Landmark Estimation of Survival and Treatment Effect in a Randomized Clinical Trial

Layla Parast[1], **Lu Tian**[2], and **Tianxi Cai**[3]

[1]RAND Corporation, Santa Monica, CA 90401

[2]Stanford University, Department of Health, Research and Policy, Stanford, CA 94305

[3]Harvard University, Department of Biostatistics, Boston, MA 02115

Summary

In many studies with a survival outcome, it is often not feasible to fully observe the primary event of interest. This often leads to heavy censoring and thus, difficulty in efficiently estimating survival or comparing survival rates between two groups. In certain diseases, baseline covariates and the event time of non-fatal intermediate events may be associated with overall survival. In these settings, incorporating such additional information may lead to gains in efficiency in estimation of survival and testing for a difference in survival between two treatment groups. If gains in efficiency can be achieved, it may then be possible to decrease the sample size of patients required for a study to achieve a particular power level or decrease the duration of the study. Most existing methods for incorporating intermediate events and covariates to predict survival focus on estimation of relative risk parameters and/or the joint distribution of events under semiparametric models. However, in practice, these model assumptions may not hold and hence may lead to biased estimates of the marginal survival. In this paper, we propose a semi-nonparametric two-stage procedure to estimate and compare *t*-year survival rates by incorporating intermediate event information observed before some landmark time, which serves as a useful approach to overcome semi-competing risks issues. In a randomized clinical trial setting, we further improve efficiency through an additional calibration step. Simulation studies demonstrate substantial potential gains in efficiency in terms of estimation and power. We illustrate our proposed procedures using an AIDS Clinical Trial Protocol 175 dataset by estimating survival and examining the difference in survival between two treatment groups: zidovudine and zidovudine plus zalcitabine.

1 Introduction

In many research settings with an event time outcome, it is often not feasible to fully observe the primary event of interest as these outcomes are typically subject to censoring. In the presence of noninformative censoring, the Kaplan Meier (KM) estimator (Kaplan & Meier, 1958) can provide consistent estimation of the event rate. However, such a simple estimate may not be extremely precise in settings with heavy censoring. To improve the efficiency, many novel procedures have been proposed to take advantage of available auxiliary information. For example, when the auxiliary information consists of a single discrete variable W, it has been shown that the nonparametric maximum likelihood estimator of survival is $n^{-1} \sum_i \hat{F}_{T|W}(t|W_i)$ where $\hat{F}_{T|W}$ is the KM estimate of survival based on the subsample with $W = W_i$ (Rotnitzky & Robins, 2005). Murray & Tsiatis (1996) considered a

nonparametric estimation procedure to incorporate a single discrete covariate and provided theoretical results on when such augmentation enables more efficient estimation than the KM estimate. However, when multiple and/or continuous covariates are available, such fully nonparametric procedures may not perform well due to the curse of dimensionality.

Additional complications may arise when auxiliary variables include intermediate event information observed over time. In many clinical studies, information on non-fatal intermediate events associated with survival may be available in addition to baseline covariates. For example, in acute leukemia patients, the development of acute graft-versus-host disease is often monitored as it is predictive of survival following bone marrow or stem cell transplantation (Lee et al., 2002; Cortese & Andersen, 2010). The occurrence of bacterial pneumonia provides useful information for predicting death among HIV-positive patients (Hirschtick et al., 1995). In these settings, incorporating intermediate event information along with baseline covariates may lead to gains in efficiency for the estimation and comparison of survival rates.

In the aforementioned examples, the primary outcome of interest is time to a terminal event such as death and the intermediate event is time to a non-terminal event. This setting is referred to as a semi-competing risk setting since the occurrence of the terminal event would censor the non-terminal event but not vice versa. With a single intermediate event and no baseline covariates, Gray (1994) proposed a kernel smoothing procedure to incorporate such event information in order to improve estimation of survival. Parast et al. (2011) proposed a nonparametric procedure for risk prediction of residual life when there is a single intermediate event and a single discrete marker. Parast et al. (2012) extended this procedure to incorporate multiple covariates using a flexible varying-coefficient model. However, such methods cannot be easily extended to allow for both multiple intermediate events and baseline covariates due to the curse of dimensionality (Robins & Ritov, 1997).

Most existing methods for analyzing semi-competing risk data focus on estimation of relative risk parameters and/or the joint distribution of events under semiparametric models (Fine et al., 2001; Siannis et al., 2007; Jiang et al., 2005). These semiparametric models, while useful in approximating the relationship between the event times and predictors, may not be fully accurate given the complexity of the disease process. Therefore, marginal survival rates derived under such models may be biased and thus lead to invalid conclusions (Lin & Wei, 1989; Hjort, 1992; Lagakos, 1988). To overcome such limitations, we propose a two-stage procedure by (i) first using a semiparametric approach to incorporate baseline covariates and intermediate event information observed before some landmark time; and (ii) then estimating the marginal survival nonparametrically by smoothing over risk scores derived from the model in the first stage. The landmarking approach allows us to overcome semi-competing risk issues and the smoothing procedure in the second stage ensures the consistency of our survival estimates.

In a randomized clinical trial (RCT) setting, there is often interest in testing for a treatment difference in terms of survival. Robust methods to incorporate auxiliary information when testing for a treatment effect have been previously proposed in the literature. Cook & Lawless (2001) discuss a variety of statistical methods that have been proposed including parametric and semiparametric models. Gray (1994) adopted kernel estimation methods which incorporate information on a single intermediate event. To incorporate multiple baseline covariates, Lu & Tsiatis (2008) used an augmentation procedure to improve the efficiency of estimating the log hazard ratio, β, from a Cox model and testing for an overall treatment effect by examining the null hypothesis $\beta = 0$ and demonstrated substantial gains in efficiency. However, if the Cox model does not hold, the hazard ratio estimate converges to a parameter that may be difficult to interpret and may depend on the censoring

distribution. It would thus be desirable to use a model-free summary measure to assess treatment difference. Furthermore, none of the existing procedures incorporate both intermediate event and baseline covariate information. In this paper, we use our proposed two-stage estimation procedure, which utilizes all available information, to make more precise inference about treatment difference as quantified by the difference in t-year survival rates.

The rest of the paper is organized as follows. Section 2 describes the proposed two-stage estimation procedure in a one-sample setting. Section 3 presents a testing procedure using these resulting estimates. We present an additional augmentation step which takes advantage of treatment randomization to further improve efficiency in an RCT setting. Simulation studies given in Section 4 demonstrate substantial potential gains in efficiency in terms of estimation and power. In Section 5 we illustrate our proposed procedures using an AIDS Clinical Trial comparing the survival rates of HIV patients treated with zidovudine and those treated with zidovudine plus zalcitabine.

2 Estimation

For the ith subject, let $T_{\mathbb{L}i}$ denote the time of the primary terminal event of interest, $\mathbf{T}_{\mathbb{S}i}$ denote the vector of intermediate event times, $\mathbf{Z}_i$ denote the vector of baseline covariates, and C_i denote the administrative censoring time assumed independent of $(T_{\mathbb{L}}, \mathbf{T}_{\mathbb{S}i}, \mathbf{Z}_i)$. Due to censoring, for $T_{\mathbb{L}i}$ and $\mathbf{T}_{\mathbb{S}i}$, we could *potentially* observe $X_{\mathbb{L}i} = \min(T_{\mathbb{L}i}, C_i)$, $\mathbf{X}_{\mathbb{S}i} = \min(\mathbf{T}_{\mathbb{S}i}, C_i)$ and $\delta_{\mathbb{L}i} = I(T_{\mathbb{L}i} \quad C_i)$, $\delta_{\mathbb{S}i} = I(\mathbf{T}_{\mathbb{S}i} \quad C_i)$. Due to semi-competing risks, $\mathbf{T}_{\mathbb{S}i}$ is additionally subject to *informative* censoring by $T_{\mathbb{L}i}$, while $T_{\mathbb{L}i}$ is only subject to administrative censoring and cannot be censored by $\mathbf{T}_{\mathbb{S}i}$. Let t_0 denote some landmark time prior to t. Our goal is to estimate $S(t) = P(T_{\mathbb{L}i} > t)$ using intermediate event collected up to t_0 and baseline covariate information, where t is some pre-specified time point such that $P(X_{\mathbb{L}} > t \mid T_{\mathbb{L}} \quad t_0) \in (0, 1)$ and $P(T_{\mathbb{L}} \quad t_0, T_{\mathbb{S}} \quad t_0) \in (0, 1)$.

2·1 Two-stage estimation procedure

For $t > t_0$, $S(t) = P(T_{\mathbb{L}i} > t)$ can be expressed as $S(t \mid t_0)S(t_0)$, where

$$S(t|t_0)=P(T_{\mathbb{L}i}>t|T_{\mathbb{L}i}>t_0)=P(T_{\mathbb{L}i}>t|X_{\mathbb{L}i}>t_0) \text{ and } S(t_0)=P(T_{\mathbb{L}i}>t_0). \quad (2\cdot1)$$

We first consider estimation of $S(t \mid t_0)$ with a two-stage procedure using subjects in the subset $\Omega_{t_0} = \{i : X_{\mathbb{L}i} > t_0\}$ as illustrated in Figure 1. Note that conditional on $X_{\mathbb{L}i} > t_0$, $\mathbf{T}_{\mathbb{S}i}$ is observable up to $\mathbf{T}_{\mathbb{S}i} \wedge t_0$ and $I(\mathbf{T}_{\mathbb{S}i} \quad t_0)$ is always observable. Let $\mathbf{W}_i = [\mathbf{Z}_i^{\mathrm{T}}, I(\mathbf{T}_{\mathbb{S}i} \leq t_0), \min(\mathbf{T}_{\mathbb{S}i}, t_0)]^{\mathrm{T}}$ for all subjects in Ω_{t_0}. Note that for any function of $\mathbf{W}$, $\zeta(\cdot)$,

$$S(t|t_0)=P(T_{\mathbb{L}i}>t|X_{\mathbb{L}i}>t_0)=E\{S_{\zeta(\mathbf{W})}(t|t_0)\}, \quad (2\cdot2)$$

where $S_{\zeta(\mathbf{W})}(t \mid t_0) = P\{T_{\mathbb{L}i} > t \mid X_{\mathbb{L}i} > t_0, \zeta(\mathbf{W}_i)\}$. Hence if $\hat{S}_{\zeta(\mathbf{W})}(t \mid t_0)$ is a consistent estimator of $S_{\zeta(\mathbf{W})}(t \mid t_0)$, $S(t \mid t_0)$ can be estimated consistently with $n_{t_0}^{-1}\sum_{i\in\Omega_{t_0}} \hat{S}_{\zeta(\mathbf{w}_i)}(t|t_0)$ where $n_{t_0}=\sum_{i=1}^{n}I(X_{\mathbb{L}i}>t_0)$. When $\mathbf{T}_{\mathbb{S}i}$ consists of a single intermediate event and $\mathbf{Z}$ consists of a single discrete marker, the nonparametric estimation procedure proposed in Parast et al. (2011) can be used to obtain a consistent estimator of $S_{\mathbf{W}}(t \mid t_0)$ with $\zeta(\mathbf{W}) = \mathbf{W}$. However, when the dimension of $\mathbf{W}$ is not small, such nonparametric estimation of $S_{\mathbf{W}}(t \mid t_0)$ may not behave well (Robins & Ritov, 1997). Instead, we propose to reduce the dimension of $\mathbf{W}$ by first approximating $S_{\mathbf{W}}(t \mid t_0)$ where $\zeta(\mathbf{W}) = \mathbf{W}$ with a working semiparametric model and use the model to construct a low dimensional function that is then used to derive

an estimator for $S(t \mid t_0)$. A useful example of such a model is the landmark proportional hazards model (Van Houwelingen & Putter, 2008; Van Houwelingen, 2007; Van Houwelingen & Putter, 2012)

$$S_{\mathbf{w}}(t|t_0)=\exp\{-\Lambda_0^{t_0}(t)\exp(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{w})\}, t>t_0 \quad (2\cdot3)$$

where $\Lambda_0^{t_0}(\cdot)$ is the unspecified baseline cumulative hazard function for $T_{\mathbb{L}i}$ among Ω_{t_0} and $\boldsymbol{\beta}$ is an unknown vector of coefficients. This is referred to as our Stage 1. Let $\hat{\boldsymbol{\beta}}$ be the maximizer of the corresponding log partial likelihood function and $\hat{\Lambda}_0^{t_0}(t)$ be the Breslow-type estimate of baseline hazard (Breslow, 1972). When (2·3) is correctly specified, $\hat{S}_{\mathbf{w}}(t|t_0)=\exp\{-\hat{\Lambda}_0^{t_0}(t)\exp(\hat{\beta}^{\mathsf{T}}\mathbf{w})\}$ could consistently estimate $S_{\mathbf{W}}(t \mid t_0)$. However, if (2·3) is not correctly specified, $\hat{S}_{\mathbf{W}}(t \mid t_0)$ would not longer be a consistent estimator of $S_{\mathbf{W}}(t \mid t_0)$ and thus $n_{t_0}^{-1}\sum_{i=1}^{n_{t_0}}\hat{S}_{\mathbf{w}_i}(t|t_0)$ may be a biased estimate of $S(t \mid t_0)$ (Lagakos, 1988; O'neill, 1986). On the other hand, the risk score $\hat{U}_i \equiv \hat{\boldsymbol{\beta}^{\mathsf{T}}}\mathbf{W}_i$ may still be highly predictive of $T_{\mathbb{L}i}$. This motivates us to overcome the difficulty arising from model misspecification and multi-dimensionality of $\mathbf{W}_i$ by using $\zeta(\mathbf{W})=\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{W}$. Therefore, $S(t \mid t_0) = E\{S_{U_i}(t \mid t_0)\}$, where $S_u(t \mid t_0) = P(T_{\mathbb{L}i} > t \mid T_{\mathbb{L}i} > t_0, U_i = u)$, $U_i=\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{W}_i$ and $\boldsymbol{\beta}_0$ is the limit of $\hat{\boldsymbol{\beta}}$, which always exists (Hjort, 1992). In essence, we use the working Cox model as a tool to construct

