Landmark Estimation of Survival Incorporating Intermediate Event Information in a Randomized Clinical Trial

Layla Parast and Tianxi Cai

Harvard University Department of Biostatistics, Boston, MA 02115

Summary

In many studies with a survival outcome, it is often too expensive, time consuming, impractical or unethical to fully observe the primary event of interest in all subjects being studied. This often leads to heavily censored data and difficulty in efficiently estimating survival. In certain diseases, baseline covariates and the event time of a non-fatal intermediate event may be associated with overall survival. In these settings, incorporating intermediate event information may lead to gains in efficiency in terms of 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 event and covariates to predict survival focus on estimation of relative risk parameters and/or the joint distribution of events under semi-parametric 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 survival by incorporating intermediate event information observed before some landmark time, which serves as a useful approach to overcome semi-competing risks issues. In addition, we present a testing procedure using these resulting estimates to test for a difference in survival between two treatment groups in a randomized clinical trial setting. 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.

Keywords: *Efficiency Augmentation, Landmark Prediction, Semi-competing Risks, Survival Analysis.*

1 Introduction

In many research settings involving survival outcomes, it is often too expensive, time consuming, impractical or unethical to fully observe the primary event of interest in all subjects being studied. This often leads to heavily censored data and difficulty in efficiently estimating survival especially when the event is not common. In certain diseases, the event time of a non-fatal intermediate event may be associated with overall survival. For example, Cortese & Andersen (2010) and Lee et al. (2002) examined how the development of acute graft-versus-host disease affects survival of acute leukemia patients following bone marrow or stem cell transplantation. Additionally, Hirschtick et al. (1995) demonstrated an association between occurrence of bacterial pneumonia and death among HIV-positive patients. Among end-stage renal disease patients, the relationship between an early hospitalization and risk of death has been examined by Collins et al. (2009) and USRDS (2010). In these settings, incorporating intermediate event information may lead to gains in efficiency for the estimation of survival and detecting a difference in survival between two treatment groups. If gains in power can be achieved, it may then be possible to decrease the sample size of patients needed for a study to achieve a particular power level or decrease the duration of the study.

In the aforementioned examples, the primary event of interest is time to a terminal event e.g. 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. Most existing methods for analyzing semi-competing risk data focus on estimation of relative risk parameters and/or the joint distribution of events under semi-parametric models (Fine et al., 2001; Siannis et al., 2007; Jiang et al., 2005). The joint distribution estimators can potentially be used to make predictions about the long term outcome given information on the short

term outcome. Methods developed specifically to incorporate short term information along with other baseline predictors for the prediction of long term survival has largely focused on semi-parametric models in a multi-state framework (Putter et al., 2006; Kay, 1986; Klein et al., 1994; Datta et al., 2000; Hatteville et al., 2002; Cortese & Andersen, 2010). For example, Klein et al. (1994) use a standard Cox regression analysis to incorporate information on whether or not the intermediate event has occurred by a certain time. Hatteville et al. (2002) extend this method to include non-proportional hazards and propose estimation. Recently, Van Houwelingen & Putter (2008) proposed a landmark prediction procedure to incorporate the status of the short term event by a landmark point.

These methods provide useful tools to leverage information about the short term outcome to improve prediction. However, these simple semi-parametric models may not capture the relationship among the two events and the baseline covariates in practice. As an example, the intermediate event may be highly associated with unknown subtypes of disease and the effect of baseline covariates on the long term outcome may vary with the disease subtype. Commonly used proportional hazards models may be inadequate for data from such a mixture of populations. Conclusions regarding the marginal survival rate of the terminal event might be invalid when inference is derived under mis-specified models. To overcome such limitations, we propose a two-stage procedure by (i) first using a semi-parametric approach to incorporating baseline covariates and intermediate event information observed before some landmark time; and (ii) then estimating the marginal survival non-parametrically 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. Section 2 describes this estimation procedure in terms of a two-stage procedure in a one-sample setting. Section 3 presents a testing procedure using these resulting estimates in a randomized clinical trial setting. We also present an augmentation

procedure similar to Lu & Tsiatis (2008), which takes advantage of treatment randomization to further improve efficiency. Simulation studies 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 dataset examining the difference in survival between two treatment groups zidovudine and zidovudine plus zalcitabine.

2 Estimation

Let $T_{\mathbb{L}i}$ denote the event time of the primary event of interest for the ith subject. As there may be information available on multiple intermediate events, let $\mathbf{T}_{\mathbb{S}\mathbf{i}}$ denote the vector of intermediate events times for the ith subject. Let C_i denote the censoring time for the ith subject. Due to censoring we observe $X_{\mathbb{L}i} = \min(T_{\mathbb{L}i}, C_i), \mathbf{X}_{\mathbb{S}\mathbf{i}} = \min(\mathbf{T}_{\mathbb{S}\mathbf{i}}, C_i)$ and $\delta_{\mathbb{L}i} = I(T_{\mathbb{L}i} \leq C_i), \delta_{\mathbb{S}\mathbf{i}} = I(\mathbf{T}_{\mathbb{S}\mathbf{i}} \leq C_i)$. We consider the semi-competing risk setting under which $\mathbf{T}_{\mathbb{S}\mathbf{i}}$ may be censored by $T_{\mathbb{L}i}$ if $T_{\mathbb{L}i}$ occurs before $\mathbf{T}_{\mathbb{S}\mathbf{i}}$, while $T_{\mathbb{L}i}$ is a terminal event subject only to administrative censoring. Namely, $T_{\mathbb{L}i}$ can only be censored by the non-informative censoring time C_i, not by $\mathbf{T}_{\mathbb{S}\mathbf{i}}$. Let $\mathbf{Z}_i$ denote the vector of baseline covariates. We assume that C_i is independent of $(T_{\mathbb{L}}, \mathbf{T}_{\mathbb{S}}, \mathbf{Z})$. In addition, let t_0 denote some landmark time. We are interested in estimating $S(t) = P(T_{\mathbb{L}i} > t)$ where $t > t_0$ using intermediate event and covariate information collected up to t_0.

