Published in final edited form as:

Lifetime Data Anal. 2010 October ; 16(4): 451–477. doi:10.1007/s10985-009-9149-x.

Estimating treatment effects on the marginal recurrent event mean in the presence of a terminating event

Abstract

In biomedical studies where the event of interest is recurrent (e.g., hospitalization), it is often the case that the recurrent event sequence is subject to being stopped by a terminating event (e.g., death). In comparing treatment options, the marginal recurrent event mean is frequently of interest. One major complication in the recurrent/terminal event setting is that censoring times are not known for subjects observed to die, which renders standard risk set based methods of estimation inapplicable. We propose two semiparametric methods for estimating the difference or ratio of treatment-specific marginal means numbers of events. The first method involves imputing unobserved censoring times, while the second methods uses inverse probability of censoring weighting. In each case, imbalances in the treatment-specific covariate distributions are adjusted out through inverse probability of treatment weighting. After the imputation and/or weighting, the treatment-specific means (then their difference or ratio) are estimated nonparametrically. Large-sample properties are derived for each of the proposed estimators, with finite sample properties assessed through simulation. The proposed methods are applied to kidney transplant data.

Keywords

Censoring; Imputation; Inverse weighting; Marginal mean; Multivariate survival analysis; Semiparametric methods

1 Introduction

Frequently in biomedical studies, the event of interest is recurrent, in that the event can potentially occur multiple times for the same subject. Often, the recurrent event sequence can be stopped permanently by a terminating event (e.g., death). This is often the case when the population being studied has a certain disease or condition. Examples include repeated myocardial infections among patients with cardiovascular disease; repeated opportunistic infections among HIV patients; and multiple hospitalizations among organ transplant recipients. In the study of diseased patients and in the setting where more than one treatment option is available, it is often of interest to compare treatments with respect to the marginal mean number of events. Such comparisons are of interest to XX patients, clinicians and, in particular, health care planners. Although the clinical trial is the preferred design for comparing treatments, in many settings it is illogical and/or unethical to assign treatment randomly. As such, any meaningful comparison XX between two competing treatments will require that any imbalances in the treatment-specific covariate distributions be factored out. In the setting of concern in this article, the recurrent events can be stopped by a terminating event; treatment options are to be compared through the treatment-specific marginal recurrent event means; treatment is not randomized; and it is anticipated that the treatment effect on the mean differs with increasing time since treatment initiation.

A rich body of work exists with respect to the analysis of recurrent event data, as summarized by Cook and Lawless (2007). Focusing on semiparametric methods, conditional approaches were developed by Prentice, Williams and Peterson (1981) and Andersen and Gill (1982), where the term ‘conditional’ refers to the conditionality on some part of the

event history. Lawless and Nadeau (1995) and Lin, Wei, Yang and Ying (2000) developed marginal methods which do not involve conditioning on any part of the event history. Pepe and Cai (1993) proposed what could be termed a partly conditional approach, since their rate model stratified on the previous number of events. Each of the afore-listed methods is built around a Cox (1972) type semiparametric model; i.e., some unspecified baseline function (be it marginal or conditional in nature) which is modulated in a multiplicative fashion by covariates. Schaubel, Zeng and Cai (2006) later proposed a marginal additive rates model, the rate function defined as the derivative of the mean.

For the setting wherein the recurrent event is subject to termination, many methods exist. For example, Li and Lagakos (1997) proposed a modification of the Wei, Lin and Weissfeld (1989) model for multivariate failure times. Cook and Lawless (1997) proposed several procedures for handling recurrent/terminal event data. Ghosh and Lin (2000, 2002) proposed nonparametric tests and semi-parametric regression methods for the marginal recurrent event mean. Various authors proposed models wherein a frailty variate was shared among recurrent and terminal events; e.g., Liu, Wolfe and Huang (2004); Huang and Wang (2004); Ye, Kalbfleisch and Schaubel (2007). Pan and Schaubel (2009) developed methods which combine semiparametric models for the terminal event hazard and conditional recurrent event rate given survival. Recently, Cook et al (2009) proposed various classes of robust methods for modeling the recurrent event mean under various censoring schemes.

XX Of the various methods for modeling recurrent events in the presence of a terminating event, the marginal mean of often of direct interest. Treatment selection by the patient may be mostly driven treatment-specific contrasts with respect to mortality. However, in the setting where different treatments are being used, comparing resource utilization by treatment may be of interest for planning and budgeting purposes. For example, the manager of a dialysis unit may be interested in comparing the mean number of exit site infections between hemodialysis patients by vascular access method. As another example, a hospital administrator may wish to estimate the mean number of hospitalizations among kidney transplant recipients by donor type (e.g., expanded criterion donor (ECD); non-ECD). Days hospitalized may be of greater concern than hospital admissions per se. Of perhaps greater interest than either admissions or hospital days may be hospitalization costs. Note that counting process techniques could be used to analyze each of admissions, days or dollars spent. Therefore the methods described in this article could be applied to any of those events.

With respect to the comparison of treatment-specific marginal mean functions, two major complications in the recurrent/terminal event setting are as follows. First, the treatment effect XX may not be constant over time. Second, censoring times are not observed for subjects observed to die.

Focusing on the first issue, if treatment was randomized, one could compute (then compare) treatment-specific means nonparametrically. When treatment is not randomized, a regression model could be used for covariate adjustment, with the treatment effect being represented by an element of the regression parameter vector. This is a parsimonious way to quantify the treatment effect, provided that the treatment effect is constant over time and hence amenable to representation by a single parameter. However, in the presence of a terminating event, the treatment effect on the mean number of recurrent events is often not constant over time. This is particularly true when the treatment-specific survival functions differ. As follow-up time increases, the composition of the study population shifts due to subjects dying, and the pattern of the shift will be treatment-specific if treatment affects survival. The marginal recurrent event mean can be interpreted as a weighted average across surviving alive and deceased subjects and, when treatment affects survival, the weights essentially change over

time. In cases where the treatment effect is not constant over time, the cumulative effect is of much greater interest for medical decision-making than the instantaneous effect.

Regarding the second issue, censoring times are not observed for subjects who die before they are censored. In settings where random loss to follow-up is not possible, censoring times are essentially known whether or not the subject dies. However, in most practical settings, loss to follow-up occurs, particularly in observational studies. Historically, censoring has been handled in the recurrent/terminal event setting by one of two techniques: by combining censored-data estimators (e.g., Kaplan-Meier estimator of survival; Nelson-Aalen type estimator for the conditional recurrent event rate given survival) or a variant of the well-established inverse probability of censoring weighting (IPCW; Robins and Rotnitzky, 1992; Robins, 1993). Multiple imputation (Rubin, 1987) is a well-known method for missing data problems, and not observing censoring times can be viewed in this fashion. However, to the best of our knowledge, no previous authors have proposed the imputation of unknown censoring times.

In this article, we develop two flexible semiparametric methods to compare treatment-specific marginal mean numbers of recurrent events in the presence of a terminating event. In both classes of methods, inverse probability of treatment weighting (IPTW; Robins et al, 2000; Hernan et al, 2000) is used to balance the treatment-specific adjustment covariate distributions. In the first method, censoring times are multiply imputed when not observed, with the imputation technique being different from that typically employed. In the second proposed method, censoring is dealt with using IPCW, the weight being different from that used previously in the recurrent/terminal event setting. Under the proposed methods, treatment-specific marginal event means can be contrasted through either the difference or the ratio of means, with each viewed as a process over time. The proposed treatment effect estimators are nonparametric in the sense that no functional form of the treatment effect is assumed. Moreover, no form of independence between the recurrent event and terminal events is assumed.

The remainder of this article is organized as follows. In Section 2, we set up notation, then describe the proposed treatment effect measure and methods of estimation. Asymptotic properties of the proposed estimators are listed in Section 3, with proofs provided in the Appendix. A simulation study is carried out in Section 4 in order to evaluate the applicability of the asymptotic results to finite samples. In Section 5, the proposed methods are used to compare mean numbers of post-transplant hospitalizations by type of kidney transplant. The article concludes with a discussion in Section 6.

2 Proposed Methods

2.1 Notation

We begin by setting up the requisite notation. Let i represent subject ($i = 1, \dots, n$). We let A_i be the treatment received by subject i , and set $A_{ij} = I(A_i = j)$. For ease of presentation, we assume that there are only two treatment groups to be compared ($j = 0, 1$). As will become evident later, the proposed methods could easily be extended to handle more than 2 treatment categories. The number of recurrent events for subject i over the $(0, t]$ interval is given by $N_i^*(t)$. For concreteness, we assume that the terminating event is death, with the time of death denoted by D_i . Note that death is assumed to stop the recurrent event process permanently, such that $N_i^*(t) = N_i^*(t \wedge D_i)$, where $a \wedge b = \min\{a, b\}$. The counting process for death is represented by $N_i^{D_i}(t) = I(D_i \leq t)$. Subjects may be right censored, and censoring time is denoted by C_i . Observation time is then given by $X_i = D_i \wedge C_i$, with the at-risk process defined as $Y_i(t) = I(X_i \geq t)$. The observed-data event counters for the recurrent

events and death are denoted by $N_i(t) = N_i^*(t \wedge X_i)$ and $N_i^D(t) = I(D_i \leq t \wedge X_i)$. Finally, we let $\mathbf{Z}_i$ represent the covariate vector, which is taken to be time-constant. We assume that $P(C_j > t/A_i) = P(C_j > t/A_j)$, and define $XX G_j(t) = P(C_j > t/A_i = j) = \exp\{-\Lambda_j^C(t)\}$.