$\zeta(\mathbf{W})=U=\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{W}$ in stage 1 and estimate $S(t \mid t_0)$ as $\hat{S}(t|t_0)=\dfrac{1}{n_{t0}}\sum_{i\in\Omega_{t_0}}\hat{S}_{\hat{U}_i}(t|t_0)$, where $\hat{S}_u(t \mid t_0)$ is a consistent estimator of $S_u(t \mid t_0)$ obtained in stage 2.

To estimate $S_u(t \mid t_0)$ in stage 2, we use a nonparametric conditional Nelson-Aalen estimator (Beran, 1981) based on subjects in Ω_{t_0}. Specifically for any given t and u, we use the synthetic data $\{(X_{\mathbb{L}i}, \delta_{\mathbb{L}i}, \hat{U}_i), i \in \Omega_{t_0}\}$ to obtain a local constant estimator for the conditional hazard $\Lambda_u(t \mid t_0) = -\log S_u(t \mid t_0)$ as

$$\hat{\Lambda}_u(t|t_0)=\int_{t_0}^t \frac{\sum_{i\in\Omega_{t_0}}K_h(\hat{\mathrm{D}}_{ui})dN_i(z)}{\sum_{i\in\Omega_{t_0}}K_h(\hat{\mathrm{D}}_{ui})Y_i(z)}$$

as in Cai et al. (2010), where $Y_i(t) = I(T_{\mathbb{L}i} \quad t)$, $N_i(t) = I(T_{\mathbb{L}i} \quad t)\delta_{\mathbb{L}i}$, $\hat{D_{ui}} = \hat{U}_i - u$, $K(\cdot)$ is a smooth symmetric density function, $K_h(x) = K(x/h)/h$, and $h = O(n^{-\upsilon})$ is a bandwidth with $1/2 > \upsilon > 1/4$. The resulting estimate for $S_u(t \mid t_0)$ is $\hat{S}_u(t \mid t_0) = \exp\{-\hat{\Lambda}_u(t \mid t_0)\}$. The uniform consistency of $\hat{S}_u(t \mid t_0)$ in u for $S_u(t \mid t_0)$ can be shown using similar arguments as in Cai et al. (2010) under mild regularity conditions. Note that this consistency still holds when $U_i=\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{W}_i$ is discrete as long as it takes only a finite number of values (Du & Akritas, 2002). Finally, we propose to estimate $S(t \mid t_0)$ as

$$\hat{S}(t|t_0)=\frac{1}{n_{t_0}}\sum_{i\in\Omega_{t_0}}\hat{S}_{\hat{U}_i}(t|t_0)=\frac{1}{n_{t_0}}\sum_{i\in\Omega_{t_0}}\exp\{-\hat{\Lambda}_{\hat{U}_i}(t|t_0)\}.$$

The uniform consistency of $\hat{S}_u(t \mid t_0)$ along with the consistency of $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$ ensures the consistency of $\hat{S}(t \mid t_0)$ for $S(t \mid t_0)$. This approach is similar to the semiparametric dimension reduction estimation procedure proposed in Hu et al. (2011) and Hu et al. (2010) for estimating the conditional mean response in the presence of missing data.

Now that we have obtained an estimate of $S(t|t_0)$ in (2·1), an estimate for $S(t_0)$ follows similarly from this same two-stage procedure replacing $\mathbf{W}_i$ with $\mathbf{Z}_i$ and Ω_{t_0} with all patients. More specifically, let $\boldsymbol{\beta}^*$ be the maximizer of the partial likelihood function corresponding to the Cox model:

$$P(T_{\mathbb{L}i} \geq t|\mathbf{Z}_i) = \exp\{-\Lambda_0(t)\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{Z}_i)\} \quad (2\cdot4)$$

where $\Lambda_0(\cdot)$ is the baseline cumulative hazard function and $\boldsymbol{\beta}^*$ is a vector of coefficients. We then define $\hat{U}_i^* \equiv \hat{\beta}^{*\mathrm{T}}\mathbf{Z}_i$ and use a local constant estimator, $\hat{\Lambda}_u(t_0)$, to obtain

$\hat{S}(t_0) = \frac{1}{n}\sum_{i=1}^{n} \hat{S}_{\hat{U}_i^*}(t_0)$ where $\hat{S}_{\hat{U}_i^*}(t_0) = \exp\{-\hat{\Lambda}_{\hat{U}_i^*}(t_0)\}$ is a consistent estimator of $S_u(t_0) = P(T_{\mathbb{L}i} > t_0|U_i^* = u)$. An estimate for our primary quantity of interest $S(t)$ incorporating intermediate event and covariate information follows as $\hat{S}_{\mathrm{LM}}(t) \equiv \hat{S}(t \mid t_0)\hat{S}(t_0)$ where LM indicates that we have used a landmark time to decompose our estimate into two components.

2·2 Theoretical Properties of the Two-Stage Estimator

In this section, we provide theoretical results on how the proposed estimator compares to the standard KM estimator denoted by $\hat{S}_{\mathrm{KM}}(t)$. We also aim to investigate how various components of the proposed procedure contribute to the gain in efficiency. To this end, we first show in Supplementary Appendix A that the variability of $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ can be ignored when making inference about $\hat{S}_{\mathrm{LM}}(t)$. In Appendix I, we show that the influence functions of $\hat{S}(t_0)$, $\hat{S}(t|t_0)$, and $\hat{S}_{\mathrm{LM}}(t)$ can be decomposed as the influence function corresponding to the full data without censoring plus an additional component accounting for variation due to censoring.

To examine the potential gain in efficiency of our proposed estimator over the KM estimator, in Appendix II, we show that the reduction in variance is always non-negative taking the form

$$\frac{S(t_0)}{S_C(t_0)}\int_{t_0}^{t}\left[E\{S_{U_i}(s|t_0)[S_{U_i}(t|s)-S(t|s)]^2\}\right]dG(s|t_0) + S(t|t_0)^2\int_{0}^{t_0}\left[E\left\{S_{U_i^*}(s)[S_{U_i^*}(t_0|s)-S(t|s)]^2\right\}\right]dG(s)$$

as defined in Appendix II. Obviously, there would not be any gain in the absence of censoring with $G(t) \equiv 1$. The maximum reduction is achieved when $S_{U_i}(t \mid t_0) = P(T_{\mathbb{L}i} \quad t \mid T_{\mathbb{L}i} \quad t_0, \mathbf{W}_i)$ and $S_{U_i^*}(s) = P(T_{\mathbb{L}i} \geq s|\mathbf{Z}_i)$. This condition for optimal efficiency is parallel to the recoverability condition described in Hu et al. (2010) for conditional mean response estimation, and is satisfied when the working models in (2·3) and (2·4) are correctly specified. It is important to note that the optimality of the proposed estimator is within the class of estimators constructed based on the specific choice of $\mathbf{W}$ and thus depends on t_0. In Sections 4 and 5, we examine the sensitivity of our estimator to the choice of t_0 and propose data-driven rules to further improve the robustness of our approach by combining information from multiple t_0s.

To shed light on where the gain in efficiency is coming from, we show in Supplementary Appendix B.1 that $\hat{S}_{\mathrm{LM}}(t) \approx \hat{S}_{\mathrm{KM}}(t) + \hat{\mathcal{R}}(t)$, where

$$\hat{\mathscr{R}}(t) = n_{t_0}^{-1} \sum_{i \in \Omega_{t_0}} \left\{ S(t_0) \int_{t_0}^{t} \left[S_{U_i}(t|t_0) \frac{dM_{C_i}(s|t_0)}{\pi_{U_i}(s|t_0)} \right. \right.$$

$$- S(t|t_0) \frac{dM_{C_i}(u|t_0)}{\pi(u|t_0)} + n^{-1} \sum_{i=1}^{n} \left\{ S(t|t_0) \int_{0}^{t_0} \left[S_{U_i^*}(t_0) \frac{dM_{C_i}(s)}{\pi_{U_i^*}(s)} \right. \right.$$

$$\left. \left. \left. - S(t_0) \frac{dM_{C_i}(u)}{\pi(u)} \right] \right\} \right\} .$$

That is, our proposed estimator can also be expressed as the KM estimator augmented by a mean zero martingale $\hat{\mathscr{R}}(t)$ corresponding to censoring weighted by baseline covariate and intermediate event information. The term $\hat{\mathscr{R}}(t)$ is negatively correlated with $\hat{S}_{KM}(t)$ and improves efficiency by leveraging independent censoring and the correlation between T_{L_i} and $\{\mathbf{T}_{S_i}, \mathbf{Z}_i\}$. The proposed two-stage modeling framework allows us to optimally use this additional information while protecting us against model misspecification. This decomposition also connects our proposed estimator to the existing augmentation procedures such as those considered in Lu & Tsiatis (2008). The advantage of our proposal is that the modeling step naturally estimates the optimal bases without necessitating the correct specification of the model. Variance estimates can be obtained using a perturbation-resampling procedure as described in Supplementary Appendix C.

3 Comparing Survival Between Two Treatment Groups

Consider a two-arm RCT setting where we use $\mathscr{G}$ to denote treatment assignment with $\mathscr{G}$ taking a value B with probability p and A otherwise. For $\mathscr{G} \in \{A, B\}$, let $S_{\mathscr{G}}(t)$ denote the survival rate of T_L at time t in group $\mathscr{G}$ and $\hat{S}_{LM,\mathscr{G}}(t)$ denote the aforementioned two-stage estimator based on data from treatment group $\mathscr{G}$. An estimator for the risk difference, $\Delta(t) = S_A(t) - S_B(t)$, may be obtained as $\hat{\Delta}_{LM}(t) = \hat{S}_{LM,A}(t) - \hat{S}_{LM,B}(t)$. The standard error of $\hat{\Delta}_{LM}(t)$ can be estimated as $\sigma(\hat{\Delta}_{LM}(t))$ using a perturbation-resampling procedure as described in Supplementary Appendix C. A normal confidence interval (CI) for $\Delta(t)$ may be constructed accordingly. To test the null hypothesis of $H_0 : \Delta(t) = 0$, a Wald-type test may be performed based on $Z_{LM}(t) = \hat{\Delta}_{LM}(t)/\sigma(\hat{\Delta}_{LM}(t))$. To compare this testing procedure to a test based on $\hat{S}_{KM}(t)$ let $\Delta_{KM}(t) = \hat{S}_{KM,A}(t) - \hat{S}_{KM,B}(t)$ denote the KM estimate for $\Delta(t)$ where $\hat{S}_{KM,\mathscr{G}}(t)$ is the KM estimate of survival at time t in group $\mathscr{G}$.