2·1 *Two-stage estimation procedure*

Observe that 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

$$S(t \mid t_0) = P(T_{\mathbb{L}i} > t \mid T_{\mathbb{L}i} > t_0) = P(T_{\mathbb{L}i} > t \mid X_{\mathbb{L}i} > t_0) \quad \text{and} \quad S(t_0) = P(T_{\mathbb{L}i} > t_0). \tag{2·1}$$

We first focus on estimation of $S(t \mid t_0)$ using a two-stage procedure. Note that conditional on $X_{\mathbb{L}i} > t_0$, intermediate event information up to t_0 has been observed since $I(\mathbf{X}_{\mathbb{S}\mathbf{i}} \leq t_0) = I(T_{\mathbb{S}i} \leq t_0)$ among $\Omega_{t_0} = \{i : X_{\mathbb{L}i} > t_0\}$. Let $\mathbf{W}_i = (\mathbf{Z}, I(\mathbf{X}_{\mathbb{S}\mathbf{i}} \leq t_0), \min(\mathbf{X}_{\mathbb{S}\mathbf{i}}, t_0))^\intercal$ for all subjects in Ω_{t_0}. Note that

$$S(t \mid t_0) = P(T_{\mathbb{L}i} > t \mid X_{\mathbb{L}i} > t_0) = E[P(T_{\mathbb{L}i} > t | X_{\mathbb{L}i} > t_0, \mathbf{W}_i)] \tag{2·2}$$

and let $q_{\mathbf{w}}(t) = P(T_{\mathbb{L}i} > t | T_{\mathbb{L}i} > t_0, \mathbf{W}_i = \mathbf{w})$. When $T_{\mathbb{S}i}$ 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 estimate a consistent estimator of $q_{\mathbf{w}}(t)$. However, when the dimension of $\mathbf{W}$ is large and the number of subjects is small, nonparametric estimation of $q_{\mathbf{w}}(t)$ may not behave well (Robins & Ritov, 1997). Instead we propose to reduce the dimension of $\mathbf{W}$ by approximating $q_{\mathbf{w}}(t)$ with a working semi-parametric model such as the proportional hazards model:

$$q_{\mathbf{w}}(t) = \exp\{-\Lambda_0^{t_0}(t) \exp(\beta^\intercal \mathbf{w})\} \tag{2·3}$$

where $\Lambda_0^{t_0}(\cdot)$ is the unspecified baseline cumulative hazard function for $T_{\mathbb{L}i}$ among Ω_{t_0} and β is an unknown vector of coefficients. This is referred to as our Stage 1. Let $\widehat{\beta}$ be the maximizer of the corresponding log partial likelihood function. When (2·3) is correctly specified, $\widehat{\beta}$ will consistently estimate β and one could estimate $q_{\mathbf{w}}(t)$ as $\widehat{q}_{\mathbf{w}}(t) = \exp\{-\widehat{\Lambda}_0^{t_0}(t) \exp(\widehat{\beta}^\intercal \mathbf{w})\}$ where $\widehat{\Lambda}_0^{t_0}(t)$ can be estimated using the Breslow estimate of baseline hazard. However, if (2·3) is not correctly specified, $\widehat{q}_{\mathbf{w}}(t)$ would not longer be a consistent estimator of $q_{\mathbf{w}}(t)$ and thus (2·2) would no longer hold. On the other hand, the risk score $\widehat{U}_i \equiv \widehat{\beta}^\intercal \mathbf{W}_i$ may still be highly predictive of $T_{\mathbb{L}i}$. This motivates us to overcome the difficulty arising from model mis-specification by using $S(t \mid t_0) = E[P(T_{\mathbb{L}i} > t | X_{\mathbb{L}i} >$

$t_0, U_i)] = E\{s_{U_i}(t)\}$, where $s_u(t) = P(T_{\mathbb{L}i} > t \mid T_{\mathbb{L}i} > t_0, U_i = u)$, $U_i = \beta_0^{\mathrm{T}}\mathbf{W}_i$ and β_0 is the limit of $\widehat{\beta}$, which always exists (Hjort, 1992).

