

Estimating the effect of a rare time-dependent treatment on the recurrent event rate

Abigail R. Smith^{ab}, Danting Zhu^a, Nathan P. Goodrich^b, Robert M. Merion^b, and Douglas E. Schaebel^{a*}

In many observational studies, the objective is to estimate the effect of treatment or state-change on the recurrent event rate. If treatment is assigned after the start of follow-up, traditional methods (e.g., adjustment for baseline-only covariates, or fully conditional adjustment for time-dependent covariates) may give biased results. We propose a two-stage modeling approach using the method of sequential stratification to accurately estimate the effect of a time-dependent treatment on the recurrent event rate. At the first stage, we estimate the pre-treatment recurrent event trajectory using a proportional rates model censored at the time of treatment. Prognostic scores are estimated from the linear predictor of this model and used to match treated patients to as-yet untreated controls based on prognostic score at the time of treatment for the index patient. The final model is stratified on matched sets and compares the post-treatment recurrent event rate to the recurrent event rate of the matched controls. We demonstrate through simulation that bias due to dependent censoring is negligible, provided the treatment frequency is low, and we investigate a threshold at which correction for dependent censoring is needed. The method is applied to liver transplant (LT), where we estimate the effect of development of post-LT End Stage Renal Disease (ESRD) on rate of days hospitalized. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: rate model; recurrent events; sequential stratification; time-dependent treatment; treatment effects

1. Introduction

Recurrent events often serve as the basis for measuring treatment effects in observational studies. A reduction in outcomes, such as repeated myocardial infarction or opportunistic infections, indicates that a treatment has a positive effect on morbidity. Reductions in hospital admission rates among the treatment group would imply that lower morbidity as well as reduced health care costs are associated with (or caused by) treatment.

Methods for analyzing recurrent events have been well described in the literature. Models have been developed that condition on the event history [1] or previous number of events [3]. Marginal models, such as those of Lawless and Nadeau [4] or Lin et al. [5], allow for an interpretation of covariate effects on the recurrent event rate that does not require patients to have similar event histories. Few papers to date have explored methods for using recurrent events as an evaluation of an experimental treatment, with exceptions being Cook et al. [6] and Schaebel and Zhang [7].

Treatment can be initiated after the beginning of follow-up, which occurs frequently in studies without randomization. While some existing recurrent event methods can incorporate time-dependent covariates [8], these traditional methods often do not give interpretations that satisfy the research question of interest. In the settings often of interest, treatment initiation depends on internal processes such as disease progression or the event history itself, violating the assumption of most time-dependent recurrent event methods that time-dependent covariates be external [9]. Ideally, we would begin follow-up of an untreated patient, and after treatment initiation we would compare the recurrent event rate to that of the same patient had they remained untreated. This counterfactual experience is unobservable in practice, however

this is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/sim.7628

^aUniversity of Michigan, Department of Biostatistics, 1415 Washington Heights, Ann Arbor, Michigan 48109-2029; ^bArbor Research Collaborative for Health, 340 E. Huron St, Suite 300, Ann Arbor, Michigan 48104

* Correspondence to: Douglas E. Schaebel, M4057 SPH II, 1415 Washington Heights, Ann Arbor, Michigan 48109-2029

In an attempt to compare each treated subject with their unobservable counterfactual treatment-free experience, this article will extend the sequential stratification method described by Schaubel et al. [10] to the recurrent event setting. For every subject treated at time s , subjects that are eligible to receive treatment at time s but do not are matched to the treated subject. Each treated subjects's post-treatment recurrent event rate is compared to the averaged matched recurrent event rate in what can be conceptualized as a subject-level experiment. Matched subjects that subsequently receive treatment are censored from experiments for which they serve as controls, and begin their own experiment as the treated subject. Note that, in every experiment, the comparison of interest begins at time s , such that recurrent events that occur in $[0, s)$ are not considered.

Schaubel et al [10] proposed combining hard covariate matching and adjustment to ensure that matched subjects were 'similar' to the treated subject in addition to the requirement that they remain untreated at time s . This method was proposed in the univariate survival setting where failure times prior to treatment are not observed for treated subjects. However, information regarding pre-treatment recurrent event trajectories are available on all subjects in the setting described above. Given that event history is a strong predictor of the recurrent event rate, we propose to leverage this information using a two-stage modeling approach. In the first stage, we use a conditional rate model to describe pre-treatment event trajectories for all subjects. We then use the linear predictor from this first stage model to caliper-match as yet untreated patients to those receiving treatment at time s . The goal is to create a control group with an event trajectory similar to that which the treated patient would have experienced had treatment not been available. The final model for the recurrent event rate includes only the treatment effect and a measure of distance between the prognostic score of the treated subject and that of the matched controls.

The method proposed is not restricted to "treatment" in the classical sense, and is in fact applicable to any state change. Often this state change is in the form of treatment such as initiation of new medication or performance of a procedure, but this is not always the case. Diagnosis of disease or experience of a medical event such as injury could constitute a state change for which comparing the recurrent event rate in the presence and absence of the state change is of clinical or policy interest. This will be discussed further in relation to the application of the method to liver transplantation.

The remainder of this article proceeds as follows. In Section 2 we introduce the notation and proposed models and describe the parameter estimation. Section 3 presents results of simulation studies to demonstrate the performance of the treatment effect estimator in moderate sized samples. An application to living donor liver transplant is described in Section 4 using data from the Adult-to-Adult Living Donor Liver Transplantation Cohort Study (A2ALL). Some concluding remarks are offered in Section 5.