Through an augmentation procedure in a RCT setting, Lu & Tsiatis (2008) demonstrated substantial gains in power when testing the log hazard ratio under a Cox model. Though Lu & Tsiatis (2008) does not offer procedures to estimate or test for a difference in t-year survival rates, the investigation of their useful and effective procedure is worthwhile in our setting. The augmentation procedure in Lu & Tsiatis (2008) (LT) leads to significant gains in efficiency by leveraging two key sets of assumptions: (1) C is independent of $\{T_L, \mathbf{Z}\}$ given $\mathscr{G}$ and (2) $\mathscr{G}$ is independent of $\mathbf{Z}$. The LT procedure takes advantage of such information by explicitly creating functional bases and augmenting over both censoring martingales and $I(\mathscr{G} = B) - p$, weighted by these specified bases. Similar to the arguments given in Section 2 and Supplementary Appendix B.1, one may interpret $\hat{\Delta}_{LM}(t)$ as augmenting $\hat{\Delta}_{KM}(t)$ with an optimal basis of censoring martingales approximated by the working models. However, our procedure does not require explicit specification of the basis and also leverages information on C being independent of $\mathbf{T}_S$.

However, $\hat{\Delta}_{LM}(t)$ does not yet leverage information on the randomization of treatment assignment (that is, $\mathscr{G}$ is independent of $\mathbf{Z}$) and hence it is possible to further improve this

estimator via augmentation. One may follow similar strategies as the LT procedure by explicitly augmenting over certain basis functions of $\mathbf{Z}$. Specifically, let $\mathbf{H}(\mathbf{Z})$ be any arbitrary vector of basis functions. An augmented estimator can be obtained as $\hat{\Delta}_{\text{LM}}(t; \varepsilon, H) = \hat{\Delta}_{\text{LM}}(t) + \varepsilon^{\top} {}_i\{I(\mathcal{G}_i = B) - p\}\mathbf{H}(\mathbf{Z}_i)$, where ε is chosen to minimize the variance of $\hat{\Delta}_{\text{LM}}(t; \varepsilon, H)$. However, a poor choice of bases and/or a large number of non-informative bases might result in a loss of efficiency in finite samples (Tian et al., 2012). Here, we propose to explicitly derive an optimal basis using the form of the influence function for $\hat{\Delta}_{\text{LM}}(t)$ and subsequently use our two-stage procedure to estimate the basis. Specifically, in Supplementary Appendix B.2 we derive an explicit form for the optimal choice for $\mathbf{H}(\mathbf{Z})$ in this setting as

$$\mathbf{H}_{\text{opt}}(\mathbf{Z}, t) = - \sum_{\mathcal{G} \in \{A,B\}} \frac{S_{\mathbf{Z},\mathcal{G}}(t) - S_{\mathcal{G}}(t)}{(1-p)^{I(\mathcal{G}=A)} p^{I(\mathcal{G}=B)}} \quad (3\cdot1)$$

where $S_{\mathbf{Z}_i,\mathcal{G}}(t) \equiv P(T_{\mathbb{L}i} > t \mid \mathbf{Z}_i, \mathcal{G}_i = \mathcal{G})$ for $\mathcal{G} = A, B$. There are various ways one may estimate $(3\cdot1)$. We choose to guard against possible model misspecification and consider a robust estimate using the proposed two-stage procedure. Specifically, we first fit a Cox model to treatment group $\mathcal{G}$ using only $\mathbf{Z}$ information to obtain an estimated log hazard ratio, $\hat{\beta}_{\mathcal{G}}^*$, and then approximate $-\log S_{\mathbf{Z},\mathcal{G}}(t)$ as $\hat{\Lambda}_{\mathbf{z}'\hat{\beta}_{\mathcal{G}}^*,\mathcal{G}}(\hat{\beta}_{\mathcal{G}}^*, t)$, where $\Lambda_{u,\mathcal{G}}(\boldsymbol{\beta}, t)$ is obtained similarly to $\Lambda_u(\boldsymbol{\beta}, t)$ but only using data in the treatment group $\mathcal{G}$. We denote the resulting estimated basis as $\hat{\mathbf{H}}_{\text{opt}}(\mathbf{Z}_i, t)$. We then obtain an augmented estimator for $\Delta(t)$ as $\hat{\Delta}_{\text{AUG}}(t) = \hat{\Delta}_{\text{LM}}(t) - \hat{\varepsilon} {}_i\{I(\mathcal{G}_i = B) - p\}\hat{\mathbf{H}}_{\text{opt}}(\mathbf{Z}_i, t)$, where $\hat{\varepsilon}$ is an estimate of $\text{cov}\{\hat{\Delta}_{\text{LM}}(t), {}_i\{I(\mathcal{G}_i = B) - p\}\hat{\mathbf{H}}_{\text{opt}}(\mathbf{Z}_i, t)\}\text{var}\{ {}_i\{I(\mathcal{G}_i = B) - p\}\hat{\mathbf{H}}_{\text{opt}}(\mathbf{Z}_i, t)\}\}^{-1}$ which can be approximated using the perturbation procedure as described in Supplementary Appendix C. Note that we use the proposed two-stage procedure to approximate the optimal basis for augmentation. Therefore, when the working models do not hold, the *estimated* basis may not be optimal but can still improve efficiency empirically by capturing important associations between $T_{\mathbb{L}i}$ and $\mathbf{Z}_i$. In addition, we estimate ε empirically as $\hat{\varepsilon}$ to guard against model misspecification. Since the variability due to $\hat{\mathbf{H}}_{\text{opt}}$ and $\hat{\varepsilon}$ has a negligible contribution asymptotically to the variance of $\hat{\Delta}_{\text{AUG}}(t)$, such calibration and weighting improve the robustness of our proposed approach. Hence, this approach could ultimately improve the efficiency of $\hat{\Delta}_{\text{AUG}}(t)$ even in the presence of model misspecification.

4 Simulations

We conducted simulation studies to examine the finite sample properties of the proposed estimation procedures. If a gain in efficiency is observed over the KM estimate, one may consider whether this advantage is due in part to incorporating covariate information or incorporating $\mathbf{T}_{\mathbb{S}}$ information or both. To address this question, let $\hat{S}_{\text{LM}}^{T_{\mathbb{S}}}(t) = \hat{S}^{T_{\mathbb{S}}}(t|t_0)\hat{S}_{\text{KM}}(t_0)$ denote an estimate of $S(t)$ incorporating only $\mathbf{T}_{\mathbb{S}}$ information where $\tilde{S}^{T_{\mathbb{S}}}(t \mid t_0)$ is obtained using the two-stage procedure with $\mathbf{W} = [I(\mathbf{T}_{\mathbb{S}i} \quad t_0), \min(\mathbf{T}_{\mathbb{S}i}, t_0)]^{\top}$. In addition, let $\tilde{S}^{\mathbf{Z}}(t)$ denote an estimate of $S(t)$ incorporating only $\mathbf{Z}$ information. In our numerical examples, we additionally estimate these quantities for comparison. In an effort to compare our results to those examined previously in the literature, we apply our method to similar simulation settings as in Gray (1994). Three settings were considered: (i) a no treatment effect setting i.e. null setting (ii) a small treatment effect setting (iii) a large treatment effect setting. For illustration, $t_0 = 1$ year and $t = 2$ years i.e. we are interested in the probability of survival past 2 years. In all settings, censoring, C_i, was generated from a Uniform(0.5, 2.5) distribution. For each treatment group, n=1000 and results summarize 2000 replications under the null setting (i) and 1000 replications under settings (ii) and (iii). Variance estimates are obtained using the perturbation-resampling method. Under the null

setting (i), the event times for each treatment group (Treatment A and Treatment B) were generated from a Weibull distribution with scale parameter b_1^{-1/a_1} and shape parameter $1/a_1$ i.e. as $T_\mathbb{S} = [-b_1^{-1} \log(1 - U_1)]^{1/a_1}$ where $a_1 = 1.5$, $b_1 = 1$, and $U_1 \sim \text{Uniform}(0,1)$. The conditional distribution $T_\mathbb{L} - T_\mathbb{S} \mid T_\mathbb{S}$ was generated from a Weibull distribution with scale parameter b_2^{-1/a_2} and shape parameter $1/a_2$ i.e. as $[-b_2^{-1} \log(1 - U_2)]^{1/a_2}$ where $a_2 = 1.5$, $b_2 = \exp(g(T_\mathbb{S}))$, $g(t) = 0.15 - 0.5t^2$ and $U_2 \sim \text{Uniform}(0,1)$. We take the covariate Z to be $0.75U_2 + .25U_3$ where $U_3 \sim \text{Uniform}(0,1)$. Note that in these simulations, $T_\mathbb{S}$ must occur before $T_\mathbb{L}$ though our procedures do not require this assumption to hold. In this setting, $P(X_{\mathbb{L}i} < t, \delta_{\mathbb{L}i} = 1) = 0.37$ and $P(X_{\mathbb{S}i} < t_0, \delta_{\mathbb{S}i} = 1) = 0.59$ in each group. In setting (ii) the Treatment A event times were generated as shown above and the Treatment B event times were generated similarly but with $b_1 = 1.2$. In this setting, $P(X_{\mathbb{L}i} < t, \delta_{\mathbb{L}i} = 1) = 0.39$ and $P(X_{\mathbb{S}i} < t_0, \delta_{\mathbb{S}i} = 1) = 0.65$ in the Treatment B group. Finally, in setting (iii) the Treatment A event times were generated as shown above and the Treatment B event times were generated similarly but with $b_1 = 1.4$. In this setting, $P(X_{\mathbb{L}i} < t, \delta_{\mathbb{L}i} = 1) = 0.43$ and $P(X_{\mathbb{S}i} < t_0, \delta_{\mathbb{S}i} = 1) = 0.70$ in the Treatment B group.

As in most nonparametric functional estimation procedures, the choice of the smoothing parameter h is critical. In our setting, the ultimate quantity of interest is the marginal survival and the estimator is effectively an integral of the conditional survival estimates. To eliminate the impact of the bias of the conditional survival function on the resulting estimator, we also require the standard undersmoothing assumption of $h = O(n^{-\delta})$ with $\delta \in (1/4, 1/2)$. To obtain an appropriate h we first use the bandwidth selection procedure given by Scott (1992) to obtain h_{opt}, and then we let $h = h_{opt} n^{-c_0}$ for some $c_0 \in (1/20, 3/10)$ to ensure the desired rate for h. In all numerical examples, we chose $c_0 = 0.10$.

Results from estimation of $S(t)$ are shown in Table 1. The standard error estimates obtained by the perturbation-resampling procedure approximate the empirical standard error estimates well. While all estimators have negligible bias and satisfactory coverage levels, our proposed estimate which incorporates both $\mathbf{Z}$ and $T_\mathbb{S}$ information is more efficient than that from the standard Kaplan Meier estimate with relative efficiencies with respect to mean squared error (MSE) around 1.20. Though extensive simulation studies may be performed, we expect that such efficiency gains would not be seen in settings with little censoring or with weak correlation between the intermediate and terminal event (Cook & Lawless, 2001). Note that in all three settings, $C_i \sim \text{Uniform}(0.5, 2.5)$ and thus $P(C < t_0) = 0.25$ and $P(C < t) = 0.75$. To examine efficiency gains under varying degrees of censoring, we simulated data as in setting (i) but with (a) more censoring where $C_i \sim \text{Uniform}(0.5, 2.15)$ and (b) less censoring $C_i \sim \text{Uniform}(0.5, 3.5)$. In (a) $P(C < t_0) = 0.30$ and $P(C < t) = 0.90$ and the relative efficiency with respect to MSE was 1.40 (not shown). In (b) $P(C < t_0) = 0.17$ and $P(C < t) = 0.50$ and the relative efficiency with respect to MSE was 1.10 (not shown).