2.2 Background and set-up

Interest lies in the marginal mean function, $\mu(t) = E[N^*(t)]$, where we temporarily omit the subscripts such that the quantities refer to any hypothetical subject. The mean function averages over the survival experience, such that $\mu(t) = E[E[N^*(t)|D]]$. This quantity is most useful from a resource utilization perspective. Note that treatment-specific differences in $\mu(t)$ could arise from differences in treatment-specific survival and/or treatment-specific differences in event rate while subjects are alive. This is made more clear when the mean is expressed as $\mu(t) = \int_0^t P(D > s) E[dN^*(s) | D > s]$. If differences in survival and differences in the event rate among surviving subjects were of primary interest, then an analysis which explicitly modeled $P(D > t)$ and $E[dN^*(t) | D > t]$ would be indicated.

The proportional means model, $\mu(t|\mathbf{Z}) = \mu_0(t) \exp\{\beta_0' \mathbf{Z}\}$, has been developed for the marginal mean function (Ghosh and Lin, 2002). This approach has several advantages, including the fact that treatment effects are concisely described. A potential disadvantage is that the proportionality assumption may be violated if treatment-specific differences in survival exist. By definition, the marginal mean is a weighted average of deceased and surviving subjects. If treatment-specific differences in survival exist, then the treatment-specific weights would change over time and at different rates, resulting in non-proportionality. In such cases, a non-proportional rates model, $d\mu(t|\mathbf{Z}) = d\mu_0(t) \exp\{\beta_0'(t)\mathbf{Z}\}$, is an alternative. However, for treatment decisions made at time $t=0$, if the treatment effect varies with t , then the cumulative effect on the mean is typically of more relevance than the instantaneous effect on the rate. This is especially true when the treatment effect changes direction over time.

Nonparametric procedures for comparing treatment-specific means have been developed (e.g., Ghosh and Lin, 2000). An appealing feature of such approaches is that no functional form is assumed with respect to the relationship between the treatment-specific mean functions. In this article, we aim to develop methods which preserve such flexibility, but allow for covariate adjustment.

Of our interest is the treatment effect on the marginal mean number of recurrent events. If treatment, A_i , was randomized, one could construct a treatment effect measure through $\mu_j^*(t) = E[N_i^*(t) | A_i = j]$ for $j=0, 1$. When treatment is not randomized, one could compare treatments through

$$\mu_j(t) = E[E[N_i^*(t) | A_i = j, \mathbf{Z}_i]], \quad (1)$$

where the outside expectation is with respect to the distribution of $\mathbf{Z}_i$ as opposed to $\mathbf{Z}_i/A_i$, such that $\mu_0(t)$ and $\mu_1(t)$ are averaged with respect to the same covariate distribution. The treatment effect could then be represented by the following measures,

$$\Delta(t) = \mu_1(t) - \mu_0(t) \quad (2)$$

$$\phi(t) = \frac{\mu_1(t)}{\mu_0(t)}, \quad (3)$$

depending on whether the investigator preferred to describe the effect of treatment on the marginal mean in terms of a difference or ratio measure. The quantity $\phi(t)$ may be appealing since mean ratios can be estimated through the regression parameters from the previously-described proportional means model (Ghosh and Lin, 2002). On the other hand, differences in means may be more interpretable and are typically advocated in the causal inference literature.

MMM Further on the subject of causal inference, consider as an example the setting wherein $j = 1$ and $j = 0$ represent treatment and placebo, respectively. We could let $N_{i1}^*(t)$ denote the number of recurrent events in $(0, t]$ is, possibly contrary to fact, subject i received treatment. Similarly, we could define $N_{i0}^*(t)$ as the counterfactual outcome when subject i received placebo. The average causal effect of treatment (vs. placebo) could then be specified as $\Delta^*(t) \equiv E[N_{i1}^*(t)] - E[N_{i0}^*(t)]$. The counterfactuals are connected to the observed data through $N_i^*(t) = A_i N_{i1}^*(t) + (1 - A_i) N_{i0}^*(t)$. Under certain assumptions, that inference targeting $\Delta^*(t)$ can be based on that for $\Delta(t)$. One such condition is the no unmeasured confounders for treatment condition (or equivalently the strong ignorability condition) (Rubin, 1978; Lunceford and Davidian, 2004), i.e., $(N_{i1}^*, N_{i0}^*) \perp A|Z$, and under this condition it can be shown that $\Delta(t) = \Delta^*(t)$. To prove this, we only need to show that $E[N_{ij}^*(t)] = \mu_j(t)$, $j=0, 1$. This follows because

$$\begin{aligned} E[N_{ij}^*(t)] &= E\left[E\left[N_{ij}^*(t) | Z_i\right]\right] \\ &= E\left[E\left[N_{i1}^*(t) | Z_i, A_i=j\right]\right] \quad \text{as } (N_{i1}^*, N_{i0}^*) \perp A|Z \\ &= E\left[E\left[N_{i1}^*(t) | Z_i, A_i=j\right]\right] \quad \text{as } N_i^*(t) = A_i N_{i1}^*(t) + (1 - A_i) N_{i0}^*(t) \\ &= \mu_j(t). \end{aligned}$$

We return to the issue of casual inference in Section 6. Note that our objective is to develop methods through which meaningful inference can be achieved, be it causal or associational.

Consistently estimating $\Delta(t)$ and $\phi(t)$ boils down to consistently estimating $\mu_f(t)$. The two main obstacles are (i) imbalances in the treatment-specific covariate distributions and (ii) censoring; in particular, the fact that censoring times are unknown for subjects observed to die. We propose two methods for estimating $\mu_f(t)$. Both methods accommodate the adjustment covariates XX by using inverse probability of treatment weighting (IPTW; Robins et al, 2000; Hernan et al, 2000; Anstrom and Tsiatis, 2001). The first proposed method employs multiple imputation (Rubin, 1987) to deal with unobserved censoring times. The second method handles censoring through inverse probability of censoring weighting (IPCW; Robins and Rotnitzky, 1992; Robins, 1993).

2.3 Ideal setting: Treatment randomized, no censoring

Suppose that treatment was randomized (such that covariates were distributed equally across treatments) and that complete follow-up was available for all subjects (no censoring). We could then estimate $\mu_f(t)$ by

$$\widehat{\mu}_j^*(t) = \frac{\sum_{i=1}^n I(A_i=j) N_{ij}^*(t)}{\sum_{i=1}^n I(A_i=j)} \quad (4)$$

Notice that the death times, D_j , are not used in (4). Since we are interested in the marginal mean number of events (averaging over the experience before and after D_j), when D_j occurs is not an issue, even though no recurrent events occur following death.

2.4 Accommodating covariate imbalances

Naturally, even without censoring, the estimator in (4) will generally be biased in the absence of randomization due to confounding. However, using inverse probability of treatment weighting (IPTW), one can essentially create treatment-specific pseudo-populations with the same adjustment covariate distribution. In general terms, IPTW weights the contributions of each subject to the estimator (or, estimating function as the case may be) by the inverse of the conditional probability the subject is assigned the treatment they did in fact receive, given the subject's adjustment covariate pattern. The weighted versions of the treatment-specific study samples have the same (joint) adjustment covariate distribution, avoiding confounding (i.e., confounding by measured covariates). Having balanced the treatment-specific adjustment covariate distributions, one could compute treatment-specific recurrent event mean functions through non-parametric procedures applied to the inverse weighted data. An IPTW-based nonparametric estimator of the mean for treatment group j would represent the mean number of recurrent events which would result if the entire study population had received treatment j .

Let $p_{ij} = P(A_j = j | Z_i)$ for $j = 0, 1$, and set $q_{ij} = 1 - p_{ij}$. We assume that the systematic component of treatment assignment is described by the following logistic model,

$$p_{i1}(\theta_0) = P(A_i = 1 | Z_i; \theta_0) = \frac{\exp\{\theta'_0 Z_i\}}{1 + \exp\{\theta'_0 Z_i\}}. \quad (5)$$

Using maximum likelihood, the parameter $\hat{\theta}$ is the solution to the score equation $U(\theta) = 0$, where

$$U(\theta) = \sum_{i=1}^n Z_i \{A_{i1} - p_{i1}(\theta)\}.$$

The estimated weights are given by $w_{ij}(\hat{\theta})$, where

$$w_{ij}(\theta) = \frac{A_{ij}}{p_{ij}(\theta)}.$$

In the absence of censoring, it can be shown that

$$\frac{\sum_{i=1}^n w_{ij}(\hat{\theta}) N_i^*(t)}{\sum_{i=1}^n w_{ij}(\hat{\theta})} \xrightarrow{P} \mu_j(t) \quad (6)$$

under mild regularity conditions.