2. Methods

2.1. Notation

In the following, i represents subject ($i = 1, \dots, n$), T_i is treatment time, with $T_i \geq 0$, and $Z_i^*(t)$ represents the time-dependent covariate for subject i . We assume for the purposes of this article that subjects treated at time T_i remain treated for the duration of follow-up. The true number of events for subject i in $[0, t]$ is defined as $N_i^*(t) = \int_0^t dN_i^*(u)$. Event and treatment times are subject to independent right censoring by C_i , assumed to be administrative in this setting without loss of generality. The number of observed events is given by $N_i(t) = \int_0^t I(C_i > u) dN_i^*(u)$.

The number of pre-treatment events in $(0, t]$ is given by the counting process

$$N_i^0(t) = \int_0^t I(T_i > u) dN_i^*(u). \quad (1)$$

If patient i receives treatment at time s ; i.e., $T_i = s$, then the post-treatment event counter is defined as

$$N_i^1(t; s) = I(T_i = s) \int_s^{s+t} dN_i^*(u). \quad (2)$$

Note that it will be our convention that $N(t; s)$ refers to the interval of length t , but starting at time s ; a single time index, as in the previously-defined $N_i^0(t)$, pertains to the $(0, t]$ time interval. Correspondingly, we define an event counter representing the events that would have been experienced in the absence of treatment, also beginning at time s ,

$$N_i^0(t; s) = \int_s^{s+t} I(T_i > u) dN_i^*(u). \quad (3)$$

Note that (3) is the pre-treatment event counter described in (1) but instead of $(0, t]$ the counter $N_i^0(t; s)$ tracks the patient on $(s, s + t]$. For a subject eligible to receive the treatment at time s , (i.e. $I(T_i \geq s)$), if $T_i = s$, the counting process (2)

takes effect; if the treatment had not been available, process (3) takes effect. The subject is untreated on $(0, s)$ under either scenario to which (2) and (3) pertain.

Finally, we define a 0/1 process for being observed to receive treatment,

$$N_i^T(t) = \int_0^t I(C_i > u) dI(T_i \leq u). \quad (4)$$

2.2. Proposed Models

As described above, the goal of this method is to compare the post-treatment recurrent event mean to the corresponding event mean under no treatment. We denote the mean of (2) by

$$\mu_i^1(t; s) = E \left[\int_0^t N_i^1(du; s) | T_i = s, \mathbf{H}_i(s) \right], \quad (5)$$

where $\mathbf{H}_i(s) = \{Z_i^*(u), N_i(u), I(T_i > u), I(C_i > u); 0 \leq u < s\}$ represents the observed pre-treatment history for subject i on $[0, s)$.

Similarly, in the absence of treatment, the mean of (3) can be written as

$$\mu_i^0(t; s) = E \left[\int_0^t N_i^0(du; s) | \mathbf{H}_i(s), T_i > u \right]. \quad (6)$$

Note, both models are partly conditional [11, 12, 13] in the sense that they condition on the history up until time s as opposed to $s + t$. We do not model either (5) or (6) directly, instead, our model of interest is given by

$$\mu_i^1(t; s) = \mu_i^0(t; s) \exp\{\beta_\star\}, \quad (7)$$

which can equivalently be expressed in terms of a rate function by

$$\mu_i^1(dt; s) = \mu_i^0(dt; s) \exp\{\beta_\star\}. \quad (8)$$

In this model $\mu_i^0(t; s)$, the treatment-free mean number of events is scaled up or down by $\exp\{\beta_\star\}$ if subject i received the experimental treatment at time s . The mean number of post-treatment events is then compared to the mean number of treatment-free events after time s . It is conceivable that the treatment effect could depend on time since treatment, t , or time of treatment, s , and this model can be extended to accommodate a time-dependent β_\star in the form $\beta_\star(t; \cdot)$, $\beta_\star(\cdot; s)$ or $\beta_\star(t; s)$. Note that these different time-dependent forms of β_\star could be any parametric function of time such as linear or log-linear; time could also be categorized to examine the functional form of the time-dependent effect.

Since we cannot observe a patients' pre-treatment experience once treatment is initiated, a patient treated at time s will be compared to similar patients who did not start treatment at follow-up time s but were eligible to do so. Similar to Schaubel et al. [10], we use the concept that each treatment time initiates an "experiment", in which the recipient of the treatment is compared to 'similar' treatment-eligible candidates. Note that 'similar', in this context, refers to current status (i.e., at time s) and history on $[0, s)$. Eligibility for the comparison is defined as

$$e_i(s) = I(T_i = s) + I(T_i > s)$$

i.e., at time s , patient i either received the treatment or remained untreated.