To examine the sensitivity of our estimator to the choice of t_0 and to further improve the robustness of our procedure by combining information from multiple t_0s, we estimate $S(t)$ with $t_0 \in \mho = \{k_{0.25}, k_{0.50}, k_{0.75}\}$, where k_p is the pth percentile of $\{X_{\mathbb{S}i} : \delta_{\mathbb{S}i} = 1\}$, in setting (i). We denote the resulting estimators and corresponding standard errors estimated using the perturbation-resampling procedure as $\hat{S}_{k0.25}(t)$, $\hat{S}_{k0.50}(t)$, $\hat{S}_{k0.75}(t)$ and $\hat{\sigma}_{k0.25}$, $\hat{\sigma}_{k0.50}$, $\hat{\sigma}_{k0.75}$, respectively. One potential data-driven approach to choose t_0 would be to set t_0 as $\hat{t}_0 = \arg \min_{t_0 \in \mho} \hat{\sigma}_{t_0}$ and let $\hat{S}_{\hat{t}_0}(t)$ denote the corresponding estimator of $S(t)$. Alternatively, one could consider an optimal linear combination of $\hat{\mathbf{S}}_{\mathbf{t_0}}(t) = (\hat{S}_{k0.25}(\mathbf{t}), \hat{S}_{k0.50}(\mathbf{t}), \hat{S}_{k0.75}(t))^\mathsf{T}$, $\hat{S}_{\text{comb}}(t) \equiv \mathbf{w}_{\text{comb}}^\mathsf{T} \hat{\mathbf{S}}_{\mathbf{t_0}}(\mathbf{t})$, where $\mathbf{w}_{\text{comb}} = \mathbf{1}^\mathsf{T} \hat{}^{-1} [\mathbf{1}^\mathsf{T} \hat{}^{-1} \mathbf{1}]^{-1}$ and $\hat{}$ is the estimated covariance matrix of $\hat{\mathbf{S}}_{\mathbf{t_0}}(\mathbf{t})$ estimated using the perturbation-resampling procedure. The variance of $\hat{S}_{\text{comb}}(t)$ can be estimated as $\mathbf{1}^\mathsf{T} \hat{}^{-1} \mathbf{1}$. Due to the high collinearity of $\hat{\mathbf{S}}_{\mathbf{t_0}}(\mathbf{t})$, we expect that in

finite samples, $\hat{}^{-1}$ may be an unstable estimate of $\mathrm{var}\{\hat{\mathbf{S}}_{\mathbf{t_0}}(\mathbf{t})\}^{-1}$. Consequently $\hat{S}_{\mathrm{comb}}(t)$ may not perform well and $\mathbf{1}^{\mathsf{T}}\hat{}^{-1}\mathbf{1}$ may be under-estimating the true variance of $\hat{S}_{\mathrm{comb}}(t)$. Another alternative would be an inverse variance weighted (IVW) estimator of $S(t)$,

$\hat{S}_{\mathrm{IVW}}(t) \equiv \mathbf{w}_{\mathrm{IVW}}^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{t_0}}(\mathbf{t})$ where $\mathbf{w}_{\mathrm{IVW}} = \mathbf{1}^{\mathsf{T}}\mathrm{diag}(\hat{\Sigma})^{-1}[\mathbf{1}^{\mathsf{T}}\mathrm{diag}(\hat{\Sigma})^{-1}\mathbf{1}]^{-1}$. Under the current simulation setting (ii), the empirical standard errors of $\hat{S}_{k0.25}(t)$, $\hat{S}_{k0.50}(t)$, $\hat{S}_{k0.75}(t)$, $\hat{S}_{\hat{t}_0}(t)$, $\hat{S}_{\mathrm{comb}}(t)$, and $\hat{S}_{\mathrm{IVW}}(t)$ are 0.0215, 0.0210, 0.0207, 0.0210, 0.0206, and 0.0206 respectively.

Although asymptotically the variance of $\hat{S}_{\hat{t}_0}(t)$ should be the same as that of $\hat{S}_{k0.75}(t)$, due to the similarity among $\{\hat{\sigma}_{t_0}, t_0 \in \mho\}$, the variability in $\hat{t}_0$ does contribute additional noise to $\hat{S}_{\hat{t}_0}(t)$. As expected, using $\mathbf{1}^{\mathsf{T}}\hat{}^{-1}\mathbf{1}$, the average variance of $\hat{S}_{\mathrm{comb}}(t)$ is estimated as 0.0202, which is slightly lower than the empirical variance of $\hat{S}_{\mathrm{comb}}(t)$. For $\hat{S}_{\mathrm{IVW}}(t)$, the average of the estimated variance is 0.0204, which is close to the empirical variance. The empirical coverage levels of the 95% confidence intervals using $\hat{S}_{\mathrm{comb}}(t)$ and $\hat{S}_{\mathrm{IVW}}(t)$ are 0.945 and 0.943, respectively, indicating that such estimators perform well when combining across a small number of choices for t_0. This suggests that our proposed estimators are not overly sensitive to the choice of t_0 and a reasonable approach in practice may be to estimate $S(t)$ using the proposed combined or IVW estimator.

For comparing treatment groups, we show results on the estimation precision and power in testing in Table 2. Similar to above, let $\hat{\Delta}^{\mathbf{Z}}(t)$ denote the estimate of $\Delta(t)$ using $\mathbf{Z}$ information only and $\hat{\Delta}_{\mathrm{LM}}^{T_{\mathbb{S}}}(t)$ denote the estimate using $\mathbf{T}_{\mathbb{S}}$ information only. Corresponding test statistics can then be constructed similar to $\Delta_{\mathrm{LM}}(t)$. With respect to estimation precision for the treatment difference in survival rate, our proposed procedure yields about 49% gain in efficiency compared to the KM estimator. For testing for a treatment difference, under the null setting (i), the type I error is close to the nominal level of 0.05 with $\hat{\Delta}_{\mathrm{AUG}}^{\mathbf{w}}(t)$ having a slightly higher value of 0.0538. When there is a small treatment difference, setting (ii), the power to detect a difference in survival at time $t = 2$ is 0.374 using a standard Kaplan Meier estimate and 0.457 when using the proposed procedure incorporating $\mathbf{Z}$ and $T_{\mathbb{S}}$ information. We gain more power through additional augmentation which results in a power of 0.520. Similarly, in setting (iii), when there is a large treatment difference, power increases from 0.846 to 0.941 after incorporating $\mathbf{Z}$ and $T_{\mathbb{S}}$ information and augmentation. These results suggest that in this setting, we can gain substantial efficiency by using the proposed procedures.

5 Example

We illustrate the proposed procedures using a dataset from the AIDS Clinical Trial Group (ACTG) Protocol 175 (Hammer et al., 1996). This dataset consists of 2467 patients randomized to 4 different treatments: zidovudine only, zidovudine + didanosine, zidovudine + zalcitabine, and didanosine only. The long term event of interest, $T_{\mathbb{L}}$, is the time to death and intermediate event information consists of two intermediate events, $\mathbf{T}_{\mathbb{S}} = (T_{\mathbb{S}}, T_{\mathbb{S}})^{\mathsf{T}}$ where $T_{\mathbb{S}} =$ time to an AIDS-defining event e.g. pneumocystis pneumonia and $T_{\mathbb{S}} =$ time to a 50% decline in CD4. If a patient experienced multiple intermediate events of one kind, for example multiple AIDS-defining events, the earliest occurrence of the event was used. For illustration, $t_0 = 1$ year and $t = 2.5$ years and we examine survival in patients from the zidovudine only (mono group, n=619) and zidovudine + zalcitabine (combo group, n=620) groups. Figure 2 displays the KM estimate of survival in each group. Baseline covariates, $\mathbf{Z}$, include the mean of two baseline CD4 counts, Karnofsky score, age at randomization, weight, symptomatic status, use of zidovudine in the 30 days prior to randomization, and days of antiretroviral therapy before randomization. Results are shown in Table 3. Our proposed procedure leads to a t-year survival rate estimate of 0.926 for the mono group and

0.953 for the combo group. The KM estimate of survival is similar though smaller in both groups.

To examine the sensitivity of our estimator to the choice of t_0 in this example, we estimate $S(t)$ with $t_0 \in \mho = \{k_{0.25}, k_{0.50}, k_{0.75}\} = \{392, 582, 818\}$, where k_p is the pth percentile of $\{X_{\mathbb{S}i} : \delta_{\mathbb{S}i} = 1, X_{\mathbb{S}i} : \delta_{\mathbb{S}i} = 1\}$, $X_{\mathbb{S}ji} = \min(T_{\mathbb{S}ji}, C_i)$ and $\delta_{\mathbb{S}ji} = I(T_{\mathbb{S}ji} \quad C_i)$ for $j = 1, 2$ in the zidovudine only group. Note that when $t_0 = 365$ as used above, $\hat{S}(t) = 0.9260$ with estimated standard error equal to 0.0105. With $t_0 \in \mho$, our estimates of S(t) are 0.9272, 0.9262, and 0.9262 and estimates of standard error are 0.0106, 0.0105, and 0.0107, respectively. As suggested above, a reasonable alternative approach would be to estimate $S(t)$ using the proposed combined or IVW estimator. Here, $\hat{S}_{\text{comb}}(t) = 0.9265$ and $\hat{S}_{\text{IVW}}(t) = 0.9266$ with standard error estimates 0.0105 and 0.0105, respectively. Therefore, we find that our estimates of survival and corresponding standard error are not sensitive to the choice of t_0 in this example and that both of the combined procedures give similar results on both the point and interval estimators.

We are interested in testing $H_0 : \Delta(2.5) \equiv S_{\text{mono}}(2.5) - S_{\text{combo}}(2.5) = 0$. Results for comparing the two treatment groups are shown in Table 4. Point estimates of $\Delta(2.5)$ from various procedures are reasonably close to each other with the KM estimator being 0.0255 and our proposed estimators being 0.0271 and 0.0278 without and with further augmentation, respectively. The proposed procedure provides an estimate which is roughly 19% more efficient than that from the KM estimate. The p-value for this test decreased from 0.0942 to 0.0533 after incorporating information on baseline covariates and $\mathbf{T}_{\mathbb{S}}$. Furthermore, augmentation in addition to the two-stage procedure provides an estimate approximately 49% more efficient than the KM estimate and a p-value of 0.0267.

6 Remarks

We have proposed a semi-nonparametric two-stage procedure to improve the efficiency in estimating $S(t)$ in a RCT setting. We demonstrate with both theoretical and numerical studies that there are substantial potential efficiency gains in both estimation and testing of t-year survival. Our approach can easily incorporate multiple intermediate events, which is appealing when multiple recurrent event information is useful for predicting the terminal events.

This procedure uses the landmark concept developed in Van Houwelingen (2007), Van Houwelingen & Putter (2012) and Parast et al. (2012) to leverage information on multiple intermediate events and baseline covariates, which overcomes complications that arise in a semi-competing risk setting. Existing work on landmark prediction has focused primarily on the prediction of residual life while this paper focuses on improving the estimation of marginal survival. Model misspecification may lead to inaccurate prediction, but would not void the validity of many of the previously proposed prediction procedures. On the contrary, model misspecification of the conditional survival could lead to a biased estimate of the marginal survival. To guard against model misspecification, we use working models to estimate the conditional survival with an additional layer of calibration which ensures the consistency of the marginal survival estimates. We derive asymptotic expansions of the proposed estimator to provide insight into why this method would always improve efficiency compared to the KM estimator and when we would expect to achieve optimal efficiency. The proposed approach can be seen as being related to Bang & Tsiatis (2000) which proposes a partitioned estimator that makes use of the cost history for censored observations to estimate overall medical costs and to Hu et al. (2011) and Hu et al. (2010) which use a semiparametric dimension reduction procedure to estimate the conditional mean response in the presence of missing data.

To avoid any bias due to model specification we use the Cox model only as a tool for dimension reduction followed by a nonparametric kernel Nelson-Aalen estimator of survival to obtain a consistent final estimate of survival. Two conditions required for the consistency of $\hat{S}_u(t|t_0)$ are independent censoring and the score derived under a given working model, $\hat{U} = \hat{\beta}'\mathbf{W}$, must converge to a deterministic score $U = \beta'\mathbf{W}$. This convergence would require the objective function used to derive $\hat{\beta}$ to have a unique optimizer. Such an objective function could involve convex functions such as a partial likelihood or an estimating functions with a unique solution such as those considered in Tian et al. (2007) and Uno et al. (2007). In this paper, we have used the Cox model as a working model to obtain $\hat{\beta}$ and it has been shown that the Cox partial likelihood has a unique solution even under model misspecification (Hjort, 1992; Van Houwelingen, 2007). Though misspecification of the working model will not affect consistency, it may affect the efficiency of the proposed estimator.

In general, our methods are not restricted to the Cox model being the working model; any working model with this property along with providing root-n convergence of $\hat{\beta}$ would be valid and ensure consistency of the final estimate. For example, a proportional odds model or other semi-parametric transformation model could be considered as long as this property holds. However, it is unclear if the maximizer of the rank correlation under the single-index models or the non-parametric transformation model for survival data (Khan & Tamer, 2009) is unique under model misspecification. Therefore, some care should be taken in the choice of the working models. It is interesting to note that it has been shown in some non-survival settings that generalized linear model-type estimators are robust to misspecification of the link function and that the resulting coefficients can be estimated consistently up to a scalar (Li & Duan, 1989). In such settings, even under model misspecification, we would not expect any loss of efficiency due to the link misspecification. Further investigation of this property in a survival setting would be valuable for future research.