2.5 Censoring times known

In most practical applications, recurrent and terminating events may be censored. In certain settings, the censoring time for each subject is known at the time the subject enters the

study. An example would be a clinical trial with complete follow-up, such that each subject was followed until a specific (perhaps common) calendar date. In such cases, the censoring time for each subject could be treated as fixed, since C_i is known when subject i begins follow-up at $t = 0$. In this case, $\mu_j(t)$ could be estimated by

$$\widehat{\mu}_j^C(t) = \int_0^t \frac{\sum_{i=1}^n w_{ij}(\widehat{\theta}) dN_i(s)}{\sum_{i=1}^n w_{ij}(\widehat{\theta}) I(C_i > s)}, \quad (7)$$

an IPTW version of the Nelson-Aalen estimator. Note that, in contrast to (6), the estimator in (7) works with the increments in order to guarantee monotonicity.

2.6 Proposed estimators

In most settings, random loss to follow-up is possible. This is a problem in the presence of a terminating event since, unless $C_1, \dots, C_n$ are known, when $D_i < C_i$ we cannot compute $I(C_i > t)$ for $t > D_i$. In such cases, we could not apply an estimator such as (7) which relies on the censoring times being known. XX Right censoring is usually discussed in the context of survival analysis studies; e.g., where time until death is of interest. In the typical survival analysis setting, subjects cease to contribute follow-up after $X_i = D_i \wedge C_i$, the time at which they either die or are censored, whichever occurs first. In such settings, one may wish to impute D_i for censored subjects (e.g., Hsu, Taylor, Murray and Hsu, 2002), but there would typically be little interest in imputing C_i for subjects observed to die. However, in estimating the marginal recurrent event mean, the process $N_i^*(t)$ inherently tracks the experience of individuals past time D_i . Even though it is known that $dN_i^*(t) = 0$ for $t > D_i$, such event-free experience is pertinent to the target parameter of interest. In the presence of random censoring, subjects do not contribute follow-up indefinitely, or even until the end of the study; valid inference on the recurrent event mean function must reflect this property of the data. Just as subjects observed to be censored cease to contribute follow-up beyond time C_i , subjects observed to die should not contribute follow-up beyond the time at which they were scheduled to be censored.

XX If it is known when subject i was scheduled to be censored, such information would be incorporated into the estimation along the lines described in Section 2.5. If censoring times are not known, there are at least two ways to estimate the marginal recurrent event mean. One option is to impute unobserved censoring times. Since imputations are intended to represent random realizations of the data, it is desirable to impute multiple time and average over the results. Multiple imputation (Rubin, 1987) has a rich history and is well-suited to our setting.

XX second option is to weight the observed events, with the intention that the weighted data reflect that which would have been observed in the absence of censoring. Such is the motivation for inverse probability of censoring weighting (IPCW). Although IPCW is typically employed to handle dependent censoring, the idea of re-weighting the uncensored data to factor out the censoring distribution applies well to our setting.

In the subsections that follow, we propose two methods to accommodate the random censoring setting. For each of the two proposed estimators, it will be necessary to estimate the censoring distribution, which can be accomplished XX through

$$\begin{aligned}\widehat{G}_j(t) &= \exp\{-\widehat{\Lambda}_j^C(t)\} \\ \widehat{\Lambda}_j^C(t) &= n^{-1} \sum_{i=1}^n \int_0^t \widehat{\pi}_j(s)^{-1} A_{ij} dN_i^C(s) \\ \widehat{\pi}_j(s) &= n^{-1} \sum_{i=1}^n A_{ij} Y_i(s),\end{aligned}$$

under our previously-stated assumptions.

2.6.1 Imputing censoring times—For our first proposed estimator, we refer back to (7), noting that estimation is straightforward if censoring times were known for all subjects. It is natural to view the censoring times for dead subjects as missing data, and multiple imputation (Rubin, 1987) is a commonly employed method for dealing with missing data. As such, we propose multiply imputing censoring times, for subjects whose death time is observed before their censoring time. For such subjects, valid imputations must incorporate the fact that $C_{\bar{j}}$ although unknown, is known to occur after $D_{\bar{j}}$. Specifically, we propose that, when $D_j < C_{\bar{j}}$ for $m = 1; \dots, M$, impute $C_{i:m}$ from $\widehat{G}_{ij}(t; D_i)$, where

$$G_{ij}(t; D_i) = \sum_{j=0}^1 A_{ij} \frac{G_j(t) I\{t > D_i\}}{G_j(D_i)},$$

the treatment group-specific conditional survival function for $C_{\bar{j}}$. The proposed mean estimator is then given by,

$$\widehat{\mu}_j^M(t) = M^{-1} \sum_{m=1}^M \sum_{i=1}^n \int_0^t \frac{w_{ij}(\widehat{\theta}) dN_i(s)}{\sum_{\ell=1}^n w_{\ell j}(\widehat{\theta}) I(C_{\ell:m} > s)}$$

Note that we are carrying out ‘improper’ imputation (Rubin, 1987), in the sense that standard or ‘proper’ multiple imputation would require (for $m = 1, \dots, M$) generating a realization, say $\widehat{G}_{j:m}(\cdot)$, from the asymptotic distribution of $\widehat{G}_j(\cdot)$; as opposed to generating all imputes from $\widehat{G}_j(\cdot)$ directly. The use of improper imputation means that the multiple imputation variance formula (Rubin, 1987; Little and Rubin, 2002) is inapplicable. Therefore, it is necessary to explicitly account for the estimation of $G_{\bar{j}}(t)$ in deriving large-sample properties of $\widehat{\mu}_j^M(t)$. Note that proper and improper imputation are sometimes referred to as Type A and Type B imputation (Wang and Robins, 1998; Robins and Wang 2000). Improper imputation has been used by several previous authors (e.g., Lu and Tsiatis, 2001; Schaubel and Cai, 2006). Studies comparing the two are quite limited, although it has been reported that improper imputation has greater asymptotic relative efficiency (Robins and Wang, 2000).

2.6.2 (Double) inverse weighting—The second proposed estimator employs the well-established inverse probability of censoring weighting (IPCW) developed by Robins and Rotnitzky (1992). The idea is to replace the unobservable $dN_i^*(t)$ by an observable quantity, which has the same expectation,

$$\frac{dN_i(t)}{G_j(t)}.$$

More specifically, we are replacing $dN_i^*(s)$ with an estimator of a quantity which, asymptotically, has the same expectation. The IPCW estimator we propose is given by

$$\widehat{\mu}_j^G(t) = n^{-1} \sum_{i=1}^n \int_0^t \frac{w_{ij}(\theta)}{\widehat{G}_j(s)} dN_i(s), \quad (8)$$

which is double-weighted (IPTW, IPCW). The quantity $A_{ij}\widehat{G}_j(s)^{-1}dN_i(s)$ is intended to represent the increment that would have been observed in the absence of censoring.

Conceptually, after applying the weight $A_{ij}\widehat{G}_j(s)^{-1}$, a risk-set based estimator, such as (7), is not required since the weighted data reflect the censoring-free setting.

Having estimated the group-specific means, $\mu_j(t)$ for $j = 0, 1$, one could estimate the mean difference through

$$\widehat{\Delta}_j^M(t) = \widehat{\mu}_1^M(t) - \widehat{\mu}_0^M(t) \quad (9)$$

$$\widehat{\Delta}_j^G(t) = \widehat{\mu}_1^G(t) - \widehat{\mu}_0^G(t), \quad (10)$$

for $t \in (0, \tau_U]$, where τ_U is chosen to avoid the well-known instability in the tail of the C_i distribution; or the mean ratio via

$$\widehat{\phi}_j^M(t) = \frac{\widehat{\mu}_1^M(t)}{\widehat{\mu}_0^M(t)} \quad (11)$$

$$\widehat{\phi}_j^G(t) = \frac{\widehat{\mu}_1^G(t)}{\widehat{\mu}_0^G(t)} \quad (12)$$

for $t \in [\tau_L, \tau_U]$, where τ_L is chosen to be far enough away from 0 to avoid infinite or unrealistically large values of $\widehat{\phi}(t)$ resulting from $\widehat{\mu}_0(t) = 0$ or $\widehat{\mu}_0(t) \approx 0$, respectively.

3 Asymptotic properties

To begin, we assume the following regularity conditions for $i = 1, \dots, n$:

- a. $\{C_i, D_i, N_i^*(\cdot), A_i, \mathbf{Z}_i\}$ are independent and identically distributed.
- b. $P(X_i > \tau_U) > 0$.
- c. $\mathbf{Z}_i$ is bounded almost surely.
- d. Positive-definiteness of the matrix, $\mathcal{J}(\theta_0) = E[\mathbf{Z}_i^{\otimes 2} p_{ij}(\theta_0) q_{ij}(\theta_0)]$.
- e. $0 < P(A_i = j | \mathbf{Z}_i) < 1$ for $j = 0, 1$.