Our method of estimating β_\star from (8) involves a stratified analysis. Each treated patient generates a stratum, which will include the index patient as well as similar treatment-eligible patients. Here we define similar as both treatment eligible at s , $e_i(s)$, and similar with regard to accumulated covariate and recurrent event history on $(0, s]$, $\mathbf{H}_i(s)$. In order to quantify each subject's history, we use a prognostic score [14] based on the pre-treatment event rate, modeled using a time-dependent proportional rates model,

$$d\mu_i^0(t) = E [dN_i^*(t) | \mathbf{H}_i(t), T_i > t] = \exp\{\alpha_0^T \mathbf{Z}_i(t)\} d\mu_0(t), \quad (9)$$

where the covariate $\mathbf{Z}_i(t)$ is chosen to capture the pertinent components of the history, $E[dN_i^*(t) | \mathbf{H}_i(t), T_i > t] = E[dN_i^*(t) | \mathbf{Z}_i(t), T_i > t]$. Model (9) resembles the marginal Lin et al. [5] model, but is more accurately interpreted as the conditional Andersen-Gill [1] model, due to the explicit dependence on the prior event history, a property avoided by Lin et al. [5]. The regression parameter α_0 from (9) can be computed by solving the unweighted Cox [15] score equation. Due to the dependence on internal covariates [9], elements of α_0 are difficult to interpret. However, the purpose of this model

is matching similar subjects on $[0, s)$, not interpretation. Note that, although we have chosen a multiplicative model, other models (e.g., an additive model with time-varying effects ([2])) may provide better fit and are reasonable options. Note also that alternatives to prognostic matching exist, and will be discussed in Section 5.

The purpose of the prognostic score is to match patients that have similar pre-treatment event rates, the rationale being that previous event rate is the most important predictor of the current event rate. Unlike a propensity score, which uses the treatment event rate to match subjects with similar probabilities of being treated, the prognostic score aims to compare the effect of treatment on the event rate among subjects that were on the same trajectory with respect to their pre-treatment event rate. The use of prognostic scores in conjunction with, or as a alternative to, propensity scores has been considered in several reports [16, 17, 18, 19] and will be discussed later. Once the prognostic scores have been estimated, caliper matching is used to assign untreated control subjects to a subject receiving treatment at time s . Caliper matching requires that the prognostic scores of matched subjects be within a certain radius of the prognostic score of the index subject. Appropriate selection of the caliper involves balancing the need for homogeneity within-stratum with the need to have an adequate number of matches for each index subject. The discrepancy between prognostic scores for experimental subject j and control subject i can be quantified through the subject-pair specific rate ratio,

$$\psi_{i,j}(s) = \frac{d\mu_i^0(s)}{d\mu_j^0(s)} = \exp\{\alpha_0^T [\mathbf{Z}_i(s) - \mathbf{Z}_j(s)]\}.$$

Subject i is ‘similar’ on $[0, s)$ to subject j if $|\log \psi_{ij}(s)| \leq \epsilon$, where $\epsilon > 0$ is a pre-determined constant.

Combining the eligibility indicators and prognostic scores, patient i is included in the stratum generated by patient j if $m_{ij}(s) = 1$, where

$$m_{ij}(s) = e_i(s)I(T_i > s)e_j(s)I(T_j = s)I(|\log \hat{\psi}_{ij}(s)| \leq \epsilon),$$

with $\hat{\psi}_{ij}(s) = \exp\{\hat{\alpha}_0^T [\mathbf{Z}_i(s) - \mathbf{Z}_j(s)]\}$. In order to account for the residual difference between patients i and j , we propose to adjust for $\log \hat{\psi}_{ij}(s)$ in the final model. Incorporating the eligibility indicator and the prognostic score distance, the final fitted model for the event mean for stratum j is then

$$\mu_{ij}^*(t; s) = m_{ij}(s)\mu_i^0(t; s) \exp\{\beta_\star I(T_i = s) + \beta_\psi \log \hat{\psi}_{ij}(s)\}. \quad (10)$$

In (10), j is the stratum (generated by patient j through $T_j = s$) and i is the patient within stratum. The model governs the treated patient through the indicator $I(T_i = s)$, which equals 1 if $i = j$. The vector of parameters to be estimated and the corresponding covariates are given by

$$\beta_{\star\psi} = \begin{bmatrix} \beta_\star \\ \beta_\psi \end{bmatrix} \quad \mathbf{Z}_i^*(s) = \begin{bmatrix} I(T_i = s) \\ \log \hat{\psi}_{ij}(s) \end{bmatrix}, \quad (11)$$

such that model (10) can be re-written as $\mu_{ij}^*(t; s) = m_{ij}(s)\mu_i^0(t; s) \exp\{\beta_{\star\psi}^T \mathbf{Z}_i^*(s)\}$.

Subjects matched to the treated subject enter the experiment without receiving any treatment, but could subsequently receive treatment. If a matched subject receives treatment after time s they are censored from all experiments in which they serve as controls and begin their own experiment as the index subject. This generally results in dependent censoring since, although treatment can be considered random given $\mathbf{H}_i(s+t)$, the model for $\mu_{ij}^*(t; s)$ from (10) only conditions on $\mathbf{H}_i(s)$, the pre-treatment history up to time s . While this could be addressed through Inverse Probability of Censoring Weighting (IPCW) [20, 21, 27], in this article we consider treatments that are relatively rare, with rates small enough such that bias due to dependent censoring is negligible. Section 3 will investigate through simulation treatment rates at which dependent censoring needs to be addressed.

2.3. Parameter Estimation

In order to estimate β_\star we define the pertinent risk set indicator for stratum j ,

$$Y_{ij}(t; s) = m_{ij}(s)I(C_i > s+t)\{I(T_i = s) + I(T_i > s+t)\}$$

If, given $\mathbf{H}_i(s)$ matched subjects are randomly assigned to treatment after time s , the process

$$m_{ij}(s) \int_0^{\tau-s} M_{ij}(du; s), \quad (12)$$

where $m_{ij}(s)$ is the matching indicator described above, $M_{ij}(du; s) = Y_{ij}(u; s)\{N_i(du; s) - \mu_{ij}(du; s)\}$, and τ is chosen to satisfy $P(C_i \geq \tau) > 0$ and often set to $\max\{C_1, \dots, C_n\}$, would have mean zero. As mentioned above, bias due to

censoring of subsequently treated controls is expected to be minimal in the setting of rare treatment, so we assume the condition above holds.

