2} \sum_{i=1}^n \left( v_i + \xi_i r_i \right) + o_p(1). \] (A.8)

It then follows from the multivariate central limit theorem that
\[ n^{-1/2} U_C(\beta_0) \xrightarrow{L} N(\mathbf{0}, \Sigma_C). \]

Let \( \eta_2(\beta, t) = \partial \eta_1(\beta, t)/\partial \beta \). Clearly,
\[ \eta_2(\beta, t) = n^{-1/2} \sum_{r=1}^n e^{\beta^T Z_r(t)} \]
\[ \times \left[ \eta_0(\alpha) \eta_1^* \frac{\alpha}{\eta_1(\alpha)} + 2\eta_0(\alpha) \eta_1(\alpha) \eta_1^* \right] \]
\[ - \eta_0(\alpha) \eta_1(\alpha) \frac{\alpha}{\eta_1(\alpha)} - \eta_0(\alpha) \eta_1(\alpha) \eta_1^* + \eta_0(\alpha) \eta_1(\alpha) \eta_1^* \eta_2 \eta_1^* \eta_2 \sum_{r=1}^n \xi_i g_i(\alpha). \]

where \( \eta_2(\alpha) = \eta_2(\beta, t)/\partial \alpha \) is a consistent estimator for \( \eta_2(\alpha) \). Note that \( \eta_2(\alpha) = -n^{-1}(\alpha) \eta_1^* \frac{\alpha}{\eta_1(\alpha)} + (2n0 \eta_0(\alpha)^{-1} \sum_{r=1}^n \xi_i g_i(\alpha)). \)

Let \( \eta_3(\beta, t) = \partial \eta_2(\beta, t)/\partial \beta \). Clearly, \( \eta_3(\beta, t) \) is obtained from \( \eta_2(\beta, t) \) by replacing \( \eta_2(\beta, t) \) with \( \eta_3(\beta, t) \).

Let \( \eta_4(\beta, t) = n^{-1} \sum_{r=1}^n \xi_i g_i(\alpha). \)

It follows from the uniform law of large numbers that, in a neighborhood \( B \),
\[ \sup_{\beta \in B} \left| \left( \eta_0^2(\beta, t) - s^2(\beta, t) \right) \right| \rightarrow 0. \]

Condition 4 guarantees that \( s(\beta, t) \) is bounded away from 0 for \( t \in [0, \tau] \) and \( \beta \in B \). Thus,
\[ \sup_{t \in [0, \tau]} \left| \eta_5(\beta, t) \right| \rightarrow 0. \]

where \( \eta_5(\beta, t) = s(\beta, t)^2 s^2(\beta, t) - (\beta, t) \eta_2(\beta, t) \).

Note that
\[ \frac{dU_C(\beta)}{d \beta} = - \int_0^t \left[ \frac{d E_C(\beta, t)}{d \beta} \right] d \Lambda_0(t) \]
\[ = - \int_0^t \left[ \frac{d E_C(\beta, t)}{d \beta} \right] \left( \sum_{i=1}^n e^{\beta^T Z_i(t)} \right) d \Lambda_0(t) \]
\[ = - \int_0^t \left[ \frac{d E_C(\beta, t)}{d \beta} \right] \left( \sum_{i=1}^n Y_i(t) \right) d \Lambda_0(t) \]
\[ \left. \right\} = \frac{d \tilde{M}_0(t)}{d \tilde{M}_0(t)}. \]

It can be shown that \( n^{-1} \int_0^t \left( |dE_C(\beta, t)/\partial \beta| \right) d \tilde{M}_0(t) \rightarrow 0 \) uniformly in \( \beta \in B \). Let \( \Gamma(\beta) = n^{-1} \int_0^t \left( |dE_C(\beta, t)/\partial \beta| \right) d \Lambda_0(t) \).

We have
\[ \sup_{\beta \in B} \left| n^{-1} \int_0^t \left( |dE_C(\beta, t)/\partial \beta| \right) d \Lambda_0(t) \right| \rightarrow 0. \]

(A.10)

The limit of \( n^{-1} \frac{dU_C(\beta, t)}{d \beta} \) is nonpositive definite everywhere and negative definite at \( \beta_0 \). Note that \( n^{-1} \frac{dU_C(\beta, t)}{d \beta} \xrightarrow{p} 0 \). It follows from the proof of Theorem 2 of Foutz (1977) that \( \tilde{B}_C \) exists and is unique in \( B \) with probability converging to 1 as \( n \rightarrow \infty \) and \( \tilde{B}_C \xrightarrow{p} \beta_0 \).