We now describe the main asymptotic results of the proposed procedures, beginning with those for the multiple imputation method.

Theorem 1

Under the previously specified regularity conditions, $\widehat{\mu}_j^M(t)$ converges in probability to $\mu_j(t)$ uniformly for in $t \in (0, \tau_U]$, while $n^{1/2} \{\widehat{\mu}_j^M(t) - \mu_j(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function $\sigma_j^M(s, t) = E \{\psi_{ij}(s) \psi_{ij}(t)\}$, where

$$\psi_{ij}(t) = \{\mathbf{b}_j'(t) + \mathbf{k}_j'(t)\} \mathcal{J}(\theta_0)^{-1} U_i(\theta_0) + \int_0^t w_{ij}(\theta_0) H_j(s; \theta_0)^{-1} dM_{ij}(s),$$

with $\mathcal{J}(\theta)$ as defined by condition (d) and

$$\begin{aligned} \mathbf{b}_j(t) &= E \left\{ \int_0^t Z_i H_j(s; \theta_0)^{-1} g_{ij}(\theta_0) dN_i(s) \right\} \\ g_{ij}(\theta) &= (-1)^j A_{ij} \frac{q_{ij}(\theta)}{p_{ij}(\theta)} \\ \mathbf{k}_j(t) &= \int_0^t \mathbf{a}_j(s) H_j(s; \theta_0)^{-1} d\mu_j(s) \\ \mathbf{a}_j(s) &= E \left\{ I(C_i > s) Z_i g_{ij}(\theta_0) \right\} \\ H_j(t; \theta) &= E \left\{ w_{ij}(\theta) I(C_i > s) \right\} \\ U_i(\theta) &= Z_i \{A_{i1} - p_{i1}(\theta)\} \\ dM_{ij}(s) &= A_{ij} \{dN_i(s) - I(C_i > s) d\mu_j(s)\}. \end{aligned}$$

The proof combines the Weak Law of Large Numbers (WLLN), the Multivariate Central Limit Theorem (MCLT) and various results from empirical processes (Pollard, 1009; van der Vaart and Wellner, 1996). A sketch of the proof is provided in the Appendix.

The covariance function in Theorem 1 can be consistently estimated by

$\widehat{\sigma}_j^M(s, t) = n^{-1} \sum_{i=1}^n \widehat{\psi}_{ij}(s) \widehat{\psi}_{ij}(t)$, where $\widehat{\psi}_{ij}(t) = M^{-1} \sum_{m=1}^M \widehat{\psi}_{ij:m}(t)$, where $\widehat{\psi}_{ij:m}(t)$ is the estimator of $\psi_{ij}(t)$ based on the m th set of imputations and is given by

$$\widehat{\psi}_{ij:m}(t) = \{\widehat{\mathbf{b}}_{j:m}(t) + \widehat{\mathbf{k}}_j(t)\} \widehat{I}(\widehat{\theta})^{-1} \widehat{U}_i(\widehat{\theta}) + \int_0^t w_{ij}(\widehat{\theta}) \widehat{H}_{j:m}(s; \widehat{\theta})^{-1} d\widehat{M}_{ij:m}(s)$$

with the empirical counterparts defined as

$$\begin{aligned} \widehat{\mathbf{b}}_{j:m}(t) &= n^{-1} \sum_{i=1}^n \int_0^t \widehat{H}_{j:m}(s; \widehat{\theta})^{-1} g_{ij}(\widehat{\theta}) dN_i(s) \\ \widehat{\mathbf{k}}_{j:m}(t) &= \int_0^t \widehat{\mathbf{a}}_j(s) \widehat{H}_{j:m}(s; \widehat{\theta})^{-1} d\widehat{\mu}_j(s) \\ \widehat{\mathbf{a}}_{j:m}(s) &= n^{-1} \sum_{i=1}^n I(C_{i:m} > s) g_{ij}(\widehat{\theta}) Z_i \\ \widehat{H}_{j:m}(t; \theta) &= n^{-1} \sum_{i=1}^n w_{ij}(\theta) I(C_{i:m} > s) \\ \widehat{I}(\theta) &= n^{-1} \sum_{i=1}^n Z_i^{\otimes 2} p_{ij}(\theta) q_{ij}(\theta) \\ \widehat{U}_i(\theta) &= Z_i \{A_{i1} - p_{i1}(\theta)\} \\ d\widehat{M}_{ij:m}(s) &= A_{ij} \{dN_i(s) - I(C_{i:m} > s) d\widehat{\mu}_{j:m}(s)\}. \end{aligned}$$

Next, we provide the main large-sample results for the group-specific marginal mean estimators based on the proposed IPTW/IPCW method.

Theorem 2

Under the previously listed regularity conditions, $\widehat{\mu}_j^G(t)$ converges in probability to $\mu_j(t)$ uniformly for $t \in (0, \tau_U]$ and $n^{1/2} \{\widehat{\mu}_j^G(t) - \mu_j(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function $\sigma_j^G(s, t) = E \{\xi_{ij}(s) \xi_{ij}(t)\}$, where

$$\begin{aligned}\xi_{ij}(t) &= d_j'(t) \mathcal{J}(\theta_0)^{-1} U_i(\theta_0) + \int_0^t \{\mu_j(t) - \mu_j(s)\} A_{ij} \pi_j(s)^{-1} dM_{ij}^C(s) + \int_0^t \left\{ \frac{w_{ij}(\theta_0) dN_i(s)}{G_j(s)} - d\mu_j(s) \right\} \\ d_j(t) &= E \left[\int_0^t g_{ij}(\theta_0) Z_i G_j(s)^{-1} dN_i(s) \right] \\ \pi_j(s) &= P(X_i \geq s | A_i = j) \\ dM_{ij}^C(s) &= A_{ij} \{dN_i^C(s) - I(X_i > s) d\Lambda_j^C(s)\}.\end{aligned}$$

The covariance function $\sigma_j^G(s, t)$ can be consistently estimated by

$$\widehat{\sigma}_j^G(s, t) = n^{-1} \sum_{i=1}^n \widehat{\xi}_{ij}(s) \widehat{\xi}_{ij}(t), \text{ where}$$

$$\begin{aligned}\widehat{\xi}_{ij}(t) &= \widehat{d}_j'(t) \widehat{I}(\widehat{\theta})^{-1} \widehat{U}_i(\widehat{\theta}) + \int_0^t \{\widehat{\mu}_j^G(t) - \widehat{\mu}_j^G(s)\} A_{ij} \widehat{\pi}_j(s)^{-1} d\widehat{M}_{ij}^C(s) + \int_0^t \left\{ \frac{w_{ij}(\widehat{\theta}) dN_i(s)}{\widehat{G}_j(s)} - d\widehat{\mu}_j^G(s) \right\} \\ \widehat{d}_j(t) &= n^{-1} \sum_{i=1}^n \int_0^t g_{ij}(\widehat{\theta}) Z_i \widehat{G}_j(s)^{-1} dN_i(s) \\ \widehat{\pi}_j(s) &= n^{-1} \sum_{i=1}^n A_{ij} Y_i(s), \\ d\widehat{W}_{ij}^C(s) &= A_{ij} \{dN_i^C(s) - Y_i(s) d\widehat{\Lambda}_j^C(s)\}.\end{aligned}$$

In the next two theorems, we list asymptotic results for $\widehat{\Delta}^M(t)$ and $\widehat{\phi}^M(t)$. Large-sample properties for $\widehat{\Delta}^G(t)$ and $\widehat{\phi}^G(t)$ are derived analogously.

Theorem 3

Under the assumed regularity conditions, $\widehat{\Delta}^M(t)$ is uniformly consistent for $\Delta(t)$ over $(0, \tau_U]$, while $n^{1/2} \{\widehat{\Delta}^M(t) - \Delta(t)\}$ converges weakly to a zero-mean Gaussian process with variance function $\sigma_{\Delta}^M(s, t) = E \{ \{\psi_{i1}(s) - \psi_{i0}(s)\} \{\psi_{i1}(t) - \psi_{i0}(t)\} \}$.

It follows from Theorem 1 that $n^{1/2} \{\widehat{\mu}_1^M(t) - \widehat{\mu}_0^M(t)\}$ is asymptotically equivalent to $n^{-1/2} \sum_{i=1}^n \{\psi_{i1}(t) - \psi_{i0}(t)\}$, a scaled sum of zero-mean Normal variates. Convergence in finite dimensions therefore follows from the Multivariate CLT. Tightness can be demonstrated using results from empirical processes, from which convergence to a zero-mean Gaussian process follows.

Analogous results hold for $\widehat{\Delta}^G(t)$, with $\psi_{ij}(t)$ replaced by $\xi_{ij}(t)$.