Aggregating across subjects for the experiment occurring at time s produces the set of zero mean processes,

$$\sum_{i=1}^n m_{ij}(s) \int_0^t M_{ij}(du; s) \tag{13}$$

$$\sum_{i=1}^n m_{ij}(s) \int_0^t \mathbf{Z}_i^*(s) M_{ij}(du; s). \tag{14}$$

We reorganize this system to solve implicitly for the baseline mean, $\mu_0^0(u; s)$ in (13), then substitute into (14). Then, aggregating across all experiments yields the final estimating function for $\beta_{*\psi}$,

$$U(\beta) = \sum_{j=1}^n \sum_{i=1}^n \int_0^\tau m_{ij}(s) \int_0^{\tau-s} \{\mathbf{Z}_i^*(s) - \bar{\mathbf{Z}}_*(u; s)\} N_i(du; s) dN_j^T(s), \tag{15}$$

where

$$\bar{\mathbf{Z}}_*(u; s) = \frac{\sum_{\ell=1}^n m_{\ell j}(s) \mathbf{Z}_\ell^*(s) \exp\{\beta_{*\psi}^T \mathbf{Z}_\ell^*(s)\}}{\sum_{\ell=1}^n m_{\ell j}(s) \exp\{\beta_{*\psi}^T \mathbf{Z}_\ell^*(s)\}}.$$

Since $U(\beta)$ from (15) behaves asymptotically like a zero-mean estimating function, the solution to $U(\beta) = \mathbf{0}$, denoted by $\hat{\beta}_{*\psi}$, should yield a consistent estimator of $\beta_{*\psi}$.

2.4. Asymptotic Properties and Inference

We propose a variation of a method proposed by Lin et al [5] to construct confidence bands for the mean function of the proportional means model. The following regularity conditions are imposed:

- (a) $[N_i(t), C_i(t), N_i^T(t), N_i^D(t), \mathbf{Z}_i(t), \gamma_i]$ are independent and identically distributed.
- (b) $P(X_i \geq \tau) > 0$ for all i .
- (c) $N_i(\tau) < \infty$ for all i .
- (d) $E[I(T_i \leq \tau)] > 0$ for all i .
- (e) $\mathbf{Z}_i(t)$ is of bounded variation.

Under these conditions, it can be shown that

$$n^{1/2}(\hat{\beta} - \beta) = \mathbf{A}^{-1}(\beta) n^{-1/2} \sum_{i=1}^n \mathbf{U}_i(\beta) + o_p(1), \tag{16}$$

where

$$\mathbf{U}_i(\beta) = \int_0^\tau m_i(s) \int_0^{\tau-s} \{\mathbf{Z}_i^*(s) - \bar{\mathbf{z}}_*(u; s)\} dM_i(s) dF^T(s),$$

$$\mathbf{A}(\beta) = E \left[\int_0^\tau m_i(s) \int_0^{\tau-s} \{\mathbf{Z}_i^*(s) - \bar{\mathbf{z}}_*(u; s)\}^{\otimes 2} \exp\{\beta^T \mathbf{Z}_i^*(s)\} dR_i^0(u; s) dF^T(s) \right],$$

$m_i(s)$ is an indicator taking value 1 when patient i is matched in terms of prognostic score to the index patient treated at time s and $F^T(s)$ is defined as above.

We approximate the distribution of (16) using the sandwich form $\hat{\mathbf{A}}^{-1}(\hat{\beta}) n^{-1/2} \sum_{i=1}^n \hat{\mathbf{U}}_i(\hat{\beta}) \hat{\mathbf{U}}_i(\hat{\beta})^T$.

3. Simulation Study

3.1. Simulations of Proposed Method

We conducted simulations to demonstrate the properties of the proposed estimator in moderate sized samples. For each scenario we simulated 1000 subjects 500 times. In addition to the observed experience, the counterfactual, treatment-free experience was generated for each subject in order to determine target values for β_* , which, given the complex data structure, were difficult to pre-specify. Target values of β_* were determined by simulating observed and counterfactual

Table 1. Results from simulations of n=1000 subjects with varying baseline pre-treatment event rates

Scenario	$d\mu_0^1$	β_*	Estimate	Bias	ESE	ASE	CP
1	0.3	-0.725	-0.731	-0.006	0.179	0.167	0.94
2	0.4	-0.553	-0.558	-0.005	0.171	0.165	0.95
3	0.5	-0.407	-0.410	-0.003	0.171	0.162	0.95
4	0.75	-0.111	-0.115	-0.004	0.172	0.157	0.94
5	1	0.113	0.111	-0.002	0.166	0.154	0.94
6	1.5	0.454	0.449	-0.005	0.166	0.149	0.93
7	1.75	0.592	0.585	-0.007	0.163	0.147	0.93
8	2	0.714	0.702	-0.012	0.166	0.145	0.92

experience for 100,000 subjects 10 times and comparing each treated subjects' observed experience to their counterfactual experience.