The ability to include multiple intermediate events is a desirable property of this proposed procedure. In settings with a small number of potential intermediate events, including all intermediate event information in stage 1 would not affect the efficiency of the overall estimate due to the fact that the variability in $\hat{\beta}$ does not contribute to the asymptotic variance (shown in Supplementary Appendix A). However, in finite samples, this additional noise could be problematic and in settings with many potential intermediate events, a regularization procedure (Friedman et al., 2001; Tian et al., 2012) may be considered for variable selection in stage 1.

As in most nonparametric functional estimation procedures, a sufficiently large sample size is needed to obtain stable estimation performance and observe efficiency gains. The degree of efficiency gain using this proposed procedure may be less than expected in small sample sizes. Alternative robust methods should be considered in such settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix

Asymptotic properties of $\hat{S}_{LM}(t)$

We explicitly examine the asymptotic properties of our estimate of survival compared to the KM estimate of survival. Recall that our proposed estimate $\hat{S}_{LM}(t) = \hat{S}(t \mid t_0)\hat{S}(t_0)$. Let $\hat{\beta}$ and $\hat{\beta}^*$ be the maximizers of the log partial likelihood functions corresponding to the proportional hazards working models (2·3) and (2·4), respectively. In addition, let β_0 and β_0^* denote the limits of $\hat{\beta}$ and $\hat{\beta}^*$, respectively. Let

$\hat{S}(\beta^*, t_0) = n^{-1}\sum_{i=1}^{n} \hat{S}_{\mathbf{z}_i'\beta^*}(\beta^*, t_0), \hat{S}(\beta, t \mid t_0) = n_{t_0}^{-1}\sum_{i=1}^{n_{t_0}} \hat{S}_{\mathbf{w}_i'\beta}(\beta, t \mid t_0)$, and $\hat{S}_{LM}(\beta^*, \beta, t) = \hat{S}(\beta^*, t_0)\hat{S}(\beta, t \mid t_0)$, where $\hat{S}_u(\beta^*, t_0) = e^{-\hat{\Lambda}_u(\beta^*, t_0)}$, $\hat{S}_v(\beta, t \mid t_0) = e^{-\hat{\Lambda}_v(\beta, t \mid t_0)}$, $\hat{\Lambda}_u(\beta^*, t_0)$ and $\hat{\Lambda}_v(\beta, t \mid t_0)$ are obtained by replacing $\hat{\beta}^*$ and $\hat{\beta}$ in $\hat{\Lambda}_u(t_0)$ and $\hat{\Lambda}_v(t \mid t_0)$ by β^* and β, respectively. Then $\hat{S}(\hat{\beta}^*, t_0) = \hat{S}(t_0)$, $\hat{S}(\hat{\beta}^*, t \mid t_0) = \hat{S}(t \mid t_0)$, and $\hat{S}_{LM}(\hat{\beta}^*, \hat{\beta}, t) = \hat{S}_{LM}(t)$. Furthermore, with a slight abuse of notation, we define $S_u(\beta^*, t_0) = P(T_{\mathbb{L}i} > t_0 \mid \mathbf{Z}_i'\beta^* = u) = e^{-\Lambda_u(\beta^*, t_0)}$ and $S_v(\beta, t \mid t_0) = P(T_{\mathbb{T}i} > t \mid \mathbf{W}_i'\beta = v, T_{\mathbb{L}i} > t_0) = e^{-\Lambda_v(\beta, t \mid t_0)}$ Hence $S_u(t_0) = S_u(\beta_0^*, t_0)$ and $S_v(t \mid t_0) = S_v(\beta_0, t \mid t_0)$. We will derive asymptotic expansions for

$\hat{\mathsf{S}}_{LM}(t, t_0) \equiv \sqrt{n}\{\hat{S}_{LM}(\hat{\beta}^*, \hat{\beta}, t) - S(t)\}$, which can be decomposed as

$$\sqrt{n}\{\hat{S}_{LM}(\hat{\beta}^*, \hat{\beta}, t) - \hat{S}_{LM}(\beta_0^*, \beta_0, t)\} + \sqrt{n}\{\hat{S}_{LM}(\beta_0^*, \beta_0, t) - S(t)\}.$$

In Appendix I below we will express

$\hat{\mathscr{W}}^{\mathbf{Z}}(t_0) = \sqrt{n}\{\hat{S}(\hat{\beta}^*, t_0) - S(t_0)\}, \hat{\mathsf{S}}^{\mathbf{w}}(t \mid t_0) = \sqrt{n_{t_0}}\{\hat{S}(\hat{\beta}^*, t \mid t_0) - S(t \mid t_0)\}$ and $\hat{\mathscr{W}}_{LM}(t, t_0)$ in martingale notation. Using these martingale representations, in Appendix II we examine the variance of $\hat{S}_{LM}(\hat{\beta}^*, \hat{\beta}, t)$ and show that this variance is always less than or equal to the variance of $\hat{S}_{KM}(t)$ and the reduction is maximized when the assumed working models in (2·3) and (2·4) are true.

Assumption A.1 Censoring, C, is independent of $\{T_{\mathbb{L}}, \mathbf{Z}\}$ given treatment assignment, $\mathscr{G}$.

Assumption A.2 Treatment assignment, $\mathscr{G}$, is independent of $\mathbf{Z}$.

Regularity Conditions (C.1) Throughout, we assume several regularity conditions: (1) the joint density of $\{T_{\mathbb{L}}, T_{\mathbb{S}}, C\}$ is continuously differentiable with density function bounded away from 0 on $[0, t]$, (2) $\mathbf{Z}^{\mathsf{T}}\beta_0^*$ and $\mathbf{W}^{\mathsf{T}}\beta_0$ have continuously differentiable densities and $\mathbf{Z}$ is bounded, (3) $h = O(n^{-\upsilon})$ with $1/4 < \upsilon < 1/2$, (4) $K(x)$ is a symmetric smooth kernel function with continuous second derivative on its support and $\int \dot{K}(x)^2 dx < \infty$, where $\dot{K}(x) = dK(x)/dx$.

Appendix I

Martingale Representation

In Supplementary Appendix A we show that the variability of $\hat{\beta}^*$ and $\hat{\beta}$ is negligible and can be ignored in making inferences on $\hat{S}(\hat{\beta}^*, t_0)$, $\hat{S}(\hat{\beta}^*, t \mid t_0)$ and $\hat{S}_{LM}(\beta_0^*, \beta_0, t)$. Therefore, we focus on the asymptotic expansions of $\hat{\mathscr{W}}^{\tilde{\mathbf{Z}}}(t_0)$, $\tilde{\mathscr{W}}(t \mid t_0)$, and $\mathscr{W}_{LM}(t, t_0)$ with β_0^* and β_0 and express these quantities in martingale notation. Throughout, we will repeatedly use the fact that $n_{t_0}/n \approx S_C(t_0)S(t_0)$, where $S_C(t_0) = P(C > t_0)$.

Theorem 1. *Assume A.1, A.2 and Regularity Conditions (C.1) hold. Note that in Supplementary Appendix A it was shown that $\tilde{\mathscr{W}}_{LM}(t)) \approx \tilde{\mathscr{W}}_{LM}(t, t_0)$.*

$$\tilde{\mathscr{W}}_{LM}(t, t_0) \approx n^{-1/2} \sum_{i=1}^{n} \phi_{1i}(t, t_0) + n^{-1/2} \sum_{i=1}^{n} \phi_{2i}(t, t_0) \quad \text{(I.1)}$$

$$where \phi_{1i}(t, t_0) \equiv \frac{I(X_{\mathbb{L}i} > t_0)}{S_C(t_0)} \left[S_{U_i}(t|t_0) \int_{t_0}^{t} \frac{dM_{C_i}(s|t_0)}{\pi_{U_i}(s|t_0)} + \frac{I(X_{\mathbb{L}i} > t)}{S_C(t|t_0)} - S(t|t_0) \right] \quad \text{(I.2)}$$

$$and \phi_{2i}(t, t_0) = S(t|t_0) \left[S_{U_i^*}(t_0) \int_{0}^{t_0} \frac{dM_{C_i}(s)}{\pi_{U_i^*}(s)} + \frac{I(X_{\mathbb{L}i} > t_0)}{S_C(t_0)} - S(t_0) \right]. \quad \text{(I.3)}$$

and $\tilde{\mathscr{W}}_{LM}(t)$ converges to a Gaussian process indexed by $t \in [t_0, \tau]$.

Proof of Theorem 1. Using a Taylor series expansion and Lemma A.3 of Bilias et al. (1997),

$$\tilde{\mathscr{W}}^{\mathbf{Z}}(t_0) \approx - \int \sqrt{n} \{\hat{\Lambda}_u(\boldsymbol{\beta}_0^*, t_0) - \Lambda_u(t_0)\} S_u(t_0) F(\boldsymbol{\beta}_0^*, du) + n^{-1/2} \sum_{i=1}^{n} \{S_{U_i^*}(t_0) - S(t_0)\}.$$

Let $M_i(s|u) = N_i(s) - \int_0^{t_0} Y_i(s) d\Lambda_u(s)$, $N_i(s) = I(X_{\mathbb{L}i} \leq s, \delta_{\mathbb{L}i} = 1)$, $Y_i(s) = I(X_{\mathbb{L}i} \geq s)$ and $\pi_u(s) = P(X_{\mathbb{L}i} \geq s | U_i^* = u)$. By a martingale representation, $\sqrt{n}\{\hat{\Lambda}_u(\boldsymbol{\beta}_0^*, t_0) - \Lambda_u(t_0)\} = n^{-1/2} \sum_{i=1}^{n} K_h(U_i^* - u) \int_0^{t_0} \pi_u(s)^{-1} dM_i(s|u)$. Using a change of variable where $\psi = \dfrac{u - U_i^*}{h}$, and a Taylor series expansion, let $\xi_i(t_0|u) = \int_0^{t_0} \dfrac{dM_i(s|u)}{\pi_u(s)}$ and $\eta(U_i^* + h\psi) = \xi_i(t_0|U_i^* + h\psi) S_{U_i^* + h\psi}(t_0) f(\boldsymbol{\beta}_0^*, U_i^* + h\psi)$, where $f(\cdot)$ is the probability density function corresponding to $S(\cdot)$,

$$\int \sqrt{n}\{\hat{\Lambda}_u(\boldsymbol{\beta}_0^*, t_0) - \Lambda_u(t_0)\} S_u(t_0) F(\boldsymbol{\beta}_0^*, du)$$

$$= n^{-1/2} \sum_{i=1}^{n} \int K(\psi)[\eta(U_i^*) + \eta'(U_i^*)h\psi + O(h^2)] d\psi \approx n^{-1/2} \sum_{i=1}^{n} \xi_i(t_0|U_i^*) S_{U_i^*}(t_0) f(\boldsymbol{\beta}_0^*, U_i^*)$$

$$= n^{-1/2} \sum_{i=1}^{n} \int_{0}^{t_0} \frac{dM_i(s|U_i^*)}{\pi_{U_i^*}(s)} S_{U_i^*}(t_0)$$

It then follows that

$$\tilde{\mathscr{W}}^{\mathbf{Z}}(t_0) \approx -n^{-1/2} \sum_{i=1}^{n} \int_{0}^{t_0} \frac{dM_i(s|U_i^*)}{\pi_{U_i^*}(s)} S_{U_i^*}(t_0) + n^{-1/2} \sum_{i=1}^{n} \{S_{U_i^*}(t_0) - S(t_0)\}. \quad \text{(I.4)}$$

To understand how $\hat{S}(\boldsymbol{\beta}^*, t_0)$ can be viewed as a KM estimator augmented by censoring martingales, we next express the influence function of $\mathscr{W}^{\mathbf{Z}}(t_0)$ in such a form. From (A.2) in Supplementary Appendix A and using martingale representation once again and integration by parts,

$$\tilde{\mathscr{W}}^{\mathbf{Z}}(t_0) \approx n^{-1/2}\sum \Big[$$

$$-\int_0^{t_0}\frac{S_{U_i^*}(t_0)}{\pi_{U_i^*}(s)}dN_i(s)$$

$$-\int_0^{t_0}\frac{S_{U_i^*}(t_0)}{\pi_{U_i^*}(s)}Y_i(s)\frac{dS_{U_i^*}(s)}{S_{U_i^*}(s)}$$

$$+S_{U_i^*}(t_0) \tag{I·5}$$

$$-S(t_0){=}n^{-1/2}\sum\left[S_{U_i^*}(t_0)\int_0^{t_0}\frac{dM_{C_i}(s)}{\pi_{U_i^*}(s)}\right.$$

$$\left.+\frac{I(X_{\mathbb{L}i}>t_0)}{S_C(t_0)}-S(t_0)\right]$$

where $M_{C_i}(t){=}N_{C_i}(t)-\int_0^t Y_i(s)d\Lambda_C(s)$ and $\Lambda_C(s)=-\log S_C(s)$. Using similar arguments, it can be shown that

$$\tilde{\mathscr{W}}(t|t_0)\approx -n_{t_0}^{-1/2}\sum_{i\in\Omega_{t_0}}\int_{t_0}^t\frac{dM_i(s|U_i,t_0)}{\pi_{U_i}(s|t_0)}S_{U_i}(t|t_0)+n_{t_0}^{-1/2}\sum_{i\in\Omega_{t_0}}\{S_{U_i}(t|t_0)-S(t|t_0)\} \tag{I·6}$$

$$\tilde{\mathscr{W}}(t|t_0)\approx n_{t_0}^{-1/2}\sum_{i\in\Omega_{t_0}}\left[S_{U_i}(t|t_0)\int_{t_0}^t\frac{dM_{C_i}(s|t_0)}{\pi_{U_i}(s|t_0)}+\frac{I(X_{\mathbb{L}i}>t)}{S_C(t|t_0)}-S(t|t_0)\right] \tag{I·7}$$

where $M_i(s|U_i,t_0){=}N_i(s|t_0)-\int_{t_0}^t Y_i(s|t_0)d\Lambda_{U_i}(s|t_0)$ and $\pi_{U_i}(s|t_0)=P(X_{\mathbb{L}i}\ \ \upsilon|U_i=u,X_{\mathbb{L}i}>t_0)$.