Theorem 4

Under the listed regularity conditions, $\widehat{\phi}^M(t)$ converges in probability to $\phi(t)$ uniformly for $t \in [\tau_L, \tau_U]$, and $n^{1/2} \{\widehat{\phi}^M(t) - \phi(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function $\sigma_{\phi}(s, t) = E \{ \gamma_{\lambda}(s) \gamma_{\lambda}(t) \}$, where

$$\gamma_i(t) = \mu_0(t)^{-1} \psi_{i1}(t) - \phi_1(t) \mu_0(t)^{-1} \psi_{i0}(t).$$

Theorem 4 follows from Theorems 1-2, continuity arguments and the Functional Delta Method (van der Vaart and Wellner, 1996).

4 Simulation study

We carried out simulation studies in order to examine the finite sample properties of the proposed methods. For each subject, i , we first generated Q_i from a $\text{Gamma}(\sigma_\phi^{-2}, \sigma_\phi^{-2})$ distribution, such that $E(Q_i) = 1$ and $V(Q_i) = \sigma_\phi^2$. We used $\sigma_\phi^2 = 0.5$ for the first half of the simulations and $\sigma_\phi^2 = 1$ for the second half. We then generated the adjustment covariate, $\mathbf{Z}_i = (Z_{i1}, Z_{i2})'$, where both Z_{i1} and Z_{i2} follow independent Bernoulli(0.5) distributions. The group indicator ($A_i = 0, 1$) was then generated through the model,

$$P(A_i = 1) = \frac{\exp\{\theta_0' \mathbf{Z}_i\}}{1 + \exp\{\theta_0' \mathbf{Z}_i\}},$$

where $\theta_0' = (-0.25, 0.25, 0.3)$. The terminal event survival function was specified as

$P(D_i > t | \mathbf{Z}_i, Q_i) = \exp\{-Q_i \lambda_0 \exp\{\delta_0' \mathbf{Z}_i\} t\}$, where $\mathbf{Z}_i' = (A_i, \mathbf{Z}_i')$, $\lambda_0 = 0.25$ and $\delta_0' = [\log(1.5), \log(1.3), \log(1.2)]$. The conditional recurrent event rate given survival was specified as

$E[dN_i^*(t) | D_i > t, A_i, \mathbf{Z}_i, Q_i] = Q_i r_0 \exp\{\beta_0' \mathbf{Z}_i\}$, where $\beta_0' = [\log(2), \log(1.3), \log(1.2)]$. We varied r_0 from 0.25, 0.5 and 0.75. Finally, the censoring time was generated as a Uniform(0,5). The sample size was set to $n = 1,000$ and each data configuration was replicated 1,000 times. In carrying out the multiple imputation method, we used $M = 10$ imputations.

Results for $\widehat{\mu}_0^M(t)$ and $\widehat{\mu}_0^G(t)$ are reported in Table 1, where *ASE* and *ESD* denote average asymptotic standard error and empirical standard deviation, respectively; while *CP* represents empirical coverage probability. It appears that both proposed marginal mean estimators and their respective asymptotic standard errors are approximately unbiased. Note that patterns in the results for $\widehat{\mu}_1^M(t)$ and $\widehat{\mu}_1^G(t)$ were very similar to those reported in Table 1 and therefore not shown. Results for $\widehat{\Delta}_M(t)$ and $\widehat{\Delta}_G(t)$ are reported in Table 2, while those for $\widehat{\phi}^M(t)$ and $\widehat{\phi}^G(t)$ are listed in Table 3. Both difference estimators and both ratio estimators appear to be approximately unbiased, with estimated standard errors which are close to their respective empirical counterparts, and coverage probabilities which approximate the nominal value.

5 Application

Data were obtained from the Scientific Registry of Transplant Recipients (SRTR) and the Pennsylvania Health Care Cost Containment Council (PHC4). Specifically, patient level demographic, clinical and mortality data were obtained from the SRTR, while information on hospitalizations were obtained from the PHC4. The two data sources were merged through a unique identifier. The study population consisted of all patients who received a deceased-donor first kidney transplant in the state of Pennsylvania between 01/01/1997 and 12/31/2002. To promote homogeneity, our inclusion criteria required that the patient be a

resident of PA, have their initial wait listing in PA, and receive a transplant at a PA facility. Pediatric cases (age <18) were excluded. The final study population consisted of $n = 2,022$ patients. Patients began follow-up upon being discharged from the hospital following kidney transplantation. Each was followed until the earliest of death, loss to follow-up or the conclusion of the observation period (12/31/2002).

The recurrent event of interest was hospital admission, while the terminating event was death. Our chief objective was to compare the marginal hospitalization mean between two types of deceased-donor transplants: expanded criteria donor (ECD) and non-ECD transplants. Kidneys classified as ECD are known to be at higher risk of graft failure than non-ECD kidneys (Port et al, 2002), although the association between ECD and hospitalization is much less well-studied. By the definition of Port et al (2002), a deceased donor is an ECD if the donor is age ≥ 60 ; or if the donor is age 50-59 with ≥ 2 of the following 3 conditions: history of hypertension, serum creatinine >1.5 , death due to stroke. The ECD classification applies to deceased donors only, which is why living donor transplants were not included in this analysis. Adjustment covariates included age at transplant, gender, diabetes status, years on dialysis prior to wait listing, years on wait list prior to transplantation, and number of pre-transplant hospitalizations.

Results presented are based on the proposed IPTW/IPCW method. The marginal mean functions for ECD and non-ECD transplant recipients are displayed in Figure 1. The mean number of hospitalizations per patient was estimated to be 4.7 for ECD patients and 3.4 for non-ECD, based on the first 4 post-transplant years of years of follow-up.

The mean difference and mean ratio estimators (1=ECD; 0=non-ECD), along with corresponding point-wise 95% confidence intervals, are plotted in Figures 2 and 3, respectively. As shown in Figure 2, the mean difference increases with increasing follow-up time, to a maximum of $\widehat{\Delta}(t) = 1.27$ hospital admissions per patient at $t = 4$ years post-transplant. Judging by the confidence intervals, the null hypothesis of a zero difference between ECD and non-ECD transplants with respect to marginal mean number of hospital admissions would be rejected for every month after month 1. As is evident from Figure 3, the mean ratio is approximately constant throughout the 0-4 year post-transplant interval. After 4 years of follow-up, it is estimated that ECD kidney recipients have 1.37 times as many hospitalizations as comparable non-ECD patients.

6 Discussion

We have developed semiparametric methods for comparing treatment-specific recurrent event marginal means in the presence of a terminating event, independent right censoring and non-randomized treatment assignment. The methods are nonparametric in the sense that no structure is imposed on the nature of the dependence among the recurrent events within subject, or on the nature of the dependence between the recurrent and terminal events. Under the proposed procedures, treatment-specific adjustment covariate distributions are balanced using inverse probability of treatment weighting, such that comparisons between the marginal means for each treatment are not confounded by measured adjustment covariates. Asymptotic properties of the proposed estimators were derived and demonstrated through simulation to be applicable in finite samples. Application of our proposed methods reveals that expanded criteria donor kidney transplantation is associated with a significantly greater number of post-transplant hospital admissions than non-ECD transplantation, in spite of the fact that ECD transplant recipients have a significantly higher death hazard (Port et al, 2002).

The treatment effect can be described through the difference or ratio of treatment-specific marginal means. It might be expected that investigators would prefer to use the mean difference, due to its ease of interpretation. The mean ratio may also be of interest, due to its connection with the regression coefficient estimated through the proportional means model (Lin et al, 2000; Ghosh and Lin, 2002).

A major complication in the analysis of recurrent and terminal events is that the censoring time is not known for subjects observed to die. We propose two methods for handling this issue. The first proposed estimation procedure involves multiply imputing censoring times which are not observed, while the second method utilizes inverse probability of censoring weighting, such that the resulting estimators are double weighted (i.e., IPTW/IPCW). Both procedures are easy to implement, in that (after some straightforward data manipulation) Cox regression software that supports weights and offsets can be used to compute the point estimator; e.g., PROC PHREG in SAS, `coxph()` function in R/S-PLUS. Although neither of the two methods is computationally intense, the IPTW/IPCW procedure is much faster than that utilizing multiple imputation.

Our proposed methods require that the censoring time be conditionally independent of the adjustment covariates, given the treatment. This assumption would hold in many practical applications. For example, if censoring is largely administrative (e.g., data collection ceases on a particular calendar date), it would be quite reasonable to assume that covariates do not affect the censoring hazard. In settings where non-administrative loss to follow-up was frequent, the assumption would be less plausible. It would be straightforward to evaluate whether or not the independent censoring assumption holds. For example, one could fit a treatment stratified Cox (1972) model with the adjustment covariates as predictors. In addition, the consistency of our estimators requires that the logistic model for treatment assignment be correct. The well-established diagnostics for generalized linear models may be helpful in this regard.