Independent adjustment covariates Z_{i1} and Z_{i2} were generated to follow a Bernoulli(0.5) distribution. A longitudinal covariate Z_3 was generated to follow a $U(0, 1)$ distribution at time 0 updated at every time unit with increment $Bernoulli(0.5) \times U(0, 1)$. Starting at time 0, we generated a new event process based on the updated value of Z_3 using a frailty model with rate parameter $Q_i d\mu_0 \exp\{\alpha_1 Z_{i1} + \alpha_2 Z_{i2} + \alpha_3 Z_{i3}(t)\}$, where $d\mu_0 = 10$, $\alpha_1 = 0.3$, $\alpha_2 = -0.1$, $\alpha_3 = 0.1$, where Q_i was distributed Gamma with mean 1 and variance 0.5. Q_i was used to induce additional correlation between recurrent events and was capped at 2, the 90th percentile. Treatment times, T_i^S , were then generated to follow the hazard $\lambda_0^T \exp\{\delta_1 Z_{i1} + \delta_2 Z_{i2} + \delta_3 \log(N_i(t^-) + 1) + \delta_4 Z_{i3}(t)\}$, where $\delta_1 = -0.2$, $\delta_2 = 0.2$, $\delta_3 = 0.2$, $\delta_4 = 0.4$. The recurrent event times post-treatment were generated from rate parameter $Q_i d\mu_0^T \exp\{\phi_1 Z_{i1} + \phi_2 Z_{i2} + \phi_3 \log(N_i(T_i) + 1) + \phi_4 Z_{i3}(t)\}$, where $\phi_1 = 0.3$, $\phi_2 = -0.1$, $\phi_3 = 0.2$, $\phi_4 = 0.1$.

Once the data were generated, prognostic scores representing pre-treatment event trajectories were obtained from the model $d\mu_i^0(t) = \exp\{\alpha_{01} Z_{i1} + \alpha_{02} Z_{i2} + \alpha_{03} N_i^0(t^-) + \alpha_{04} Z_{i3}(t)\}$. Subjects were matched if $|\log \hat{\psi}_{ij}| \leq 0.025$.

Parameters used in the simulation studies are as follows. For the pre-treatment event rates, we set $d\mu_0 = 3$, $\alpha_1 = 0.3$, and $\alpha_2 = -0.1$ for the observed process and $d\mu_0 = 6$, $\alpha_1 = 0.3$, and $\alpha_2 = -0.1$ for the unobserved process. For the treatment hazard, $\lambda_0^T = 0.01$, $\delta_1 = -0.2$, $\delta_2 = 0.1$, $\delta_3 = 0.2$, and $\delta_4 = 0.5$. Finally, for the post-treatment event rate, $d\mu_0^T$ was given values of 1, 1.5, 2, 3, 4, and 5, $\phi_1 = 0.3$, $\phi_2 = -0.1$, and $\phi_3 = 0.2$. This resulted in values of β_* ranging from -0.725 to 0.714. In the simulated data 8% of the sample received treatment and the mean number of events was 20.7.

Results from the simulations are show in Table 1. Absolute bias ranged from 0.002 to 0.012. Empirical and asymptotic standard errors were similar, and coverage probabilities ranged from 0.92 to 0.95, close to the target level of 0.95.

3.2. Investigation of dependent censoring

Recall that subjects are censored from strata in which they serve as controls if they subsequently receive treatment. Since treatment depends on, among other things, the event history, this will result in dependent censoring in cases where treatment is not rare. We used simulation to explore the point at which more common treatments result in substantial bias. To do this we used a similar set up to that of the previous section except that $d\mu_0^T$ was set at 1 and λ_0^T took on values of 0.01, 0.05, 0.15, and 0.3. This resulted in percentages of subjects treated ranging from 8% to 64%. Results of these simulations are shown in Table 2.

As shown in Figure 1, increasing the proportion of subjects treated increases bias and decreases coverage probability. At approximately one-third of subjects treated bias is at 0.03, and this almost triples to 0.086 when 64% of subjects are treated, with increased bias resulting in lower coverage. Bias and coverage are similar from 8-24% treated, likely due to the smaller absolute number of treated subjects. However, as the percentage of subjects treated increases, censoring of subjects serving as controls due to subsequent treatment results in dependent censoring that begins to have a non-negligible effect on the treatment effect. Given this trajectory it is important to keep in mind that methods such as IPCW aimed at correcting dependent censoring are not necessary only if the proportion of treated subjects relatively rare; however, once the proportion of treated subjects becomes more prevalent weighting is necessary to correct bias.

Table 2. Results from simulations with varying percent treated

Scenario	% Treated	β_*	Estimate	Bias	ESE	ASE	CP
1	8%	0.714	0.706	-0.008	0.151	0.145	0.95
2	16%	0.693	0.681	-0.012	0.115	0.103	0.93
3	24%	0.674	0.651	-0.023	0.095	0.086	0.92
4	33%	0.650	0.620	-0.030	0.079	0.075	0.92
5	41%	0.626	0.585	-0.041	0.076	0.068	0.88
6	46%	0.603	0.553	-0.050	0.071	0.064	0.84
7	54%	0.578	0.518	-0.060	0.065	0.061	0.81
8	64%	0.536	0.450	-0.086	0.059	0.054	0.72