Using (I·5), (I·7) and (A.2) in Supplementary Appendix A, we now examine the form of our complete proposed estimator for $S(t)$, $\hat{S}_{\mathrm{LM}}(\hat{\boldsymbol{\beta}}^*,\hat{\boldsymbol{\beta}},t)$, and obtain

$$\tilde{\mathscr{W}}_{\mathrm{LM}}(t,t_0)\approx S(t_0)\tilde{\mathscr{W}}(t|t_0)+S(t|t_0)\tilde{\mathscr{W}}^{\mathbf{Z}}(t_0)\approx n^{-1/2}\sum_{i=1}^n\phi_{1i}(t,t_0)+n^{-1/2}\sum_{i=1}^n\phi_{2i}(t,t_0) \tag{I·8}$$

$$\text{where}\phi_{1i}(t,t_0)\equiv\frac{I(X_{\mathbb{L}i}>t_0)}{S_C(t_0)}\left[S_{U_i}(t|t_0)\int_{t_0}^t\frac{dM_{C_i}(s|t_0)}{\pi_{U_i}(s|t_0)}+\frac{I(X_{\mathbb{L}i}>t)}{S_C(t|t_0)}-S(t|t_0)\right] \tag{I·9}$$

$$\text{and}\phi_{2i}(t,t_0){=}S(t|t_0)\left[S_{U_i^*}(t_0)\int_0^{t_0}\frac{dM_{C_i}(s)}{\pi_{U_i^*}(s)}+\frac{I(X_{\mathbb{L}i}>t_0)}{S_C(t_0)}-S(t_0)\right]. \tag{I·10}$$

It is interesting to note that these two terms, $n^{-1/2}\sum_{i=1}^n\phi_{1i}(t,t_0)$ and $n^{-1/2}\sum_{i=1}^n\phi_{2i}(t,t_0)$, converge to two independent normal distributions and that $E\{\varphi_{1i}(t,t_0)\varphi_{2i}(t,t_0)\}=0$. Since $S_{U_i}(t|t_0)$ is a bounded smooth monotone function in t and $\int_{t_0}^t\frac{dM_{C_i}(s|t_0)}{\pi_{U_i}(s|t_0)}$ is a martingale process, $\left\{S_U(t|t_0)\int_{t_0}^t\frac{dM_{C_i}(s|t_0)}{\pi_{U_i}(s|t_0)}t\in[t_0,\tau]\right\}$ is Donsker. This coupled with the fact that the

simple class $\{I(X_{\mathbb{L}i} > t)|t\}$ is Donsker, implies that $\{\varphi_{ji}(t, t_0)|t\}$, $j = 1, 2$ are Donsker by Theorem 9.30 of Kosorok (2008). Therefore $\hat{\mathscr{W}}_{LM}(t))$ converges to a Gaussian process indexed by $t \in [t_0, \tau]$.

Appendix II

Variance of $\hat{S}_{KM}(\hat{\beta}^*, \hat{\beta}, t)$

In this section we compare the variance of the proposed estimator to that of the KM estimator, as described in the next theorem.

Theorem 2. *Assume A.1, A.2, and Regularity Conditions (C.1) hold. The proposed estimator, $\hat{S}_{LM}(\hat{\beta}^*, \hat{\beta}, t)$, is always at least as efficient as the Kaplan Meier estimate. Specifically,*

$$
\begin{aligned}
\Delta Var_{LM} = {} & n\, Var\{\hat{S}_{KM}(t) \\
& - S(t)\} \\
& - n\, Var\{\hat{S}_{LM}(\hat{\beta}^*, \hat{\beta}, t) \\
& - S(t)\} \approx \frac{S(t_0)}{S_C(t_0)} \int_{t_0}^{t} \left[E\left\{ S_{U_i}(s|t_0)[S_{U_i}(t|s) - S(t|s)]^2 \right\} \right] dG(s|t_0) \\
& + S(t|t_0)^2 \int_{0}^{t_0} E\left\{ S_{U_i^*}(s)[S_{U_i^*}(t_0|s) - S(t_0|s)]^2 \right\} dG(s) \geq 0
\end{aligned}
$$

where ΔVar_{LM} denotes the difference in variance.

Proof of Theorem 2. We first note that it can be shown that $\hat{S}(\hat{\beta}^*, t_0)$ and $\hat{S}(\hat{\beta}, t)$ are asymptotically independent and hence can be handled separately. From (I·4), we can easily examine the variance of $\hat{S}(\hat{\beta}^*, t_0)$ and compare it to the variance of $\hat{S}_{KM}(t_0)$. The resulting form of the difference in variance allows us to make some interesting conclusions regarding the reduction in variance when using the proposed marginal estimate. Note that this result only concerns one component of our complete estimate since $\hat{S}_{LM}(\hat{\beta}^*, \hat{\beta}, t_0) = \hat{S}(\hat{\beta}^*, t_0)\hat{S}(\hat{\beta}, t|t_0)$, however we consider the results and conclusions worthy of discussion. Similar to Murray & Tsiatis (1996), we use a simple application of the conditional variance formula along with martingale properties, integration by parts and independent censoring,

$$
\begin{aligned}
Var\{\hat{\mathscr{W}}^Z(t_0)\} \approx {} & n^{-1}E\left\{ \sum_{i=1}^{n} S_{U_i^*}(t_0)^2 \int_{0}^{t_0} \frac{d\Lambda_{U_i^*}(s)}{\pi_{U_i^*}(s)} \right\} \\
& + Var\{S_{U_i^*}(t_0)\} = n^{-1}\sum_{i=1}^{n} E\left\{ S_{U_i^*}(t_0)G(t_0) \right. \\
& \left. - S_{U_i^*}(t_0)^2 - \int_{0}^{t_0} \frac{G'(s)ds}{S_{U_i^*}(s)} S_{U_i^*}(t_0)^2 \right\} \\
& + Var\{S_{U_i^*}(t_0)\} \\
= {} & S(t_0)G(t_0) \\
& - E\left\{ \int_{0}^{t_0} \frac{G'(s)ds}{S_{U_i^*}(s)} S_{U_i^*}(t_0)^2 \right\} - S(t_0)^2
\end{aligned}
\tag{II·1}
$$

where $G(s) = 1/S_c(s)$. Murray & Tsiatis (1996) showed that

$n\text{Var}\{\hat{S}_{KM}(t_0) - S(t_0)\} = S(t_0)G(t_0) - \int_0^{t_0} \frac{G'(s)ds}{S(s)} S(t_0)^2 - S(t_0)^2$. Comparing these two variances, we have

$$\Delta\text{Var} = n\text{Var}\{\hat{S}_{KM}(t_0) - S(t_0)\}$$

$$-\text{Var}\{\hat{\mathscr{W}}^{\mathbf{Z}}(t_0)\} = -\int_0^{t_0} \frac{G'(s)ds}{S(s)}S(t_0)^2 + E\left\{\int_0^{t_0} \frac{G'(s)ds}{S_{U_i^*}(s)}S_{U_i^*}(t_0)^2\right\} = \int_0^{t_0}\left\{E\left[\frac{S_{U_i^*}(t_0)^2}{S_{U_i^*}(s)} - \frac{S(t_0)^2}{S(s)}\right]\right\}dG(s)$$

$$= \int_0^{t_0} E\{S_{U_i^*}(s)[S_{U_i^*}(t_0|s) - S(t_0|s)]^2\}dG(s) \geq 0$$

Here and in the sequel, for any random variable $\mathscr{U}$, we use notation $S_{\mathscr{U}}(t \mid s)$ to denote $P(T_{\mathbb{L}} > t \mid T_{\mathbb{L}} > s, \mathscr{U})$ and $S(t \mid s) = P(T_{\mathbb{L}} > t \mid T_{\mathscr{U}} > s)$. Provided that $\text{cov}(U_i^*, T_{\mathbb{L}i}|T_{\mathbb{L}i} \leq t_0) > 0$, $\hat{S}(\hat{\beta}^*, t_0)$ is always more efficient than $\hat{S}_{KM}(t_0)$, asymptotically. Furthermore, ΔVar is maximized when the model used to obtain the score, U_i^*, is the true model. To see this, we note that if $S_{U_i^*}(t)$ is replaced with the true conditional survival $S_{\mathbf{Z}_i}(t)$, the corresponding variance reduction over the KM estimator is

$\Delta\text{Var}(\mathbf{Z}) \equiv \int_0^{t_0} E[I(T_{\mathbb{L}i} > s)\{S_{\mathbf{Z}_i}(t_0|s) - S(t_0|s)\}^2]dG(s)$. Comparing the variance reduction using $S_{\mathbf{Z}_i}(t)$ versus using an arbitrary function of $\mathbf{Z}_i$, denoted by $\mathscr{U}_i$, we have

$$\Delta\text{Var}(\mathbf{Z}) - \Delta\text{Var}(\mathscr{U})$$

$$= \int_0^{t_0} E\{I(T_{\mathbb{L}i} > s)[S_{\mathbf{Z}_i}(t_0|s) - S_{\mathscr{U}_i}(t_0|s)]^2\}dG(s) + \int_0^{t_0} 2E\{[S_{\mathscr{U}_i}(t_0|s) - S(t|s)][S_{\mathbf{Z}_i}(t_0|s) - S_{\mathscr{U}_i}(t_0|s)]|T_{\mathbb{L}i} > s\}P(T_{\mathbb{L}i} > s)dG(s)$$

Since $E\{S_{\mathbf{Z}_i}(t_0|s) \mid T_i > s, \mathscr{U}_i\} = E\{E(I(T_{\mathbb{L}i} > t)|T_{\mathbb{L}i} > s, \mathbf{Z}_i)|T_{\mathbb{L}i} > s, \mathscr{U}_i\} = E\{T_{\mathbb{L}i} > t|T_{\mathbb{L}i} > s, \mathscr{U}_i\} = S_{\mathscr{U}_i}(t_0|s)$, and hence $E\{[S_{\mathscr{U}_i}(t_0|s) - S(t|s)][S_{\mathbf{Z}_i}(t_0|s) - S_{\mathscr{U}_i}(t_0|s)]|T_{\mathbb{L}i} > s\} = 0$. Therefore, $\Delta\text{Var}(\mathbf{Z}) - \Delta\text{Var}(\mathscr{U}) \geq 0$. Since $P(T_{\mathbb{L}i} > s|U_i^*) = S_{\mathbf{Z}_i}(s)$ under the correct specification of the working model, $\hat{\mathscr{S}}^{\mathbf{Z}}(t)$ achieves maximal efficiency when the fitted Cox model holds.