MMM As stated in Section 2.2, the treatment effects targeted in this article [e.g., $\Delta(t)$] will have a causal interpretation under the no unmeasured confounders condition (described in Section 2.2). In addition to this, for the proposed methods to make valid causal inference, it is required that $P(A_i = j | \mathbf{Z}_i) \in (0, 1)$. This is known as the positivity assumption, and is listed as regularity Condition (e) in Section 3. This assumption would be violated if certain treatment types were never (or always) assigned to subjects with specific covariate patterns. The average causal effect is intended to contrast two hypothetical scenarios; one in which treatment is assigned to all subjects, and one in which placebo is given to all subjects. If the positivity assumption fails, then the contrast of interest cannot be identified using the proposed methods. With respect to censoring, we require that censoring not depend on $\mathbf{Z}_i$ given A_i . The assumptions required for valid causal inference may appear to be quite strong, but very similar assumptions are made for valid associational inference. In the end, one would not expect these assumptions to be violated to some extent in any observational study. When such violations are not pronounced (e.g., provided that data are available on the most important predictors of treatment), meaningful inference contrasting the marginal recurrent event means should be possible through the methods we propose.

APPENDIX

Proof of Theorem 1

With respect to the consistency of $\widehat{\mu}_{j:m}(t)$, since $\widehat{\theta} \xrightarrow{P} \theta_0$ and hence $w_{ij}(\widehat{\theta}) \xrightarrow{P} w_{ij}(\theta_0)$ by the Weak Law of Large Numbers (WLLN) and Continuous Mapping Theorem, it follows that

$$\widehat{\mu}_{j:m}(t) \xrightarrow{P} \int_0^t \frac{E[w_{ij}(\theta_0) dN_i(s)]}{E[w_{ij}(\theta_0) I(C_{i:m} > s)]}. \quad (13)$$

Regarding the numerator in (13),

$$\begin{aligned} E[w_{ij}(\theta_0) dN_i(s)] &= E\left[\frac{A_{ij}I(C_i > s)dN_i^*(s)}{P(A_i=j|Z_i)}\right] \\ &= E\left[E\left[\frac{A_{ij}I(C_i > s)dN_i^*(s)}{P(A_i=j|Z_i)} \middle| Z_i\right]\right] \\ &= E\left[\frac{1}{P(A_i=j|Z_i)} E[I(C_i > s) dN_i^*(s) | Z_i, A_i=j] P(A_i=j|Z_i)] \quad (14) \\ &= E\left[E[I(C_i > s) dN_i^*(s) | Z_i, A_i=j]\right] \\ &= E\left[P(C_i > s | Z_i, A_i=j) E[dN_i^*(s) | Z_i, A_i=j]\right] \\ &= G_j(s) E\left[E[dN_i^*(s) | Z_i, A_i=j]\right] \\ &= G_j(s) d\mu_j(s), \end{aligned}$$

where the outside expectation, in all iterated expectations, is with respect to the marginal distribution of Z_i . Regarding the denominator of (13),

$$E[w_{ij}(\theta_0) I(C_{i:m} > s)] = E\left[\frac{A_{ij}I(C_{i:m} > s)}{P(A_i=j|Z_i)}\right] = G_j(s). \quad (15)$$

Combining (14) and (15), then using applying continuity once more, $\widehat{\mu}_{j:m}(t) \xrightarrow{P} \mu_j(t)$. Since an average of consistent estimators will also be consistent, $\widehat{\mu}_j^M(t) \xrightarrow{P} \mu_j(t)$.

With respect to asymptotic normality, we begin by expressing the estimator based on the m th imputation as follows,

$$n^{1/2} \{\widehat{\mu}_{j:m}(t) - \mu_j(t)\} = n^{1/2} \{\widehat{\mu}_{j:m}(t; \widehat{w}, \widehat{H}) - \widehat{\mu}_{j:m}(t; w_0, \widehat{H})\} \quad (16)$$

$$+ n^{1/2} \{\widehat{\mu}_{j:m}(t; w_0, \widehat{H}) - \widehat{\mu}_{j:m}(t; w_0, H_0)\} \quad (17)$$

$$+ n^{1/2} \{\widehat{\mu}_{j:m}(t; w_0, \widehat{H}) - \mu_j(t)\}, \quad (18)$$

where $\widehat{\mu}_{j:m}(t; \widehat{w}, \widehat{H}) = \widehat{\mu}_{j:m}(t)$ and

$$\begin{aligned} \widehat{\mu}_{j:m}(t; w_0, \widehat{H}) &= \sum_{i=1}^n \int_0^t \widehat{H}_{j:m}(s; \widehat{\theta})^{-1} w_{ij}(\theta_0) dN_i(s) \\ \widehat{\mu}_{j:m}(t; w_0, H_0) &= \sum_{i=1}^n \int_0^t H_{j:m}(s; \theta_0)^{-1} w_{ij}(\theta_0) dN_i(s). \end{aligned}$$

Asymptotically, we can write (16) as

$$n^{1/2} \{\widehat{\mu}_{j:m}(t; \widehat{w}, \widehat{H}) - \widehat{\mu}_{j:m}(t; w_0, \widehat{H})\} = n^{-1/2} \sum_{i=1}^n \widehat{H}_{j:m}(s; \widehat{\theta})^{-1} \{w_{ij}(\widehat{\theta}) - w_{ij}(\theta_0)\} dN_i(s). \quad (19)$$

Using a linear Taylor series expansion, we have

$n^{1/2}(\widehat{\theta} - \theta_0) = \mathcal{J}(\theta_0)^{-1} n^{-1/2} \sum_{i=1}^n U_i(\theta_0) + o_p(1)$, where $\mathcal{J}(\theta)$ and $U_i(\theta)$ are as defined in Theorem 1. By the Delta method, we then have

$$n^{1/2} \{w_{ij}(\widehat{\theta}) - w_{ij}(\theta_0)\} = g_{ij}(\theta_0) Z_i' \mathcal{J}(\theta_0)^{-1} n^{-1/2} \sum_{\ell=1}^n U_{\ell}(\theta_0) + o_p(1), \quad (20)$$

with $g_{ij}(\theta)$ as defined in Theorem 1. Substituting (20) into (19), we obtain

$$\begin{aligned} & n^{1/2} \{ \widehat{\mu}_{j:m}(t; \widehat{w}, \widehat{H}) - \widehat{\mu}_{j:m}(t; w_0, \widehat{H}) \} \\ &= n^{-1} \sum_{i=1}^n \widehat{H}_{j:m}(s; \widehat{\theta})^{-1} g_{ij}(\theta_0) Z_i' \mathcal{J}(\theta_0)^{-1} n^{-1/2} \sum_{\ell=1}^n U_{\ell}(\theta_0) + o_p(1). \end{aligned}$$

Applying the WLLN and the Continuous Mapping Theorem,

$$n^{1/2} \{ \widehat{\mu}_{j:m}(t; \widehat{w}, \widehat{H}) - \widehat{\mu}_{j:m}(t; w_0, \widehat{H}) \} = \mathbf{b}_j'(t) \mathcal{J}(\theta_0)^{-1} n^{-1/2} \sum_{i=1}^n U_i(\theta_0) + o_p(1), \quad (21)$$

with $\mathbf{b}_j(t)$ defined in Theorem 1.

We can express (17) as follows,

$$\begin{aligned} & n^{1/2} \{ \widehat{\mu}_{j:m}(t; w_0, \widehat{H}) - \widehat{\mu}_{j:m}(t; w_0, H_0) \} \\ &= n^{-1/2} \sum_{i=1}^n \int_0^t w_{ij}(\theta_0) \{ \widehat{H}_{j:m}(s; \widehat{\theta})^{-1} - H_j(s; \theta_0)^{-1} \} dN_i(s) + o_p(1). \end{aligned} \quad (22)$$

Using the fact that $n^{1/2} \{ \widehat{H}_{j:m}(s; \widehat{\theta}) - H_j(s; \theta_0) \} = n^{-1/2} \sum_{i=1}^n I(C_{i:m} > s) \{ w_{ij}(\widehat{\theta}) - w_{ij}(\theta_0) \} + o_p(1)$, then applying the Delta Method, we have

$$\begin{aligned} & n^{1/2} \{ \widehat{H}_{j:m}(s; \widehat{\theta})^{-1} - H_j(s; \theta_0)^{-1} \} \\ &= -H_j(s; \theta_0)^{-2} n^{-1/2} \sum_{k=1}^n I(C_{k:m} > s) g_{kj}(\theta_0) Z_k' \mathcal{J}(\theta_0)^{-1} n^{-1/2} \sum_{\ell=1}^n U_{\ell}(\theta_0) + o_p(1), \end{aligned}$$

such that we can re-write (22) as

$$\begin{aligned} & n^{1/2} \{ \widehat{\mu}_{j:m}(t; w_0, \widehat{H}) - \widehat{\mu}_{j:m}(t; w_0, H_0) \} \\ &= -n^{-1} \sum_{i=1}^n \int_0^t w_{ij}(\theta_0) H_j(s; \theta_0)^{-2} n^{-1} \sum_{k=1}^n I(C_{k:m} > s) g_{kj}(\theta_0) Z_k' \mathcal{J}(\theta_0)^{-1} n^{-1/2} \sum_{\ell=1}^n U_{\ell}(\theta_0) dN_i(s) + o_p(1). \end{aligned}$$

After switching the order of summation and applying the WLLN and continuity arguments, we then have

$$n^{1/2} \{ \widehat{\mu}_{j:m}(t; w_0, \widehat{H}) - \widehat{\mu}_{j:m}(t; w_0, H_0) \} = \mathbf{k}_j'(t) \mathcal{J}(\theta_0)^{-1} n^{-1/2} \sum_{i=1}^n U_i(\theta_0) + o_p(1), \quad (23)$$

with $\mathbf{k}_j(t)$ as defined by Theorem 1.