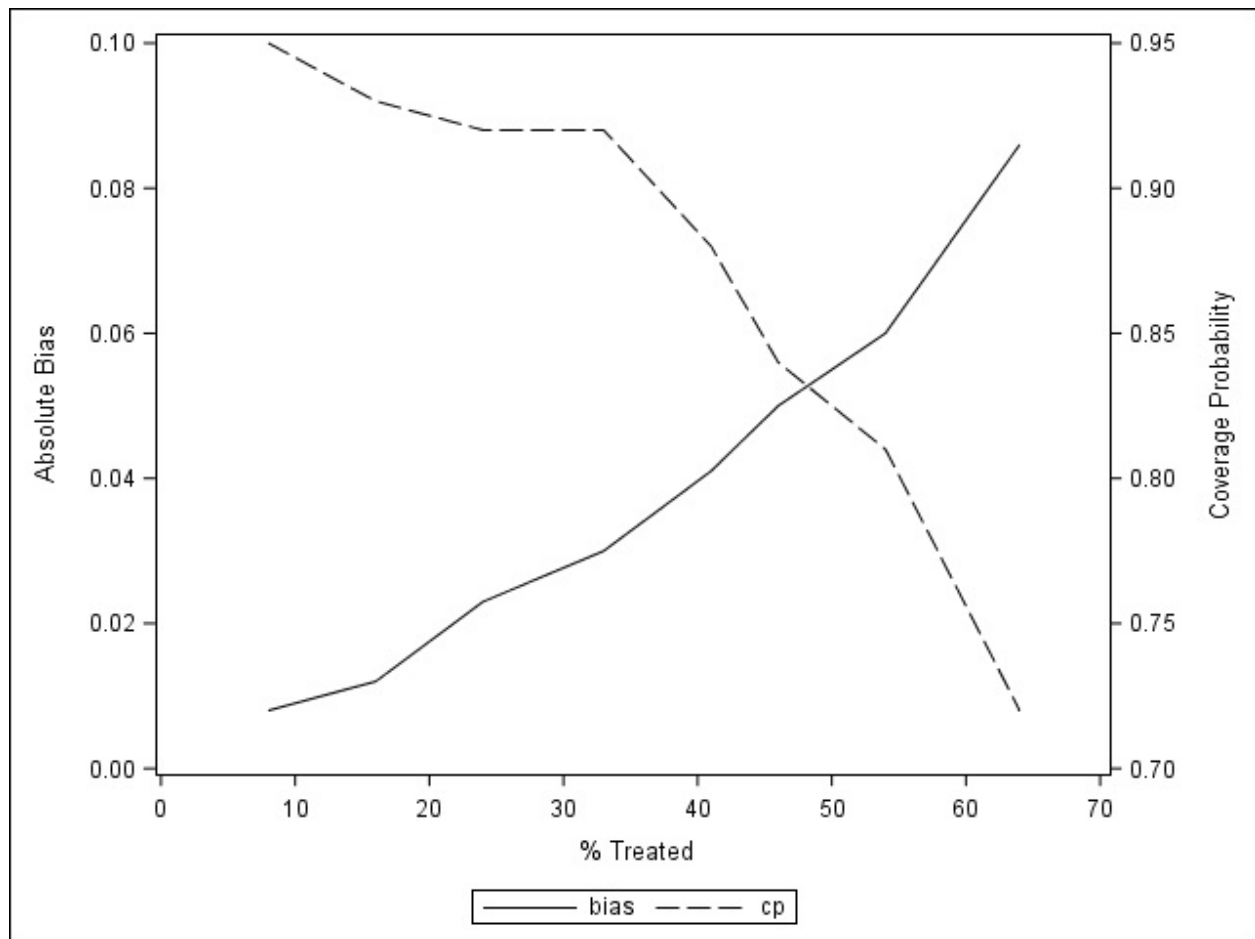


Figure 1. Bias and coverage probability with increasing percentage treated: The left y-axis shows bias and corresponds to the solid line. The right y-axis shows coverage probability and corresponds to the dashed line.

4. Application to Liver Transplantation

Development of End Stage Renal Disease (ESRD) post-liver transplant leads to increased patient morbidity and mortality, and places increased burden on health care resources. We will use the proposed method to evaluate effect of ESRD development post-liver transplant on the number of days hospitalized in the Adult-to-Adult Living Donor Liver Transplantation Cohort Study (A2ALL). In this setting the “treatment” of interest is development of ESRD, defined as initiation of dialysis or kidney transplant post-liver transplant. As mentioned previously, the proposed method is

Table 3. Prognostic model: Pre-treatment rate of days hospitalized

Parameter	Rate Ratio	95% Confidence Interval	p-value
Recipient age at Transplant (ref=65+)			
18–40	1.00	(0.95, 1.06)	0.873
40–50	0.77	(0.73, 0.81)	<.001
50–55	0.79	(0.75, 0.83)	<.001
55–60	0.78	(0.74, 0.82)	<.001
60–65	0.87	(0.82, 0.92)	<.001
Recipient diagnosis: HCV	1.07	(1.04, 1.10)	<.001
African–American (ref=all others)	0.75	(0.71, 0.79)	<.001
Diabetes	0.89	(0.86, 0.92)	<.001
Ln(creatinine) (time–dependent)	1.31	(1.27, 1.35)	<.001
Ln(bilirubin) (time–dependent)	1.09	(1.07, 1.10)	<.001
Ln(albumin) (time–dependent)	0.36	(0.34, 0.38)	<.001
Donor age (ref=70+)			
<18	0.64	(0.55, 0.74)	<.001
18–40	0.76	(0.67, 0.87)	<.001
40–50	0.78	(0.69, 0.89)	<.001
50–60	1.02	(0.90, 1.16)	0.722
60–70	0.72	(0.63, 0.83)	<.001
DCD (ref=non–DCD)	1.33	(1.21, 1.47)	<.001
Regional (ref=Local)	1.32	(1.25, 1.40)	<.001
National (ref=Local)	1.41	(1.32, 1.51)	<.001
Split Liver	1.18	(1.08, 1.29)	<.001
Living Donor (ref=Deceased Donor)	0.96	(0.88, 1.05)	0.397
Hospitalization History (per day)	1.02	(1.02, 1.02)	<.001