To estimate $s_u(t)$ in Stage 2, we propose to use a nonparametric kernel Nelson-Aalen estimator for the conditional hazard function $\Lambda_u(t) = -\log s_u(t)$ based on subjects in Ω_{t_0}. We consider a local constant estimator by assuming that for u' in a small neighborhood of u, $\Delta\Lambda_{u'}(t) \equiv \Delta\Lambda_u(t)$. Specifically for any given t and u, we obtain $\Delta\widehat{\Lambda}_u(t)$ as the following minimizer:

$$\arg\min_a \sum_{i\in\Omega_{t_0}} K_h(\hat{D}_{ui})\{\Delta N_i(t) - Y_i(t)a\}^2 = \frac{\sum_{i\in\Omega_{t_0}} K_h(\hat{D}_{ui})\Delta N_i(t)}{\sum_{i\in\Omega_{t_0}} K_h(\hat{D}_{ui})Y_i(t)}, \tag{2·4}$$

where $Y_i(t) = I(T_{\mathbb{L}i} \geq t)$, $\Delta N_i(t) = N_i(t) - N_i(t-) = \lim_{x\to 0}\int_{t-x}^{t} dN_i(s)$, $N_i(t) = I(T_{\mathbb{L}i} \leq t)\delta_{\mathbb{L}i}$, $\hat{D}_{ui} = \widehat{U}_i - u$, $K(\cdot)$ is a smooth symmetric density function, $K_h(x) = K(x/h)/h$, and $h = O(n^{-v})$ is a bandwidth with $1/2 > v > 0$ (Wand et al., 1991; Park et al., 1997). Based on (2·4), we obtain a local constant estimator for $\Lambda_u(t)$ as

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

The resulting estimate for $s_u(t)$ is $\widehat{s}_u(t) = \exp\{-\widehat{\Lambda}_u(t)$. It can be shown using similar arguments as in Cai et al. (2010) that $\widehat{s}_u(t)$ is consistent for $s_u(t)$ under mild regularity conditions. Finally, we propose to estimate $S(t \mid t_0)$ as

$$\widehat{S}^{\mathbf{w}}(t \mid t_0) = \frac{1}{n_{t0}} \sum_{i\in\Omega_{t_0}} \widehat{s}_{\widehat{U}_i}(t) = \frac{1}{n_{t0}} \sum_{i\in\Omega_{t_0}} \exp\{-\widehat{\Lambda}_{\widehat{U}_i}(t)\}$$

where n_{t0} is the number of subjects in Ω_{t_0} (see Figure 2). The uniform consistency of $\widehat{s}_u(t)$ along with

the consistency of $\widehat{\beta}$ for β_0 ensures the consistency of $\widehat{S}^{\mathbf{W}}(t \mid t_0)$ for $S(t \mid t_0)$.

Now that we have obtained an estimate of $S(t|t_0)$ in (2·1), an estimate for $S(t)$ follows similarly from this same two-stage procedure replacing $\mathbf{W}_i = \mathbf{Z}_i$ throughout and using all patients. Let $\widehat{S}^{\mathbf{Z}}(t_0)$ denote this resulting estimate. An estimate for our primary quantity of interest $S(t)$ incorporating intermediate event and covariate information follows as $\widehat{S}^{\mathbf{W}}_{\text{LM}}(t) \equiv \widehat{S}^{\mathbf{W}}(t \mid t_0)\widehat{S}^{\mathbf{Z}}(t_0)$ where LM stands for landmark.

To compare the performance of our proposed estimation procedure, let $\widehat{S}_{\text{KM}}(t)$ denote the Kaplan-Meier estimate of $S(t)$. If a gain in efficiency is observed over the Kaplan-Meier estimate, one may consider whether this advantage is due to incorporating covariate information or incorporating $\mathbf{T}_{\mathbb{S}}$ information. To address this question, let $\widehat{S}^{T_{\mathbb{S}}}_{\text{LM}}(t) = \widehat{S}^{T_{\mathbb{S}}}(t \mid t_0)\widehat{S}_{\text{KM}}(t_0)$ denote an estimate of $S(t)$ incorporating only $\mathbf{T}_{\mathbb{S}}$ information where $\widehat{S}^{T_{\mathbb{S}}}(t \mid t_0)$ is obtained using the two-stage procedure with $\mathbf{W}_i = \mathbf{T}_{\mathbb{S}}$. In addition, let $\widehat{S}^{\mathbf{Z}}(t)$ denote an estimate of $S(t)$ incorporating only $\mathbf{Z}$ information.

3 Comparing Survival Between Two Treatment Groups

In a randomized clinical trial setting with treatments A and B, there is often interest in testing whether there is a difference in survival between the two groups. Methods to incorporate intermediate event or auxiliary information when testing for a treatment effect have been previously proposed in the literature. Cook & Lawless (2001) discusses a variety of statistical methods that have been proposed including parametric and semi-parametric models. Fully parametric models (Lagakos, 1977) allow the greatest potential gains in efficiency but may not be robust to departures from the assumed model. Fleming et al. (1994) propose using augmented score and likelihood methods for incorporating data on auxiliary endpoints. Gray (1994) adopted kernel estimation methods which incorporate intermediate event information only and proposed a test statistic for equality of survival between treatment groups.

Through extensive simulation studies, they showed that efficiency gains vary depending on the amount of censoring and the correlation between intermediate and long term events. Recently, 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 is mis-specified, β in this setting may be difficult to interpret. Here, we are interested in making more efficient inference about the contrast between two survival rates by incorporating the baseline and intermediate event information.