Finally, it is straightforward that (18) can be written

$$n^{1/2} \left\{ \widehat{\mu}_{j:m}(t; w_0, H_0) - \mu_j(t) \right\} = n^{-1/2} \sum_{i=1}^n \int_0^t H_j(s; \theta_0)^{-1} w_{ij}(\theta_0) \left\{ dN_i(s) - I(C_{i:m} > s) d\mu_j(s) \right\} \quad (24)$$

Combining (21), (23) and (24), then averaging over the M imputations, it follows that $n^{1/2} \left\{ \widehat{\mu}_j^M(t) - \mu_j(t) \right\} = n^{-1/2} \sum_{i=1}^n \bar{\psi}_{ij}(t)$, where, as $n \rightarrow \infty$ the $\bar{\psi}_{ij}(t)$ are independent and identically distributed mean 0 variates. Therefore, by the Central Limit Theorem, for fixed t , $n^{1/2} \left\{ \widehat{\mu}_j^M(t) - \mu_j(t) \right\}$ converges in distribution to a zero-mean Normal distribution with variance $E \left[\bar{\psi}_{ij}(t)^2 \right]$. Each of the variates, $\bar{\psi}_{ij}(t)$, is manageable (Pollard, 1990) since each can be written as sums of monotone functions. Hence, $n^{1/2} \left\{ \widehat{\mu}_j^M(t) - \mu_j(t) \right\} = n^{-1/2} \sum_{i=1}^n \bar{\psi}_{ij}(t)$ is tight and therefore converges to a zero-mean Gaussian process, thus completing the proof.

Proof of Theorem 2

With respect to consistency, using the facts that $\widehat{G}_j(s) \xrightarrow{P} G_j(s)$ (Fleming and Harrington, 1991) and $w_{ij}(\widehat{\theta}) \xrightarrow{P} w_{ij}(\theta_0)$, then applying the WLLN and using continuity,

$$\widehat{\mu}_j^G(t) \xrightarrow{P} \int_0^t \frac{E \left[w_{ij}(\theta_0) dN_i(s) \right]}{G_j(s)} = \mu_j(t),$$

where the equality follows from arguments similar to those used in the proof of Theorem 1.

To demonstrate asymptotic normality, we begin with the following decomposition,

$$n^{1/2} \left\{ \widehat{\mu}_j^G(t) - \mu_j(t) \right\} = n^{1/2} \left\{ \widehat{\mu}_j^G(t; \widehat{G}, \widehat{\theta}) - \widehat{\mu}_j^G(t; \widehat{G}, \theta_0) \right\} \quad (25)$$

$$+ n^{1/2} \left\{ \widehat{\mu}_j^G(t; \widehat{G}, \theta_0) - \widehat{\mu}_j^G(t; G, \theta_0) \right\} \quad (26)$$

$$+ n^{1/2} \left\{ \widehat{\mu}_j^G(t; G, \theta_0) - \mu_j(t) \right\}, \quad (27)$$

where

$$\widehat{\mu}_j^G(t; F, \theta) = \sum_{i=1}^n \int_0^t w_{ij}(\theta) F_j(s)^{-1} dN_i(s), \quad (28)$$

where F is set to either $\widehat{G}$ or G , as required by a particular calculation. The term (25) can be expressed as

$$n^{1/2} \left\{ \widehat{\mu}_j^G(t; \widehat{G}, \widehat{\theta}) - \widehat{\mu}_j^G(t; \widehat{G}, \theta_0) \right\} = n^{-1/2} \sum_{i=1}^n \int_0^t \left\{ w_{ij}(\widehat{\theta}) - w_{ij}(\theta_0) \right\} \widehat{G}_j(s)^{-1} dN_i(s). \quad (29)$$

Subbing (20) into (29), we have

$$n^{1/2} \left\{ \widehat{\mu}_j^G(t; \widehat{G}, \widehat{\theta}) - \widehat{\mu}_j^G(t; \widehat{G}, \theta_0) \right\} = n^{-1} \sum_{i=1}^n \int_0^t g_{ij}(\theta_0) Z_i' \mathcal{J}(\theta_0)^{-1} n^{-1/2} \sum_{\ell=1}^n U_\ell(\theta_0) \widehat{G}_j(s)^{-1} dN_i(s) + o_p(1).$$

Switching the order of summation, then applying the WLLN and continuity, we have

$$n^{1/2} \left\{ \widehat{\mu}_j^G(t; \widehat{G}, \widehat{\theta}) - \widehat{\mu}_j^G(t; \widehat{G}, \theta_0) \right\} = \mathbf{d}_j'(t) \mathcal{J}(\theta_0)^{-1} n^{-1/2} \sum_{i=1}^n U_i(\theta_0) + o_p(1), \quad (30)$$

with $\mathbf{d}_j(t)$ defined in Theorem 1.

We can write (26) as follows,

$$n^{1/2} \left\{ \widehat{\mu}_j^G(t; \widehat{G}, \theta_0) - \widehat{\mu}_j^G(t; G, \theta_0) \right\} = n^{-1/2} \sum_{i=1}^n \int_0^t w_{ij}(\theta_0) \left\{ \widehat{G}_j(s)^{-1} - G_j(s)^{-1} \right\} dN_i(s). \quad (31)$$

Using Martingale theory (Andersen et al, 1993), it can be shown that

$$n^{1/2} \left\{ \widehat{\Lambda}_j^C(t) - \Lambda_j^C(t) \right\} = n^{-1/2} \sum_{i=1}^n \int_0^t \pi_j(s)^{-1} dM_{ij}^C(s) + o_p(1),$$

such that, applying the Functional Delta Method (van der Vaart and Wellner, 1996; van der Vaart, 2000),

$$n^{1/2} \left\{ \widehat{G}_j(s)^{-1} - G_j(s)^{-1} \right\} = -G_j(s)^{-1} n^{-1/2} \sum_{i=1}^n \int_0^s \pi_j(r)^{-1} dM_{ij}^C(r) + o_p(1). \quad (32)$$

Substituting (32) into (31), we obtain

$$\begin{aligned} & n^{1/2} \left\{ \widehat{\mu}_j^G(t; \widehat{G}, \theta_0) - \widehat{\mu}_j^G(t; G, \theta_0) \right\} \\ &= n^{-1/2} \sum_{i=1}^n \int_0^t w_{ij}(\theta_0) G_j(s)^{-1} n^{-1} \sum_{\ell=1}^n \int_0^s \pi_j(r)^{-1} dM_{\ell j}^C(r) dN_i(s) \\ &= n^{-1/2} \sum_{\ell=1}^n \int_0^t n^{-1} \sum_{i=1}^n \int_r^t w_{ij}(\theta_0) G_j(s)^{-1} dN_i(s) \pi_j(r)^{-1} dM_{\ell j}^C(r), \end{aligned}$$

where the last expression is obtained by switching the orders of summation and integration.

Using the consistency of $\widehat{\mu}_j^G(t)$, we have

$$n^{1/2} \left\{ \widehat{\mu}_j^G(t; \widehat{G}, \theta_0) - \widehat{\mu}_j^G(t; G, \theta_0) \right\} = n^{-1/2} \sum_{i=1}^n \int_0^t \left\{ \mu_j(t) - \mu_j(s) \right\} A_{ij} \pi_j(s)^{-1} dM_{ij}^C(s) + o_p(1) \quad (33)$$

Finally, (27) can be written as

$$n^{1/2} \left\{ \widehat{\mu}_j^G(t; G, \theta_0) - \mu_j(t) \right\} = n^{-1/2} \sum_{i=1}^n \int_0^t \left\{ w_{ij}(\theta_0) G_j(s)^{-1} dN_i(s) - d\mu_j(s) \right\}. \quad (34)$$

Combining (30), (33) and (34), $n^{1/2} \left\{ \widehat{\mu}_j^G(t) - \mu_j(t) \right\} = n^{-1/2} \sum_{i=1}^n \xi_{ij}(t)$, a scaled sum of independent and identically distributed mean zero variates, asymptotically. Analogous to the proof of Theorem 1, the finite dimensional distributions converge to a multivariate Normal

by the Multivariate CLT. The demonstration of tightness (Pollard, 1990), which follows from various empirical process results (van der Vaart and Wellner, 1996; van der Vaart, 2000) completes the proof of convergence to a zero-mean Gaussian process.