generalizable to time-dependent state changes such as development of post-LT ESRD. In this setting we utilize time-dependent markers of kidney function such as creatinine to estimate the hospitalization trajectory from the time of transplant to the development of ESRD, and use these to match with patients on similar trajectories that do not develop ESRD. Comparing the rate of days hospitalized for a patient that develops post-LT ESRD compared to the rate that would have been observed had the patient not developed ESRD is a critical component to the estimation of the costs of post-LT care.

A2ALL is a multi-center NIH-funded consortium composed of 12 North American transplant centers. Potential living donor liver transplant (LDLT) recipients transplanted between January 1, 1998 and January 31, 2014 were enrolled. Retrospective and prospective data collection included post-transplant vital status and laboratory information as well as hospitalization admission and discharge information. Data were supplemented from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donors, wait-listed candidates, and transplant recipients in the United States; these data are submitted by the members of OPTN and have been described elsewhere. The Health Resources and Services Administration (US Department of Health and Human Services) provides oversight for the activities of the OPTN and SRTR contractors.

There were 55 ESRD events out of 1447 transplanted patients in A2ALL. Median post-transplant follow-up time was 5 years, and the average number of days hospitalized per patient year was 14.9 for non-ESRD patients and 37.2 for ESRD patients (median days 2.3 and 5.1, respectively). Hospitalization admissions that occurred after discharge from the transplant hospitalization but before onset of ESRD were used to build the prognostic model, which was adjusted for the event history as well as other transplant and post-transplant time-dependent predictors. Results from the prognostic model are shown in Table 3. Each additional day of hospitalization history was associated with a 2% increase in the rate of future hospitalization days ($p < 0.001$).

Using prognostic scores derived from the model in Table 3, the distribution of prognostic score distance from index patient with and without matching is shown in Figure 2. Prior to matching on prognostic score the range of distance between the index subject and matched controls spans from -7.2 to 5.4, with 98% of matched controls within the interval [-1.5, 3.5] from the index subject. When the 55 patients that developed ESRD post-transplant were matched to patients that had not yet developed ESRD based on prognostic score, with all control subjects within ± 0.02 , the distribution of score distance is much tighter around zero. The matching resulted in a median of 14 matches, with 6 (11%) of patients