We next derive the variance form of our complete estimate, $\tilde{\mathscr{W}}_{LM}(t, t_0)$. From (I·8), $\text{Var}\{\tilde{\mathscr{W}}_{LM}(t, t_0)\} \approx E[\{\varphi_{1i}(t, t_0) + \varphi_{2i}(t, t_0)\}^2] = E\{\varphi_{1i}(t, t_0)^2\} + E\{\varphi_{2i}(t, t_0)^2\}$ since it can be shown that $E\{\varphi_{1i}(t, t_0)\varphi_{2i}(t, t_0)\} = 0$. It follows from (I·5) and (II·1), that

$$E\{\phi_{2i}(t, t_0)^2\} \approx S(t|t_0)^2\left[S(t)G(t) - E\left\{\int_0^{t_0} \frac{G'(s)ds}{S_{U_i^*}(s)}S_{U_i^*}(t_0)^2\right\} - S(t_0)^2\right]$$

Using similar arguments as above and the expression in (I·6), we can show that

$$\text{Var}\{\tilde{\mathscr{W}}(t|t_0)\} = S(t|t_0)G(t|t_0) - E\left\{\int_{t_0}^t \frac{G'(s|t_0)ds}{S_{U_i}(s|t_0)}S_{U_i}(t|t_0)^2\right\} - S(t|t_0)^2$$

where $G(s|t_0) = 1/P(C_i > s|t_0)$. It follows along with (I·7) that

$$E[\phi_{1i}(t,t_0)^2] = \frac{S(t_0)}{S_C(t_0)} \left[S(t|t_0)G(t|t_0) - E\left\{ \int_{t_0}^t \frac{G'(s|t_0)ds}{S_{U_i}(s|t_0)} S_{U_i}(t|t_0)^2 \right\} - S(t|t_0)^2 \right]$$

Therefore, $\mathrm{Var}\{\widehat{\mathscr{H}}_{\mathrm{LM}}(t,t_0)\} \approx$

$$\frac{S(t_0)}{S_C(t_0)} \left[S(t|t_0)G(t|t_0) - E\left\{ \int_{t_0}^t \frac{G'(s|t_0)ds}{S_{U_i}(s|t_0)} S_{U_i}(t|t_0)^2 \right\} - S(t|t_0)^2 \right] + S(t|t_0)^2 \left[S(t)G(t) - E\left\{ \int_0^{t_0} \frac{G'(s)ds}{S_{U_i^*}(s)} S_{U_i^*}(t_0)^2 \right\} - S(t)^2 \right]$$

Note that $\hat{S}_{\mathrm{KM}}(t) = \hat{S}_{\mathrm{KM}}(t|t_0)\hat{S}_{\mathrm{KM}}(t_0)$ where $\hat{S}_{\mathrm{KM}}(t|t_0)$ is the KM estimate of survival at t using only Ω_{t_0}. Similarly to (A.2) in Supplementary Appendix A, using a Taylor series expansion, one can show that $\hat{S}_{\mathrm{KM}}(t) - S(t) \approx S(t_0)[\hat{S}_{\mathrm{KM}}(t|t_0) - S(t|t_0)] + S(t|t_0)[\hat{S}_{\mathrm{KM}}(t_0) - S(t_0)]$ and thus,

$$\begin{aligned} n\mathrm{Var}[\hat{S}_{\mathrm{KM}}(t) & - S(t)] \approx \frac{S(t_0)}{S_C(t_0)} \left[G(t|t_0)S(t|t_0) - \int_{t_0}^t \frac{G'(s|t_0)ds}{S(s|t_0)} S(t|t_0)^2 - S(t|t_0)^2 \right] \\ & + S(t|t_0)^2 \left[G(t_0)S(t_0) \right. \\ & - \int_0^{t_0} \frac{G'(s)ds}{S(s)} S(t_0)^2 \\ & \left. - S(t_0)^2 \right] \end{aligned}$$

where we have used the independence of $\hat{S}_{\mathrm{KM}}(t|t_0)$ and $\hat{S}_{\mathrm{KM}}(t_0)$, and KM variance estimates from Murray & Tsiatis (1996). Examining the difference which we denote by $\Delta\mathrm{Var}_{\mathrm{LM}}$,

$$\begin{aligned} \Delta\mathrm{Var}_{\mathrm{LM}} &= n\mathrm{Var}\{\hat{S}_{\mathrm{KM}}(t) \\ & - S(t)\} \\ & - n\mathrm{Var}\{\hat{S}_{\mathrm{LM}}(\hat{\beta}^*, \hat{\beta}, t) \\ & - S(t)\} \approx \frac{S(t_0)}{S_C(t_0)} \int_{t_0}^t \left[E\left\{ S_{U_i}(s|t_0)[S_{U_i}(t|s) - S(t|s)]^2 \right\} \right] dG(s|t_0) \\ & + S(t|t_0)^2 \int_0^{t_0} E\left\{ S_{U_i^*}(s)[S_{U_i^*}(t_0|s) - S(t_0|s)]^2 \right\} dG(s) \geq 0. \end{aligned}$$

Thus, the proposed estimator, $\hat{S}_{\mathrm{LM}}(\hat{\beta}^*, \hat{\beta}, t)$, is at least as efficient as the Kaplan Meier estimate. In addition, the reduction is maximized when the assumed working models in (2·3) and (2·4) are true. Note that the gain in efficiency depends on t_0 and the true values of β and β^*. The maximum reduction is within the class of estimators constructed using $\mathbf{W}$ and t_0. The amount of reduction would also depend on the magnitude of β and β^*. In particular, if β and β^* are equal to zero, there would be no efficiency gain, but the larger β and β^* are, the

more variation we would expect in $S_{U_i}(t|s)$ and $S_{U_i^*}(t_0|s)$ which would lead to a larger expected reduction in variance.

References

Bang H, Tsiatis A. Estimating medical costs with censored data. Biometrika. 2000; 87:329–343.

Parast et al.
Page 18

Beran, R. Technical report. University of California Berkeley; 1981. Nonparametric regression with randomly censored survival data.

Bilias Y, Gu M, Ying Z. Towards a general asymptotic theory for Cox model with staggered entry. The Annals of Statistics. 1997; 25:662–682.

Breslow N. Contribution to the discussion of the paper by dr cox. Journal of the Royal Statistical Society, Series B. 1972; 34:216–217.

Cai T, Tian L, Uno H, Solomon S, Wei L. Calibrating parametric subject-specific risk estimation. Biometrika. 2010; 97:389–404. [PubMed: 23049123]

Cook R, Lawless J. Some comments on efficiency gains from auxiliary information for right-censored data. Journal of Statistical Planning and Inference. 2001; 96:191–202.

Cortese G, Andersen P. Competing risks and time-dependent covariates. Biometrical Journal. 2010; 52:138–158. [PubMed: 20029852]

Du Y, Akritas M. Uniform strong representation of the conditional kaplan-meier process. Mathematical Methods of Statistics. 2002; 11:152–182.

Fine J, Jiang H, Chappell R. On semi-competing risks data. Biometrika. 2001; 88:907–919.

Friedman, J.; Hastie, T.; Tibshirani, R. The elements of statistical learning. Vol. volume 1. Springer Series in Statistics; 2001.

Gray R. A kernel method for incorporating information on disease progression in the analysis of survival. Biometrika. 1994; 81:527–539.

Hammer S, Katzenstein D, Hughes M, Gundacker H, Schooley R, Haubrich R, Henry W, Lederman M, Phair J, Niu M, et al. A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. New England Journal of Medicine. 1996; 335:1081–1090. [PubMed: 8813038]

Hirschtick R, Glassroth J, Jordan M, Wilcosky T, Wallace J, Kvale P, Markowitz N, Rosen M, Mangura B, Hopewell P, et al. Bacterial pneumonia in persons infected with the human immunodeficiency virus. New England Journal of Medicine. 1995; 333:845. [PubMed: 7651475]

Hjort NL. On inference in parametric survival data models. International Statistical Review/Revue Internationale de Statistique. 1992; 60:355–387.

Hu Z, Follmann D, Qin J. Semiparametric dimension reduction estimation for mean response with missing data. Biometrika. 2010; 97:305–319. [PubMed: 23049121]

Hu Z, Follmann D, Qin J. Dimension reduced kernel estimation for distribution function with incomplete data. Journal of statistical planning and inference. 2011; 141:3084–3093. [PubMed: 21731174]

Jiang H, Fine J, Chappell R. Semiparametric analysis of survival data with left truncation and dependent right censoring. Biometrics. 2005; 61:567–575. [PubMed: 16011706]

Kaplan E, Meier P. Nonparametric estimation from incomplete observations. Journal of the American statistical association. 1958:457–481.

Khan S, Tamer E. Inference on endogenously censored regression models using conditional moment inequalities. Journal of Econometrics. 2009; 152:104–119.

Kosorok, MR. Introduction to empirical processes and semiparametric inference. Springer: 2008.

Lagakos S. The loss in efficiency from misspecifying covariates in proportional hazards regression models. Biometrika. 1988; 75:156–160.

Lee S, Klein J, Barrett A, Ringden O, Antin J, Cahn J, Carabasi M, Gale R, Giralt S, Hale G, et al. Severity of chronic graft-versus-host disease: association with treatment-related mortality and relapse. Blood. 2002; 100:406–414. [PubMed: 12091329]

Li K-C, Duan N. Regression analysis under link violation. The Annals of Statistics. 1989; 17:1009–1052.

Lin D, Wei L. The robust inference for the cox proportional hazards model. Journal of the American Statistical Association. 1989:1074–1078.

Lu X, Tsiatis A. Improving the efficiency of the log-rank test using auxiliary covariates. Biometrika. 2008; 95:679–694.

Murray S, Tsiatis A. Nonparametric survival estimation using prognostic longitudinal covariates. Biometrics. 1996:137–151. [PubMed: 8934589]

O'neill T. Inconsistency of the misspecified proportional hazards model. Statistics & probability letters. 1986; 4:219–222.

Parast L, Cheng SC, Cai T. Incorporating short-term outcome information to predict long-term survival with discrete markers. Biometrical Journal. 2011; 53:294–307. [PubMed: 21337601]

Parast L, Cheng S-C, Cai T. Landmark prediction of long term survival incorporating short term event time information. Journal of the American Statistical Association. 2012; 107:1492–1501. [PubMed: 23293405]

Robins J, Ritov Y. Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. Statistics in Medicine. 1997; 16:285–319. [PubMed: 9004398]

Rotnitzky A, Robins J. Inverse probability weighted estimation in survival analysis. Encyclopedia of Biostatistics. 2005:2619–2625.

Scott, D. Multivariate Density Estimation. Vol. 1. New York: Wiley; 1992. Multivariate density estimation. 1992

Siannis F, Farewell V, Head J. A multi-state model for joint modeling of terminal and non-terminal events with application to whitehall ii. Statistics in medicine. 2007; 26:426–442. [PubMed: 16220522]

Tian L, Cai T, Goetghebeur E, Wei L. Model evaluation based on the sampling distribution of estimated absolute prediction error. Biometrika. 2007; 94:297–311.

Tian L, Cai T, Zhao L, Wei L. On the covariate-adjusted estimation for an overall treatment difference with data from a randomized comparative clinical trial. Biostatistics. 2012

Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored regression models. Journal of the American Statistical Association. 2007; 102:527–537.

Van Houwelingen H. Dynamic prediction by landmarking in event history analysis. Scandinavian Journal of Statistics. 2007; 34:70–85.

Van Houwelingen H, Putter H. Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. Lifetime Data Analysis. 2008; 14:447–463. [PubMed: 18836831]

Van Houwelingen, J.; Putter, H. Dynamic Prediction in Clinical Survival Analysis. CRC Press; 2012.