References

- Andersen, PK.; Borgan, O.; Gill, RD.; Keiding, N. Statistical models based on counting processes. Springer; New York: 1993.
- Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. *Ann Stat*. 1982; 10:1100–1120.
- Anstrom KJ, Tsiatis AA. Utilizing propensity scores to estimate causal treatment effects with censored time-lagged data. *Biometrics*. 2001; 57:1207–1218. [PubMed: 11764262]
- Biliyas Y, Gu M, Ying Z. Towards a general asymptotic theory for the Cox model with staggered entry. *Annals of Statistics*. 1997; 25:662–682.
- Cook RJ, Lawless JF. Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine*. 1997; 16:911–24. [PubMed: 9160488]
- Cook, RJ.; Lawless, JF. The analysis of recurrent event data. Springer; New York: 2007.
- Cook RJ, Lawless JF, Lakhal-Chaieb L, Lee K-A. Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: Application to skeletal complications in cancer metastatic to bone. *J Am Stat Assoc*. 2009; 104:6075.
- Cox DR. Regression Models and Life-Tables. *J Roy Stat Soc*. 1972; B34:187–202.
- Fleming, TR.; Harrington, DP. Counting Processes and survival analysis. Wiley; New York: 1991.
- Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics*. 2000; 56:554–62. [PubMed: 10877316]
- Ghosh D, Lin DY. Marginal regression models for recurrent and terminal event. *Stat Sinica*. 2002; 12:663–88.
- Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of Zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000; 11:561–570. [PubMed: 10955409]
- Taylor JMG, Marray S, Hsu C-H. Survival estimation and testing via multiple imputation. *Statistics and Probability Letters*. 2002; 3:221–232.
- Huang CY, Wang MC. Joint modeling and estimation for recurrent event processes and failure time data. *J Am Stat Assoc*. 2004; 99:1153–1165.
- Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics*. 1995; 37:158–168.
- Li Q, Lagakos S. Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine*. 1997; 16:925–940. [PubMed: 9160489]
- Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J Roy Stat Soc*. 2000; B62:711–730.
- Little, RJA.; Rubin, DB. Statistical analysis with missing data. Wiley; New York: 2002.
- Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics*. 2004; 60:747–56. [PubMed: 15339298]
- Lu K, Tsiatis AA. Multiple imputation methods for estimating regression coefficients in proportional hazards models with missing cause of failure. *Biometrics*. 2001; 57:1191–1197. [PubMed: 11764260]
- Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*. 2004; 23:2937–2960. [PubMed: 15351954]
- Pan Q, Schaubel DE. Flexible estimation of differences in treatment-specific recurrent event means in the presence of a terminating event. *Biometrics*. 2009 in press.
- Pepe M, Cai J. Some Graphical Displays and Marginal Regression Analysis for Recurrent Failure Times and Time Dependent Covariates. *J Am Stat Assoc*. 1993; 88:811–820.

- Pollard, D. Empirical processes: Theory and applications. Institute of Mathematical Statistics; Hayward: 1990.
- Port FK, Bragg-Gresham JL, Metzger RA, Dykstra DM, Gillespie BW, Young EW, Delmonico FL, Wynn JJ, Merion RM, Wolfe RA, Held PJ. Donor characteristics associated with reduced graft survival: An approach to expanding the pool of donor kidneys. *Transplantation*. 2002; 74:1281–1286. [PubMed: 12451266]
- Prentice RL, Williams B, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika*. 1981; 68:373–379.
- Robins, JM. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers; Proceedings of the Biopharmaceutical Section, American Statistical Association; 1993; p. 24-33.
- Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
- Robins, JM.; Rotnitzky, A. Recovery of Information and Adjustment for Dependent Censoring using Surrogate Markers”, *AIDS Epidemiolgy-Methodological Issues*, Proceedings of the Biopharmaceutical Section. In: Jewell, N.; Dietz, K.; Farewell, V., editors. American Statistical Association. 1992. p. 24-33.N.
- Robins JM, Wang N. Inference for imputation estimators. *Biometrika*. 2000; 87:113–124.
- Rubin DB. Bayesian inference for causal effects: The role of randomization. *Ann Stat*. 1978; 6:34–58.
- Rubin, DB. Multiple imputation for nonresponse in surveys. Wiley; New York: 1987.
- Schaubel DE, Cai J. Multiple imputation methods for recurrent event data with missing event category. *Canadian Journal of Statistics*. 2006; 34:677–692.
- Schaubel DE, Zeng D, Cai J. A semiparametric additive rates model for recurrent event data. *Lifetime Data Anal*. 2006; 12:389–406. [PubMed: 17031499]
- van der Vaart, AW. Asymptotic statistics. Cambridge; Melbourne: 2000.
- van der Vaart, AW.; Wellner, JA. Weak convergence and empirical processes. Springer; New York: 1996.
- Wang N, Robins JM. Large-sample theory for parametric multiple imputation estimators. *Biometrika*. 1998; 85:935–948.
- Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *J Am Stat Assoc*. 1989; 84:1065–1073.
- Ye Y, Kalbfleisch JD, Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics*. 2007; 63:78–87. [PubMed: 17447932]

Figure 1.

Analysis of SRTR/PHC4 kidney transplant data: $\tilde{\mu}_j^G(t)$, by month. $j=1$ (solid line) pertains to ECD transplants; $j=0$ (dashed line) pertains to non-ECD transplants.

Figure 2.
Analysis of SRTR/PHC4 kidney transplant data: $\widehat{\Delta}^G(t)$ and pointwise 95% confidence intervals.

Figure 3.

Analysis of SRTR/PHC4 kidney transplant data: $\widehat{\phi}^G(t)$ and pointwise 95% confidence intervals

Table 1

Simulation results: $\mu_0(\delta)$ estimators

r_0	t	$\mu_0(\delta)$	$\hat{\mu}_0^M(t)$				$\hat{\mu}_0^G(t)$			
			BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP
0.50	2	0.42	-0.001	0.0366	0.0366	0.935	-0.000	0.0381	0.0377	0.947
	3	0.53	0.000	0.0457	0.0454	0.944	-0.002	0.0465	0.0467	0.950
	1	0.51	-0.003	0.0467	0.0467	0.944	0.000	0.0306	0.0421	0.961
	2	0.84	0.000	0.0692	0.0686	0.946	0.002	0.0578	0.0617	0.953
	3	1.08	0.001	0.0870	0.0865	0.954	0.004	0.0714	0.0767	0.965
	1	0.76	0.004	0.0505	0.0503	0.945	0.003	0.0508	0.0526	0.953
0.75	2	1.26	0.006	0.0747	0.0755	0.952	0.001	0.0771	0.0795	0.952
	3	1.61	0.005	0.0975	0.0964	0.943	0.003	0.0977	0.1020	0.963

Table 2

Simulation results: $\Delta(t)$ estimators

r_0	$\mu_0(t)$	$\Delta^M(t)$				$\Delta^G(t)$			
		BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP
0.25	1	0.21	0.001	0.0449	0.0464	0.961	0.001	0.0437	0.0446
	2	0.30	0.003	0.0637	0.0661	0.959	-0.001	0.0636	0.0623
	3	0.34	0.004	0.0768	0.0810	0.956	-0.001	0.0763	0.0756
0.50	1	0.41	-0.002	0.0791	0.0887	0.969	-0.002	0.0669	0.0672
	2	0.59	-0.004	0.118	0.131	0.973	-0.004	0.0994	0.0964
	3	0.67	-0.007	0.148	0.163	0.968	-0.006	0.124	0.119
0.75	1	0.62	-0.004	0.0882	0.0995	0.969	0.000	0.0508	0.0526
	2	0.89	-0.003	0.129	0.151	0.975	0.001	0.0771	0.0800
	3	1.01	-0.002	0.163	0.189	0.978	0.003	0.0977	0.1020

Table 3

Simulation results: $\phi(t)$ estimators

r_0	t	$\mu_0(t)$	$\hat{\phi}^M(t)$				$\hat{\phi}^G(t)$			
			BIAS	ESD	ASE	CP	BIAS	ESD	ASE	CP
0.25	1	1.83	0.024	0.243	0.243	0.952	0.027	0.243	0.242	0.955
	2	1.71	0.022	0.200	0.200	0.955	0.012	0.199	0.199	0.952
	3	1.63	0.019	0.184	0.188	0.948	0.013	0.185	0.186	0.951
0.50	1	1.81	0.016	0.212	0.226	0.965	0.006	0.178	0.192	0.961
	2	1.70	0.007	0.180	0.193	0.959	0.002	0.152	0.157	0.966
	3	1.63	0.004	0.172	0.185	0.957	-0.001	0.144	0.147	0.955
0.75	1	1.82	-0.002	0.152	0.164	0.960	0.014	0.156	0.163	0.968
	2	1.70	-0.000	0.129	0.145	0.968	0.008	0.140	0.143	0.958
	3	1.63	0.003	0.127	0.140	0.972	0.004	0.135	0.138	0.951