By the Taylor series expansion,
\[ n^{-1/2} U_C(\beta) = n^{-1/2} U_C(\beta_0) + n^{-1/2} \frac{dU_C(\beta^*)}{d \beta} n^{-1/2} (\beta - \beta_0), \]
where \( \beta^* \) lies on the line segment between \( \tilde{B}_C \) and \( \beta_0 \). Therefore,
\[ n^{-1/2} (\beta - \beta_0) \]
\[ = - \left[ n^{-1} \frac{dU_C(\beta^*)}{d \beta} \right] n^{-1/2} U_C(\beta_0) \xrightarrow{L} N(0, \Sigma_C). \]

Proof of Theorem 2. We make the decomposition: \( \Lambda_C(t) = \Lambda_0(t) + D_1(t) + D_2(t) + D_3(t) + D_4(t) \), where
\[ D_1(t) = \sum_{i=1}^n \right| \left. \frac{1}{\sum_{j=1}^n Y_j(s) \hat{R}_{0j}(\beta, s)} \right| dN_i(s). \]
\[ D_2(t) = \sum_{i=1}^n \right| \left. \frac{1}{\sum_{j=1}^n Y_j(s) \hat{R}_{0j}(\beta, s)} \right| dN_i(s). \]
\[ D_3(t) = \sum_{i=1}^n \right| \left. \frac{1}{\sum_{j=1}^n Y_j(s) \hat{R}_{0j}(\beta, s)} \right| dN_i(s). \]
\[ D_4(t) = \Lambda_0(t). \]

and \( \Lambda_0(t) = \sum_{i=1}^n \int_0^t \left( \frac{dN_i(s)}{dN_i(s)} \right) \right| dN_i(s). \)

Let \( \beta^* \) lies on the line segment between \( \tilde{B}_C \) and \( \beta_0 \). It can be seen that
\[ \frac{dG(\beta, t)}{d \beta} = \int_0^t \left( \frac{S_C^2(\beta, s)}{[S_C^2(\beta, s)]^2} \right) \left( \frac{dN_i(s)}{dN_i(s)} \right) \]
For $D_2(t)$ first note that $R_{k0}^{(1)}(\beta_0, t) = \tilde{\gamma}_k^{-1}(\alpha_0) \eta_0(\alpha_0) R_{k0}^{(1)}(\beta_0, t)$. Thus, it follows from (4.1) that

$$D_2(t) = \frac{\tilde{\gamma}_k(\alpha_0) - \eta_0(\alpha_0)}{\eta_0(\alpha_0)} \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{j=1}^n Y_j(s) R_{k0}^{(1)}(\beta_0, s)} d\Lambda_0(s)$$

$$= 2 \rho \tilde{\gamma}_k^{-1}(\alpha_0) - \frac{\Lambda_0(t)n^{-1}}{} - \sum_{i=1}^n \xi_i [\alpha_0 \tilde{\gamma}_k^{-1}(\alpha_0) - \eta_0^2(\alpha_0)] + o_p(n^{-1/2}).$$

Clearly,

$$D_3(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s) - \sum_{j=1}^n Y_j(s) R_{k0}^{(1)}(\beta_0, s)}{\sum_{j=1}^n Y_j(s) R_{k0}^{(1)}(\beta_0, s)} d\Lambda_0(s)$$

$$= n^{-1} \sum_{i=1}^n \int_0^t dN_i(s) - \int_0^t Y_i(s) R_{k0}^{(1)}(\beta_0, s) d\Lambda_0(s).$$

Let $M_i(t) = N_i(t) - \int_0^t R_{k0}^{(1)}(\beta_0, s) Y_i(s) d\Lambda_0(s)$. Because $dN_i(t) = Y_i(t) d\tilde{\gamma}_k(t)$,

$$E[dN_i(t)I_i(t), Y_i(t)] = Y_i(t) E[d\tilde{\gamma}_k(t)I_i(t), Y_i(t)]$$

$$= Y_i(t) \exp[\beta_0^T Z_i(t)] d\Lambda_0(t).$$

In addition, $E[Y_i(t) R_{k0}^{(1)}(\beta_0, t) d\Lambda_0(t) | Z_i(t), Y_i(t), \eta_i(t)] = Y_i(t) \exp[\beta_0^T Z_i(t)] d\Lambda_0(t).$ Thus,

$$E[dM_i(t)I_i(t), Y_i(t)] = 0,$$

which implies that $E[dM_i(t)] = 0$ and $E[M_i(t)] = 0$. Note that

$$D_3(t) = n^{-1} \int_0^t \frac{d\tilde{M}(s)}{n^{-1} \sum_{j=1}^n Y_j(s) R_{k0}^{(1)}(\beta_0, s)}.$$

where $\tilde{M}(t) = \sum_{i=1}^n M_i(t)$. Using the empirical process theory as in the proof of the asymptotic normality of $U_C(\beta_0)$, we can show that

$$n^{1/2} \int_0^t \left[ n^{-1} \sum_{j=1}^n Y_j(s) R_{k0}^{(1)}(\beta_0, s) \right]^{-1}$$

$$- \left[ \tilde{\gamma}_k^{(0)}(\beta_0, s) \right]^{-1} d\tilde{M}(s) \overset{P}{\rightarrow} 0.$$ 