Let $S_A(t)$ and $S_B(t)$ denote the survival rate of $T_{\mathbb{L}}$ at time t for treatment groups A and B respectively. Using the proposed estimation procedure, one may obtain $\widehat{S}^{\mathbf{W}}_{\text{LM}}(t)$ for treatment groups A and B denoted by $\widehat{S}^{\mathbf{W}}_{\text{LM},A}(t)$ and $\widehat{S}^{\mathbf{W}}_{\text{LM},B}(t)$, respectively. An estimator for the risk difference, $\Delta(t) = S_A(t) - S_B(t)$, may be obtained as $\widehat{\Delta}^{\mathbf{W}}_{\text{LM}}(t) = \widehat{S}^{\mathbf{W}}_{\text{LM},A}(t) - \widehat{S}^{\mathbf{W}}_{\text{LM},B}(t)$. A confidence interval for $\Delta(t)$ may be constructed using a normal approximation by centering at $\widehat{\Delta}^{\mathbf{W}}_{\text{LM}}(t)$ with standard error estimated using the bootstrap. To test the null hypothesis of $H_0 : S_A(t) = S_B(t)$, we propose a Wald-type test statistic :

$$Z^{\mathbf{W}}_{\text{LM}}(t) = \frac{\widehat{\Delta}^{\mathbf{W}}_{\text{LM}}(t)}{\widehat{\sigma}(\widehat{\Delta}^{\mathbf{W}}_{\text{LM}}(t))}$$

where $\widehat{\Delta}^{\mathbf{W}}_{\text{LM}}(t) = \widehat{S}^{\mathbf{W}}_{\text{LM},A}(t) - \widehat{S}^{\mathbf{W}}_{\text{LM},B}(t)$ and $\widehat{\sigma}(\widehat{\Delta}^{\mathbf{W}}_{\text{LM}}(t))$ is the estimated standard error of $\widehat{\Delta}^{\mathbf{W}}_{\text{LM}}(t)$. Thus we would reject the null hypothesis at α level when $|Z^{\mathbf{W}}_{\text{LM}}(t)| > \mathcal{Z}_{1-\alpha/2}$, where $\mathcal{Z}_p$ is the pth percentile of the standard normal. To compare this testing procedure to a test based on $\widehat{S}_{\text{KM}}(t)$ let $\widehat{\Delta}_{\text{KM}}(t)$ denote the corresponding estimate for $\Delta(t)$. Similarly, let $\widehat{\Delta}^{\mathbf{Z}}(t)$ denote the estimate of $\Delta(t)$ using $\mathbf{Z}$ information only and $\widehat{\Delta}^{T_{\mathbb{S}}}_{\text{LM}}(t)$ denote the estimate using $\mathbf{T}_{\mathbb{S}}$ information only. Corresponding test statistics can then be constructed similar to $Z^{\mathbf{W}}_{\text{LM}}(t)$.

As mentioned previously, Lu & Tsiatis (2008) observe substantial gains in efficiency through an augmentation procedure. This procedure makes two assumptions: 1) conditional on the treatment assignment, the censoring time is independent of the failure time and of baseline covariates and 2) treatment assignment is independent of the baseline covariates. Through landmarking, we have taken advantage of the additional information obtained when the censoring time is independent of the failure time and of baseline covariates. To determine whether our estimation procedure may gain additional efficiency by taking advantage of independent treatment assignment in a randomized clinical trial setting, we examine an additional augmentation step. Let

$$\widehat{\Delta}^{\mathbf{w}}_{\text{AUG}}(t) = \widehat{\Delta}^{\mathbf{w}}_{\text{LM}}(t) - a^{\intercal}\epsilon \quad \text{where} \quad \epsilon = \sum_{i=1}^{n}(R_i - \pi)\mathbf{B}(\mathbf{Z}_i)$$

where R_i = treatment assignment, $\pi = P(R_i = 1)$, and $\mathbf{B}(\cdot)$ is a vector of basis functions for $\mathbf{Z}$. As a result of randomization we can assume π is known, though in numerical studies, it can be estimated as $\widehat{\pi}$ = the observed proportion of treatment assignment. The minimizer of $\text{var}[\Delta_{\text{AUG}}(t)]$, $a^* = \text{var}(\epsilon)^{-1}\text{cov}\{\widehat{\Delta}^{\mathbf{w}}_{\text{LM}}(t), \epsilon\}$, and $\widehat{\Delta}^{\mathbf{w}}_{\text{AUG}}(t) = \widehat{\Delta}^{\mathbf{w}}_{\text{LM}}(t) - a^{*\intercal}\widehat{\epsilon}$ where $\widehat{\epsilon} = \sum_{i=1}^{n}(R_i - \hat{\pi})\mathbf{B}(\mathbf{Z}_i)$. When the influence function of $\widehat{\Delta}^{\mathbf{w}}_{\text{LM}}(t)$ is known, an explicit form of a^* may be obtained. Due to the complexity of our proposed estimation procedures and potential model mis-specification in stage I, the influence function would be unknown and even difficult to estimate explicitly. To overcome this difficulty, one may use resampling methods such as the bootstrap to obtain perturbed realizations of $\{\widehat{\Delta}^{\mathbf{w}}_{\text{LM}}(t), \epsilon\}$ and then estimate the variance and covariance of $\widehat{\Delta}^{\mathbf{w}}_{\text{LM}}(t)$ and ϵ. When the covariates are predictive of survival, such augmentation is likely to further improve the estimation precision of $\Delta(t)$. To capture the potential non-linear effects of the covariates, one may choose basis functions that allow for quadratic and interactive effects. Throughout our numerical studies, we chose $\mathbf{B}(\mathbf{Z}) = (1, \mathbf{Z}, \mathbf{Z}^2)^{\intercal}$ as used in Lu

& Tsiatis (2008).