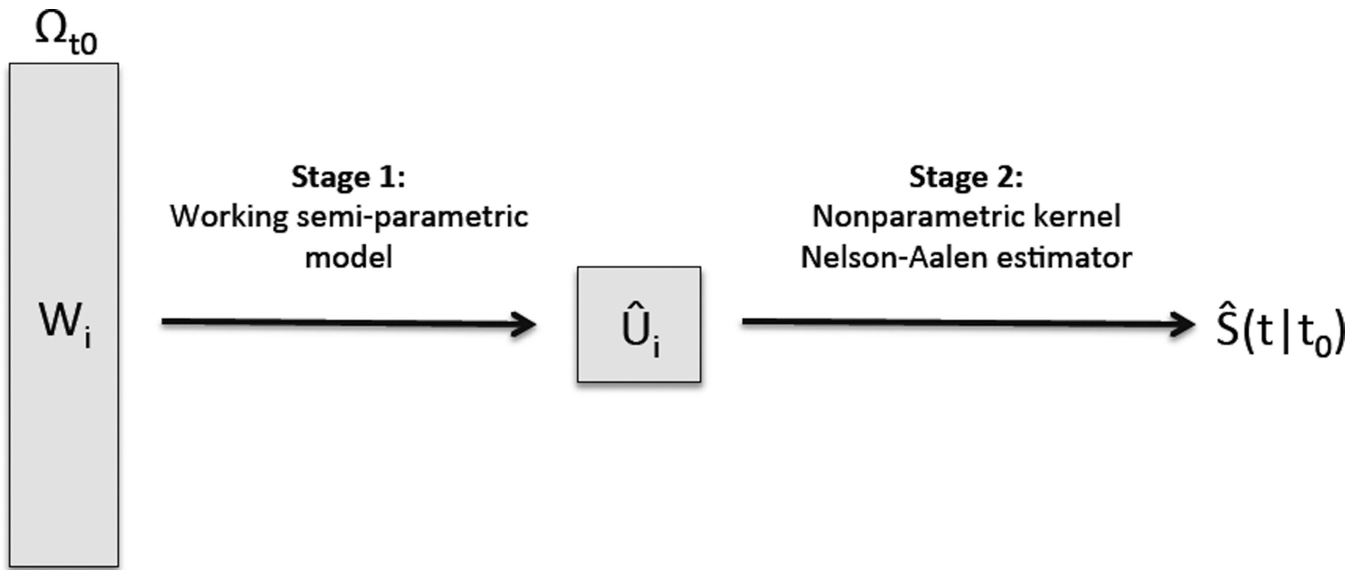**Figure 1.**
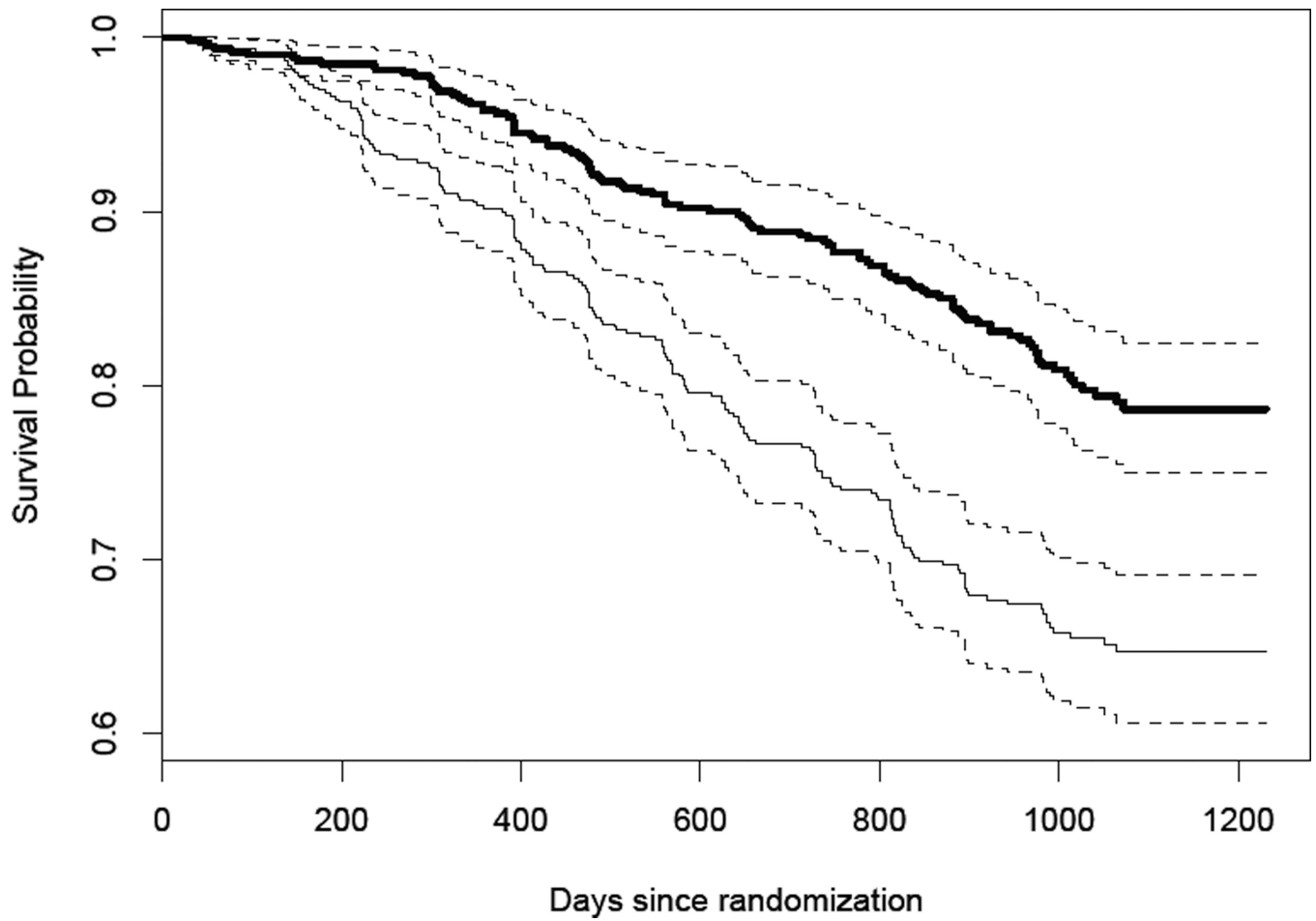Illustration of proposed two-stage procedure

Figure 2.
Kaplan Meier estimate of survival for the mono group (thin black line) and combo group
(thick black line) with corresponding 95% confidence intervals (dashed lines).

Table 1

Estimates of $S(t)$ for $t = 2$ years in treatment group B in settings (i),(ii) and (iii) using a Kaplan Meier estimator, $\hat{S}_{KM}(t)$, the proposed estimator using $\mathbf{Z}$ information only, $\tilde{S}^{\mathbf{Z}}(t)$, $T_{\mathbb{S}}$ information only, $\tilde{S}^{\mathbf{Z}}(t)$, and $\mathbf{Z}$ and $T_{\mathbb{S}}$ information, $\hat{S}_{LM}^{\mathbf{W}}(t)$, with corresponding empirical standard error (ESE), average of the standard error estimates from the perturbation-resampling method (ASE), efficiency relative to $\hat{S}_{KM}(t)$ with respect to mean squared error (MSE) (RE_{MSE}), and empirical coverage (Cov) of the 95% CIs.

	$\hat{S}_{KM}(t)$	$\tilde{S}^{\mathbf{Z}}(t)$	$\hat{S}_{LM}^{T_{\mathbb{S}}}(t)$	$\hat{S}_{LM}^{\mathbf{W}}(t)$
		Setting (i)		
Estimate	0.4205	0.4267	0.4238	0.4246
Bias	−0.0004	0.0060	0.0031	0.0039
ESE	0.0225	0.0220	0.0220	0.0207
ASE	0.0231	0.0218	0.0218	0.0206
RE_{MSE}	-	1.07	1.04	1.20
Cov	0.95	0.95	0.94	0.95
		Setting (ii)		
Estimate	0.3676	0.3745	0.3716	0.3720
Bias	−0.0008	0.006	0.0032	0.0036
ESE	0.0224	0.0222	0.0219	0.0210
ASE	0.0227	0.0213	0.0215	0.0201
RE_{MSE}	-	1.07	1.03	1.19
Cov	0.96	0.94	0.94	0.93
		Setting (iii)		
Estimate	0.3248	0.3322	0.3289	0.3292
Bias	0.0001	0.0074	0.0040	0.0043
ESE	0.0224	0.0222	0.0218	0.0206
ASE	0.0222	0.0208	0.0210	0.0196
RE_{MSE}	-	1.06	1.01	1.20
Cov	0.94	0.94	0.91	0.93

Table 2

Results from testing for a difference in survival at time $t = 2$ years between the two treatment groups using a Kaplan Meier estimator, $\hat{\Delta}_{KM}(t)$, the proposed estimator using $\mathbf{Z}$ information only, $\hat{\Delta}^{\mathbf{Z}}(t)$, $T_{\mathbb{S}}$ information only, $\hat{\Delta}^{T_{\mathbb{S}}}_{LM}(t)$, $\mathbf{Z}$ and $T_{\mathbb{S}}$ information, $\hat{\Delta}^{W}_{LM}(t)$, and the augmented estimator, $\hat{\Delta}^{W}_{AUG}(t)$ with corresponding empirical standard error (ESE), average of the standard error estimates from the perturbation-resampling method (ASE), efficiency relative to $\hat{S}_{KM}(t)$ with respect to MSE, and power.

	$\hat{\Delta}_{KM}(t)$	$\hat{\Delta}^{\mathbf{Z}}(t)$	$\hat{\Delta}^{T_{\mathbb{S}}}_{LM}(t)$	$\hat{\Delta}^{W}_{LM}(t)$	$\hat{\Delta}^{W}_{AUG}(t)$
	Setting (i), no treatment difference				
Estimate	−0.0002	−0.0002	−0.0004	−0.0001	0.0003
ESE	0.0323	0.0318	0.0316	0.0300	0.0282
ASE	0.0326	0.0308	0.0309	0.0291	0.0267
RE$_{MSE}$	-	1.12	1.12	1.26	1.49
α	0.0415	0.0462	0.0544	0.0467	0.0538
	Setting (ii), small treatment difference				
Estimate	0.0525	0.0519	0.0520	0.0523	0.0525
ESE	0.0318	0.0312	0.0312	0.0292	0.0276
ASE	0.0323	0.0305	0.0306	0.0287	0.0263
RE$_{MSE}$	-	1.12	1.11	1.27	1.51
Power	0.3740	0.4110	0.4060	0.4570	0.5200
	Setting (iii), large treatment difference				
Estimate	0.0954	0.0942	0.0947	0.0952	0.0952
ESE	0.0317	0.0312	0.0310	0.0288	0.0270
ASE	0.0320	0.0301	0.0303	0.0283	0.0259
RE$_{MSE}$	-	1.13	1.11	1.27	1.53
Power	0.8460	0.8690	0.8650	0.9090	0.9410

Table 3

Estimates of $S(t)$ for $t = 2.5$ years in two treatment groups from ACTG Protocol 175 using a Kaplan Meier estimator, $\hat{S}_{KM}(t)$, the proposed estimator using $\mathbf{Z}$ information only, $\hat{S}^{\mathbf{Z}}(t)$, $T_{\mathbb{S}}$ information only, $\hat{S}^{\mathbf{Z}}(t)$, and $\mathbf{Z}$ and $T_{\mathbb{S}}$ information, $\hat{S}^{\mathbf{W}}_{LM}(t)$, with corresponding standard error from the perturbation-resampling method (SE).

	$\hat{S}_{KM}(t)$	$\hat{\mathcal{S}}(t)$	$\hat{S}^{T_{\mathbb{S}}}_{LM}(t)$	$\hat{S}^{\mathrm{W}}_{LM}(t)$
Treatment: Zidovudine				
Estimate	0.92	0.93	0.93	0.93
SE	0.0114	0.0108	0.0107	0.0105
Treatment: Zidovudine and Zalcitabine				
Estimate	0.95	0.96	0.95	0.95
SE	0.0101	0.0091	0.0094	0.0091

Table 4

Results from testing for a difference in survival at time $t = 2.5$ years between the two treatment groups from ACTG Protocol 175 using a Kaplan Meier estimator, $\hat{\Delta}_{KM}(t)$, the proposed estimator using $\mathbf{Z}$ information only, $\Delta^{\mathbf{Z}}(t)$, T_S information only, $\hat{\Delta}_{LM}^{T_S}(t)$, $\mathbf{Z}$ and T_S information, $\hat{\Delta}_{LM}^{\mathbf{W}}(t)$, and the augmented estimator, $\hat{\Delta}_{AUG}^{\mathbf{W}}(t)$ with corresponding standard error from the perturbation-resampling method (SE) and efficiency relative to $\hat{S}_{KM}(t)$ with respect to variance (RE) and p-value

	$\hat{\Delta}_{KM}(t)$	$\Delta^{\mathbf{Z}}(t)$	$\hat{\Delta}_{LM}^{T_S}(t)$	$\hat{\Delta}_{LM}^{\mathbf{W}}(t)$	$\hat{\Delta}_{AUG}^{\mathbf{W}}(t)$
Estimate	0.03	0.03	0.02	0.03	0.03
SE	0.0153	0.0143	0.0143	0.0140	0.0125
RE	1.00	1.13	1.14	1.19	1.49
p-value	0.09	0.04	0.10	0.05	0.03