We then have

$$D_3(t) = n^{-1} \sum_{i=1}^n \int_0^t \frac{d\tilde{M}(s)}{\tilde{\gamma}_k^{(0)}(\beta_0, s)} + o_P(n^{-1/2}).$$

Finally,

$$D_4(t) = \Lambda_0(t) - \Lambda_0(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s)} d\Lambda_0(s).$$

Thus, we have $P[\sup_{t \in [0, 1]} |D_4(t)| \neq 0] \leq P[Y_i(t) = 0, i = 1, \ldots, n] \overset{P}{\rightarrow} 0$ by condition 4. Hence, $n^{1/2}D_4 \overset{P}{\rightarrow} 0$ uniformly in $t \in [0, 1].$

Summarizing the preceding results for $D_k(t)$ ($k = 1, \ldots, 4$), we obtain

$$n^{1/2} \left[ \tilde{\lambda}_C(t) - \Lambda_0(t) \right] = n^{1/2} \sum_{i=1}^n u_i(t) + o_P(1),$$

where

$$u_i(t) = -\int_0^t e^T(\beta_0, s) d\lambda_0(s) \frac{1}{n} \sum_{j=1}^n \left[ \{ \beta_0, s \} - \beta_0, s \right] + \int_0^t \frac{d\tilde{M}(s)}{\tilde{\gamma}_k^{(0)}(\beta_0, s)}.$$ 

Clearly, the mean of $u_i(t)$ is 0. Let $\phi(t, s) = E[u_i(t)u_i(s)]$. It can be shown that $u_i(t)$ is the sum of monotone functions in $t$. It then follows from Example 2.11.16 of van der Vaart and Wellner (1996) that $n^{1/2}[\tilde{\lambda}_C(t) - \Lambda_0(t)]$ converges weakly to a zero-mean Gaussian process with covariance function $\phi(t, s)$ at $(t, s)$. By the same decomposition $\Lambda_0(t) - \Lambda_0(t) = \sum_{i=1}^n D_i(t)$ and the uniform law of large numbers, we can show the uniform consistency of $\tilde{\lambda}_C(t)$.

Proof of the consistency of $\tilde{\phi}_C$. By (A.2) and condition 4, suppose $\sup_{t \in [0, 1]} |\phi_0(T, t) - \phi(T, t)| \overset{P}{\rightarrow} 0$. The consistency of $\tilde{\phi}$ for $\Gamma$ then follows easily from (A.2), (A.9), the consistency of $\beta_C$, and the uniform consistency of $\tilde{\lambda}_C$. For the consistency of $\tilde{\phi}_C$, we write $\tilde{\phi}_C = \sum_{i=1}^n (v_i + \xi_i) + \tilde{\phi}_C$. By the Cauchy-Schwarz inequality, it suffices to show that $n^{-1} \sum_{i=1}^n (v_i - \xi_i) \overset{P}{\rightarrow} 0$ and $n^{-1} \sum_{j=1}^n (v_j - \xi_j) \overset{P}{\rightarrow} 0$. The latter can be proved easily by using the consistency of $\tilde{\phi}_C$. Let $E = \sum_{i=1}^n (v_i - \xi_i)$. Define $\tilde{\phi}_C = \sum_{i=1}^n (v_i - \xi_i) \overset{P}{\rightarrow} 0$. For $j = 1, \ldots, 4$, the convergence follows easily from condition 2 and the uniform convergence of $E_C$. For $j = 2, 3, 4$, the claim is established by using conditions 3, 5, and 7, the uniform consistency of $\tilde{\lambda}_C$, and Lemma 1 of Lin et al. (2000).