4 Simulations

We conducted simulation studies to examine the finite sample properties of the proposed estimation procedures. 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 (i) 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 1000 replications under these settings. Variance estimates are obtained using the bootstrap. Under the null setting (i), the event times for each treatment group (Treatment A and Treatment B) were generated as $T_{\mathbb{S}i} = [-b_1^{-1}\log(1-U_1)]^{1/a_1}$ where $a_1 = 1.5, b_1 = 1$, and $U_1 \sim$ Uniform(0,1). The conditional distribution $T_{\mathbb{L}} - T_{\mathbb{S}} \mid T_{\mathbb{S}}$ was generated 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 $U_2 \sim$ Uniform(0,1). We take the covariate Z to be $0.75U_2 + .25U_3$ where $U_3 \sim$ 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, $P(X_{\mathbb{L}i} < t, \delta_{\mathbb{L}i} = 1) = 0.63$ and $P(X_{\mathbb{S}i} < t_0, \delta_{\mathbb{S}i} = 1) = 0.70$ 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. Results from estimation of $S(t)$ are shown in Table 1. The bootstrapped standard error estimates approximate

the empirical standard error estimates well. While all estimators have negligible bias, 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) ranging from 1.20-1.24.

For comparing treatment groups, we show results on the estimation precision and power in testing in Table 2. With respect to estimation precision for the treatment difference in survival rate, our proposed procedure yields about 45% gain in efficiency compared to the KM estimator. For testing in detecting the treatment difference, under the null setting (i), type 1 error is close to the nominal level of 0.05 with $\widehat{\Delta}^{\mathbf{w}}_{\text{AUG}}(t)$ having a slightly higher value of 0.058. When there is a small treatment difference, setting (ii), the power to detect a difference in survival at time $t = 2$ is 0.376 using a standard Kaplan-Meier estimate and 0.454 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 0.497. Similarly, in setting (iii), when there is a large treatment difference, power increases from 0.849 to 0.948 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 procedure.

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 and didanosine, zidovudine and zalcitabine, and didanosine only. The long term event of interest $T_{\mathbb{L}}$ = time to death and intermediate event information consists of two intermediate events, $\mathbf{T}_{\mathbb{S}} = (T_{\mathbb{S}1}, T_{\mathbb{S}2})^{\intercal}$ where $T_{\mathbb{S}1}$ = time to an AIDS-defining event e.g. pneumocystis pneumonia and $T_{\mathbb{S}2}$ = 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 or 365 days and $t = 2.5$ years or 915 days and we examine survival in patients from the zidovudine only (mono group, n=619) and zidovudine and zalcitabine (combo group, n=620) groups. Figure 1 displays the Kaplan-Meier 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, zidovudine in the 30 days prior to randomization, days of antiretroviral therapy before randomization. Results are shown in Table 3. Using our proposed estimation procedure the probability of survival incorporating $\mathbf{Z}$ and $\mathbf{T}_{\mathbb{S}}$ information is 0.9276 for the zidovudine treatment group and 0.9531 for the zidovudine and zalcitabin treatment group. The Kaplan-Meier estimate of survival is similar though smaller in both groups.

We are interested in testing the null hypothesis that there is no difference in survival between the two treatment groups at time $t = 2.5$ years: $H_0 : S_{\text{mono}}(2.5) = S_{\text{combo}}(2.5)$ or $H_0 : \Delta(2.5) = S_{\text{combo}}(2.5) - S_{\text{mono}}(2.5) = 0$. 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 0.0252 and our proposed estimators being 0.0255 with or without further augmentation. The proposed procedure provides an estimate which is roughly 25% more efficient than that from the Kaplan-Meier estimate. The p-value for this test decreased from 0.0906 to 0.0552 after incorporating information on baseline covariates, time to an AIDS-defining event, and time to a decrease of 50% in CD4. In this example, augmentation in addition to the two-stage procedure does not seem to alter estimates compared to the two-stage procedure alone.

6 Remarks

We have proposed a two-stage procedure to improve the efficiency in estimating $S(t)$. The proposed procedure leads to increased efficiency in terms of estimation of survival and testing for a difference in survival between two treatment groups. Our key idea is to incorporate intermediate event information and baseline covariate information through landmarking, which overcomes complications that arise in a semi-competing risk setting. When interest lies in comparing survival rates of two treatment groups in a randomized clinical trial setting, we improve efficiency further by using the augmentation concept from Lu & Tsiatis (2008). Our approach can easily incorporate multiple intermediate events, which is appealing when multiple recurrent event information is useful for predicting the terminal events.

We do not discuss the choice of t_0; however, further investigation into optimal selection of t_0 in terms of gaining efficiency is warranted. In practice, if t_0 is too close to t, the subset of patients in Ω_{t_0} may be quite small and could lead to difficulty in estimation. On the other hand, if t_0 is too close to baseline, very little $T_{\mathbb{S}}$ information will be available. A possible approach to overcome this difficulty is to consider several choices of t_0 and then construct a linear combinations of the corresponding $\widehat{\Delta}^{\mathbf{w}}_{\text{AUG}}(t)$ which attains the minimum variance. In some studies, there might be a large number of baseline covariates and many of those may not be helpful in improving the estimation of survival. For such cases, it would be interesting to explore variable selection procedures to identify the important candidates for efficiency improvement. Further research in these area is warranted.

	$\widehat{S}_{\text{KM}}(t)$	$\widehat{S}^{\mathbf{z}}(t)$	$\widehat{S}^{T_{\mathbb{S}}}_{\text{LM}}(t)$	$\widehat{S}^{\mathbf{w}}_{\text{LM}}(t)$
	Treatment A			
Estimate	0.4193	0.4258	0.4227	0.4241
Bias	-0.0022	0.0044	0.0013	0.0027
ESE	0.0233	0.0226	0.0229	0.0214
ASE	0.0231 (1)	0.0221 (1.10)	0.0220 (1.09)	0.0208 (1.24)
MSE	0.0011 (1)	0.0010 (1.07)	0.0010 (1.06)	0.0009 (1.20)
	Treatment B, setting (i)			
Estimate	0.4210	0.4274	0.4240	0.4245
Bias	0.0001	0.0063	0.0030	0.0034
ESE	0.0231	0.0225	0.0224	0.0209
ASE	0.0231 (1)	0.0221 (1.10)	0.0221 (1.09)	0.0208 (1.24)
MSE	0.0011 (1)	0.0010 (1.07)	0.0010 (1.03)	0.0009 (1.21)
	Treatment B, setting (ii)			
Estimate	0.3677	0.3745	0.3706	0.3711
Bias	-0.0010	0.0058	0.0019	0.0024
ESE	0.0226	0.0221	0.0217	0.0202
ASE	0.0227 (1)	0.0217 (1.09)	0.0217 (1.10)	0.0204 (1.25)
MSE	0.0010 (1)	0.0010 (1.09)	0.0009 (1.04)	0.0008 (1.24)
	Treatment B, setting (iii)			
Estimate	0.3249	0.3323	0.3284	0.3285
Bias	-0.0006	0.0067	0.0028	0.0030
ESE	0.0220	0.0214	0.0212	0.0198
ASE	0.0222 (1)	0.0212 (1.09)	0.0213 (1.10)	0.0198 (1.26)
MSE	0.0010 (1)	0.0010 (1.07)	0.0009 (1.03)	0.0008 (1.24)

Table 1: Estimates of $S(t)$ for $t = 2$ years using a Kaplan-Meier estimator, $\widehat{S}_{\text{KM}}(t)$, the proposed estimator using $\mathbf{Z}$ information only, $\widehat{S}^{\mathbf{z}}(t)$, $T_{\mathbb{S}}$ information only, $\widehat{S}^{\mathbf{z}}(t)$, and $\mathbf{Z}$ and $T_{\mathbb{S}}$ information, $\widehat{S}^{\mathbf{w}}_{\text{LM}}(t)$, with corresponding empirical standard error (ESE), average bootstrap standard error (ASE), and mean squared error (MSE). Entries in parentheses are relative efficiencies with respect to $\widehat{S}_{\text{KM}}(t)$

	$\widehat{\Delta}_{\text{KM}}(t)$	$\widehat{\Delta}^{\mathbf{z}}(t)$	$\widehat{\Delta}^{T_{\mathbb{S}}}_{\text{LM}}(t)$	$\widehat{\Delta}^{\mathbf{w}}_{\text{LM}}(t)$	$\widehat{\Delta}^{\mathbf{w}}_{\text{AUG}}(t)$
Setting (i), no treatment difference					
Estimate	-0.0018	-0.0016	-0.0014	-0.0004	-0.0004
ESE	0.0328	0.0319	0.0318	0.0298	0.0274
ASE	0.0326 (1)	0.0313 (1.09)	0.0312 (1.10)	0.0294 (1.23)	0.0273 (1.43)
MSE	0.0011 (1)	0.0010 (1.09)	0.0010 (1.10)	0.0009 (1.24)	0.0007 (1.43)
α	0.0480	0.0520	0.0550	0.0490	0.0580
Setting (ii), small treatment difference					
Estimate	0.0516	0.0513	0.0521	0.0531	0.0531
ESE	0.0327	0.0319	0.0316	0.0293	0.0271
ASE	0.0324 (1)	0.0310 (1.09)	0.0310 (1.09)	0.0291 (1.24)	0.0269 (1.45)
MSE	0.0011 (1)	0.0010 (1.09)	0.0010 (1.09)	0.0008 (1.24)	0.0007 (1.45)
Power	0.3760	0.4020	0.3950	0.4540	0.4970
Setting (iii), large treatment difference					
Estimate	0.0943	0.0935	0.0943	0.0956	0.0956
ESE	0.0319	0.0309	0.0311	0.0288	0.0268
ASE	0.0320 (1)	0.0306 (1.10)	0.0306 (1.09)	0.0287 (1.25)	0.0265 (1.46)
MSE	0.0010 (1)	0.0009 (1.09)	0.0009 (1.09)	0.0008 (1.25)	0.0007 (1.46)
Power	0.8490	0.8720	0.8620	0.9110	0.9480