Proof of Theorem 3. For any $m = 1, \ldots, M$ and $l = 1, \ldots, L$, we have

$$n^{-1/2} \sum_{i=1}^n \int_0^t \left[ \tilde{\psi}_{\text{inti}}(t) - \tilde{\psi}_{\text{int}}(T, t) \right] d\psi_{\text{inti}}(t)$$

$$= n^{-1/2} \sum_{i=1}^n (v_{\text{inti}} + \xi_{\text{inti}}) + o_P(1)$$

by the proof of Theorem 1. Thus,

$$n^{-1/2} \psi_{\text{inti}}(\beta_0) = n^{-1/2} \sum_{i=1}^n \sum_{j=1}^m \sum_{l=1}^L (v_{\text{intj}} + \xi_{\text{intj}}) + o_P(1),$$

and $n^{-1/2} \psi_{\text{inti}}(\beta_0) \overset{P}{\rightarrow} N(0, \Gamma_\beta).$ As in the proof of Theorem 1, there exists a compact neighborhood of $B$ of $\beta_0$ such that $\sup_{t \in [0, 1]} |\phi(T, t) - \phi(T, t)| \overset{P}{\rightarrow} 0$. The existence, uniqueness, and consistency of $\beta_C$ can be proved in the same manner as in the proof of Theorem 1. The Taylor series expansion can then be used to show that

$$n^{-1/2} \beta_C - \beta_0 \overset{P}{\rightarrow} N(0, \Gamma_\beta).$$

Proof of Theorem 4. Let $\Lambda_0(t) = \int_0^t \{ \lambda_0(s) - \beta_0 R_{k0}(s) \} d\lambda_0(s)$ and $\tilde{\lambda}_C(t) = \sum_{i=1}^n D_i(t)$. Then

$$D_1(t) = n^{-1} \sum_{i=1}^n \int_0^t \left[ \tilde{\phi}^{(0)}(\beta_0, s) - \phi^{(0)}(\beta_0, s) \right] dN_i(s).$$
\[ D_2(t) = \sum_{i=1}^{n} \sum_{l=1}^{L} \int_0^t \left[ n^{-1} S_{\theta}^{(0)}(\beta_0, s)^{-1} \right] dN_{\text{lm}}(s), \]
\[ D_3(t) = \sum_{i=1}^{n} \sum_{l=1}^{L} \int_0^t \left[ \sum_{j=1}^{n} \sum_{k=1}^{L} Y_{\text{lm}j}(s) R_{\text{lm}j}^{(0)}(\beta_0, s) \right] dN_{\text{lm}}(s), \]
\[ = \Lambda_{\text{lm}}^*(t), \]
and \( D_4(t) = \Lambda_{\text{lm}}^*(t) - \Lambda_{\text{lm}}(t). \) Also, let \( M_{\text{lm}}^*(t) = N_{\text{lm}}(t) - \int_0^t R_{\text{lm}}^{(0)}(\beta_0, s) dN_{\text{lm}}(s). \) Following the steps in the proof of Theorem 2, we see that \( n^{1/2} \{ \Lambda_{\text{lm}}(t) - \Lambda_{\text{lm}}(t) \} = n^{-1/2} \sum_{i=1}^{n} \sum_{l=1}^{L} u_{\text{lm}}(t) + o_p(1), \)
where
\[ u_{\text{lm}}(t) = -\int_0^t e_{\text{lm}}^T(\beta_0, s) d\Lambda_{\text{lm}}(s) \Gamma_n^{-1} \sum_{k=1}^{M} \sum_{\ell=1}^{M} \left( \psi_{\ell k}^T + \xi_{\ell k}^T \right) \psi_{\ell k} \]
\[ + \left\{ 2 \rho_n \ln \eta_{\text{lm}}^*(\alpha_0) \right\}^{-1} \Lambda_{\text{lm}}(t) \xi_{\text{lm}} \psi_{\text{lm}}^{(0)} R_{\text{lm}}^{(0)} - \delta_{\text{lm}}(\alpha_0) \}
\[ + \int_0^t \Lambda_{\text{lm}}^*(t) dM_{\text{lm}}^*(s). \]

Let \( u_{\text{lm}}(t) = \sum_{l=1}^{L} u_{\text{ml}}(t) \) and \( \phi_{\text{lm}}(t, s) = E(u_{\text{lm}}(t) u_{\text{lm}}(s)). \) By the arguments used in the proof of Theorem 2, we can show that \( n^{1/2} \{ \Lambda_{\text{lm}}(t) - \Lambda_{\text{lm}}(t) \} \) converges weakly to a zero-mean Gaussian process with covariance function \( \phi_{\text{lm}}(t, s), \) and \( \Lambda_{\text{lm}}^*(\cdot) \) converges to \( \Lambda_{\text{lm}}(\cdot) \) uniformly.

(Received July 2002. Revised July 2003.)

REFERENCES