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, $\widehat{\Delta}_{\text{KM}}(t)$, the proposed estimator using $\mathbf{Z}$ information only, $\widehat{\Delta}^{\mathbf{z}}(t)$, $T_{\mathbb{S}}$ information only, $\widehat{\Delta}^{T_{\mathbb{S}}}_{\text{LM}}(t)$, $\mathbf{Z}$ and $T_{\mathbb{S}}$ information, $\widehat{\Delta}^{\mathbf{w}}_{\text{LM}}(t)$, and an augmented estimator, $\widehat{\Delta}^{\mathbf{w}}_{\text{AUG}}(t)$ with corresponding empirical standard error (ESE), average bootstrap standard error (ASE), and mean squared error (MSE) and power. Entries in parentheses are relative efficiencies with respect to $\widehat{S}_{\text{KM}}(t)$

	$\widehat{S}_{KM}(t)$	$\widehat{S}^{\mathbf{Z}}(t)$	$\widehat{S}^{T_{\mathbb{S}}}_{LM}(t)$	$\widehat{S}^{\mathbf{W}}_{LM}(t)$
Treatment: Zidovudine				
Estimate	0.9245	0.9258	0.9297	0.9276
SE	0.0120	0.0114	0.0114	0.0111
Treatment: Zidovudine and Zalcitabine				
Estimate	0.9498	0.9545	0.9526	0.9531
SE	0.0081	0.0075	0.0079	0.0074

Table 3: Estimates of $S(t)$ for $t = 2.5$ years in two treatment groups from ACTG Protocol 175 using a Kaplan-Meier estimator, $\widehat{S}_{\text{KM}}(t)$, the proposed estimator using $\mathbf{Z}$ information only, $\widehat{S}^{\mathbf{Z}}(t)$, $T_{\mathbb{S}}$ information only, $\widehat{S}^{\mathbf{Z}}(t)$, and $\mathbf{Z}$ and $T_{\mathbb{S}}$ information, $\widehat{S}^{\mathbf{W}}_{\text{LM}}(t)$, with corresponding bootstrap standard error (SE).

	$\widehat{\Delta}_{\text{KM}}(t)$	$\widehat{\Delta}^{\mathbf{Z}}(t)$	$\widehat{\Delta}^{T_{\mathbb{S}}}_{\text{LM}}(t)$	$\widehat{\Delta}^{\mathbf{W}}_{\text{LM}}(t)$	$\widehat{\Delta}^{\mathbf{W}}_{\text{AUG}}(t)$
Estimate	0.0252	0.0287	0.0229	0.0255	0.0255
SE	0.0149	0.0139	0.0138	0.0133	0.0133
RE	1.0000	1.1691	1.1536	1.2541	1.2544
p-value	0.0906	0.0390	0.0972	0.0552	0.0552

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, $\widehat{\Delta}_{\text{KM}}(t)$, the proposed estimator using $\mathbf{Z}$ information only, $\widehat{\Delta}^{\mathbf{Z}}(t)$, $T_{\mathbb{S}}$ information only, $\widehat{\Delta}^{T_{\mathbb{S}}}_{\text{LM}}(t)$, $\mathbf{Z}$ and $T_{\mathbb{S}}$ information, $\widehat{\Delta}^{\mathbf{W}}_{\text{LM}}(t)$, and an augmented estimator, $\widehat{\Delta}^{\mathbf{W}}_{\text{AUG}}(t)$ with corresponding bootstrap standard error (SE) and relative efficiency (RE) with respect to $\widehat{S}_{\text{KM}}(t)$ and p-value

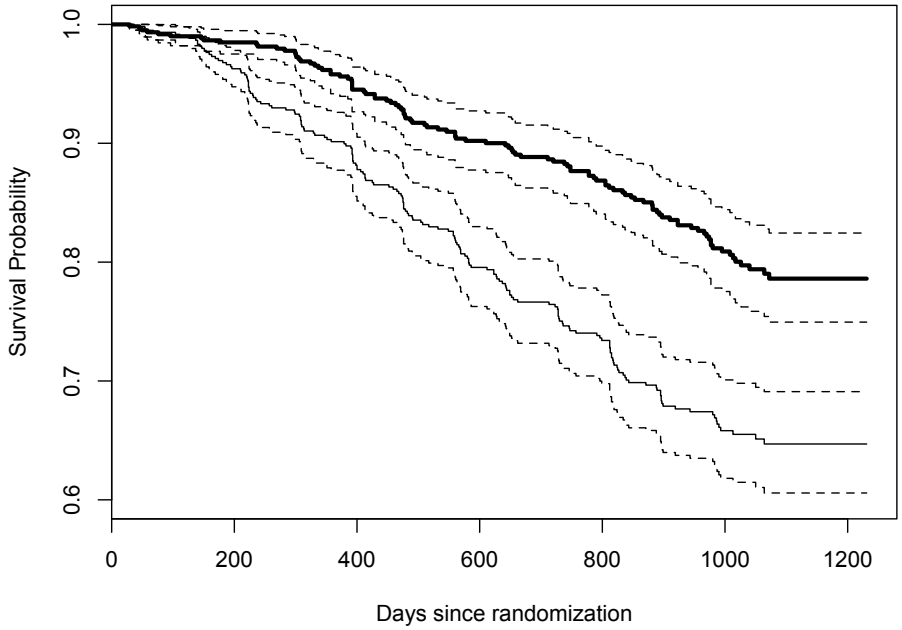

Figure 1: Kaplan-Meier estimate of survival for the mono group (thin black line) and combo group (thick black line) with corresponding 95% confidence intervals (dashed lines).

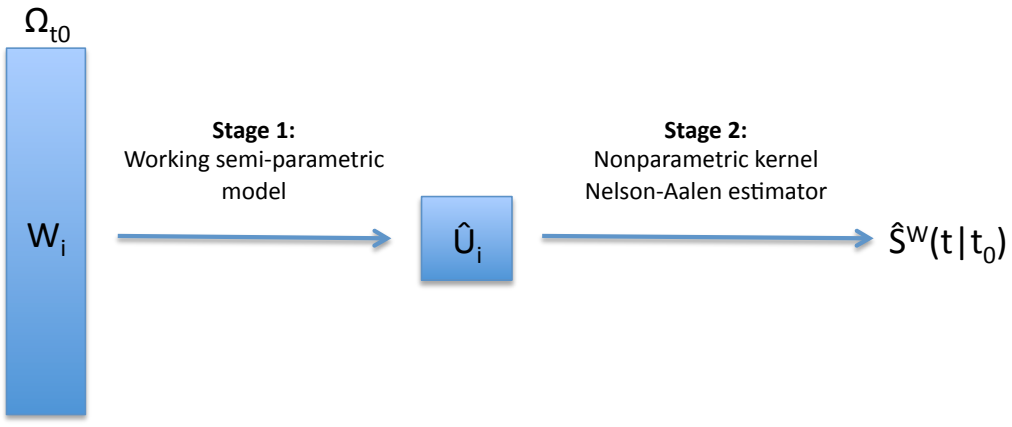

Figure 2: Illustration of proposed two-stage procedure

References

Cai, T., Tian, L., Uno, H., Solomon, S. & Wei, L. (2010). Calibrating parametric subject-specific risk estimation. *Biometrika* **97**, 389–404.

Collins, A., Foley, R., Gilbertson, D. & Chen, S. (2009). The state of chronic kidney disease, ESRD, and morbidity and mortality in the first year of dialysis. *Clinical Journal of the American Society of Nephrology* **4**, S5.

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

Cortese, G. & Andersen, P. (2010). Competing risks and time-dependent covariates. *Biometrical Journal* **52**, 138–158.

Datta, S., Satten, G. & Datta, S. (2000). Nonparametric Estimation for the Three-Stage Irreversible Illness–Death Model. *Biometrics* **56**, 841–847.

Fine, J., Jiang, H. & Chappell, R. (2001). On semi-competing risks data. *Biometrika* **88**, 907–919.

Fleming, T., Prentice, R., Pepe, M. & Glidden, D. (1994). Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and aids research. *Statistics in Medicine* **13**, 955–968.

Gray, R. (1994). A kernel method for incorporating information on disease progression in the analysis of survival. *Biometrika* **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. (1996). 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* **335**, 1081–1090.

Hatteville, L., Mahe, C. & Hill, C. (2002). Prediction of the long-term survival in breast cancer patients according to the present oncological status. *Statistics in Medicine* **21**, 2345–2354.

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

Hjort, N. L. (1992). On inference in parametric survival data models. *International Statistical Review/Revue Internationale de Statistique* **60**, 355–387.

Jiang, H., Fine, J. & Chappell, R. (2005). Semiparametric analysis of survival data with left truncation and dependent right censoring. *Biometrics* **61**, 567–575.

Kay, R. (1986). A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics* **42**, 855–865.

Klein, J., Keiding, N. & Copelan, E. (1994). Plotting summary predictions in multistate survival models: Probabilities of relapse and death in remission for bone marrow transplantation patients. *Statistics in Medicine* **12**, 2315–2332.

Lagakos, S. (1977). Using auxiliary variables for improved estimates of survival time. *Biometrics* pages 399–404.

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

Lu, X. & Tsiatis, A. (2008). Improving the efficiency of the log-rank test using auxiliary covariates.

Biometrika **95**, 679–694.

Parast, L., Cheng, S. C. & Cai, T. (2011). Incorporating short-term outcome information to predict long-term survival with discrete markers. *Biometrical Journal* .

Park, B., Kim, W., Ruppert, D., Jones, M., Signorini, D. & Kohn, R. (1997). Simple transformation techniques for improved non-parametric regression. *Scandinavian Journal of Statistics* **24**, 145–163.

Putter, H., van der Hage, J., de Bock, G., Elgalta, R. & van de Velde, C. (2006). Estimation and Prediction in a Multi-State Model for Breast Cancer. *Biometrical Journal* **48**, 366–380.

Robins, J. & Ritov, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16**, 285–319.

Siannis, F., Farewell, V. & Head, J. (2007). A multi-state model for joint modeling of terminal and non-terminal events with application to whitehall ii. *Statistics in medicine* **26**, 426–442.

USRDS (2010). US Renal Data System 2010 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.

Van Houwelingen, H. & Putter, H. (2008). Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime Data Analysis* **14**, 447–463.

Wand, M., Marron, J. & Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association* **86**, 343–353.