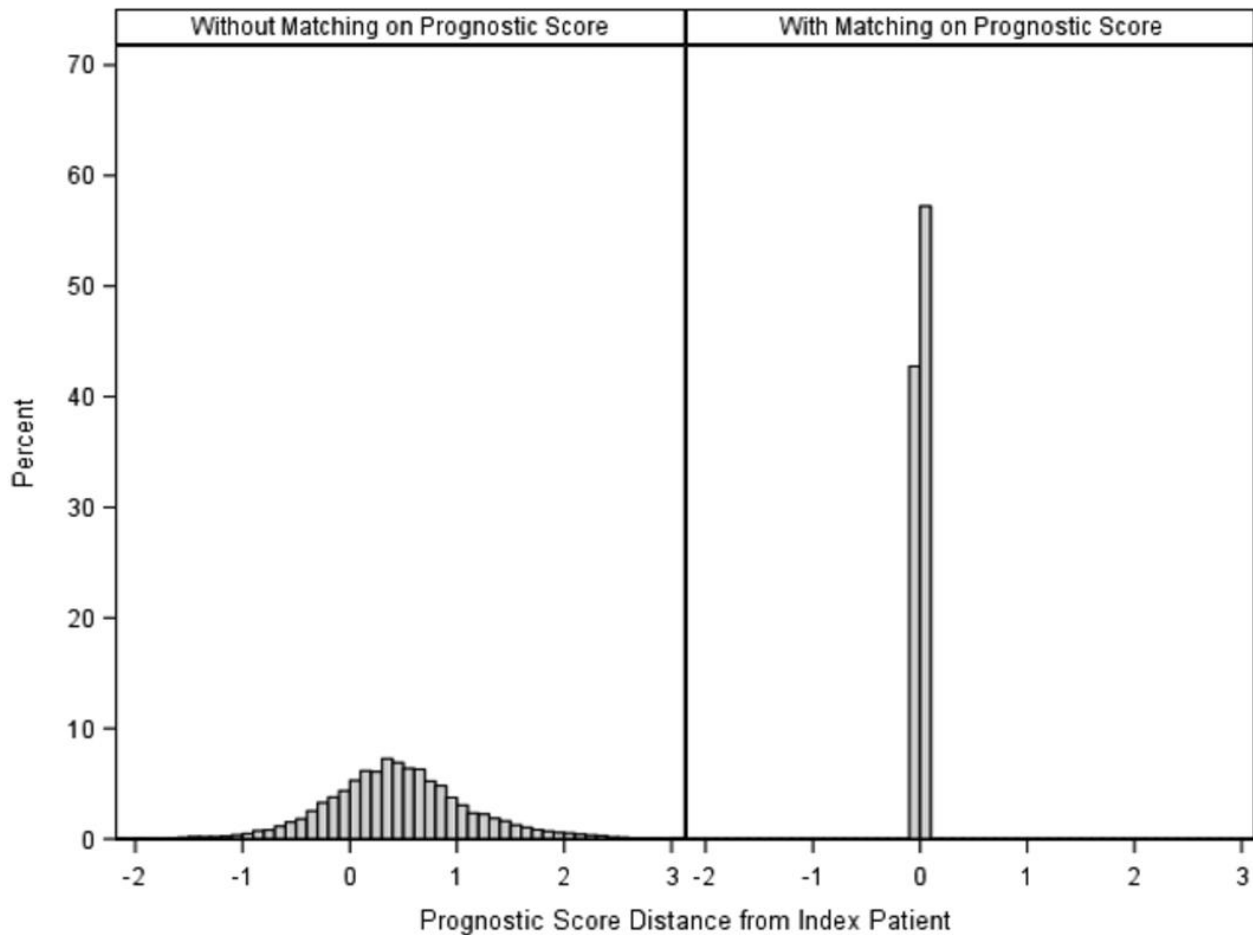


Figure 2. Distribution of prognostic score distance from index patient with and without matching: The left panel shows the difference in prognostic scores between the index patient and all potential controls. The right panel shows the difference in prognostic scores between the index patient and matched controls, i.e. controls with prognostic scores within the caliper.

that developed ESRD being excluded due to lack of matches.

The proposed method was then used to fit a stratified model to determine the effect of ESRD on the rate of days hospitalized (Model I) using model (10). The following traditional time-dependent proportional rates models were also fitted where ESRD status was treated as a time-dependent predictor adjusted for the same predictors in the prognostic model (Table 2).

$$\mu_i(t) = \mu_i^0(t) \exp\{\theta_{II}I(T_i \leq t) + \beta^T \mathbf{Z}_i(t)\}. \quad (17)$$

$$\mu_i(t) = \mu_i^0(t) \exp\{\theta_{III}I(T_i \leq t) + \beta^T \mathbf{Z}_i\}. \quad (18)$$

In the first model, model (17), additional time-dependent predictors thought to be associated with the progression to ESRD were included, such as lab values and hospitalization history, while in second model, (18), only baseline, i.e. at transplant, values of these predictors were used. The results from all three models are shown in Table 4. In Model I, which uses the proposed method, patients that develop ESRD have a rate of days hospitalized that is 2.9 times higher than patients that have not yet developed ESRD. By contrast, results from Model II give a rate that is only 1.4 times higher for patients that have developed ESRD, but this comparison is to patients without ESRD that have the same lab values and hospitalization history at time t . By contrast, adjusting only for baseline values of factors associated with the development of ESRD (Model III) estimates that patients that develop ESRD have a days hospitalized rate 3.2 times higher than patients that do not have ESRD at time t and were similar at transplant. This comparison demonstrates how use of the proposed method balances the opposing biases of over- and under-adjustment. As a sensitivity analysis we also tested interactions

Table 4. Rate of days hospitalized post-liver transplant: comparison of proposed method with traditional baseline and time-dependent Cox models

Model	Equation	Parameter	RR	95% CI	p-value
I: Proposed Method	(10)	β_*	2.90	(1.69, 4.97)	<0.001
II: Time-dependent Adjustment Covariates	(17)	θ_{II}	1.44	(1.35, 1.53)	<0.001
III: Baseline Adjustment Covariates	(18)	θ_{III}	3.17	(3.01, 3.35)	<0.001

RR=Rate Ratio; CI = Confidence Interval

with time since development of ESRD and time of development of ESRD, but no significant variations in the effect ESRD development were found (both rate ratios ≈ 1 , $p = 0.23$ and $p = 0.68$, respectively).

5. Discussion

In this report we lay out a two-stage method for estimating the effect of time-dependent treatments on recurrent events using an extension of the method of sequential stratification. The method proposed is partly conditional in the sense that information up until treatment time, s , is used in the prognostic model, but the final model for $\mu_{ij}^*(t; s)$ does not condition on covariates after s . A purely conditional model (e.g. Anderson and Gill [1]), which would include covariate information on $[s, s + t)$, would tend to dampen the effect of treatment because it would require comparison subjects to have the same history at the time of treatment. A marginal analysis such as that of Lin et al [5], on the other hand, would exaggerate the effect of treatment because since treatment depends on the history, subjects that receive treatment at time s may differ from those that do not. Our method ensures subjects are similar up to s through conditioning, and is marginal thereafter, so that the post-treatment comparison averages over the treatment-free experience of the matched controls.

Note that there are two practical considerations to consider when applying this method. First, in order to fit the prognostic model there needs to be sufficient observed time-at-risk prior to treatment, and a sufficient number of events prior to treatment. In the repeated measures setting this is generally not a concern since a given subject can have multiple events, however, if the recurrent event is rare and many subjects never experience the recurrent event or don't experience their first recurrent event prior to treatment this must be considered. Second, there must be enough strata for sequential stratification to be appropriately applied. In the proposed method each treated subject forms their own stratum and is matched to patients with similar prognostic scores at the time of treatment. In theory this implies that there does not need to be large variation in treatment times, however the method was developed for the setting in which subjects receive treatment at varying times during follow-up, and in the setting in which all patients are treated at the same time (i.e. at or close to baseline) alternative, traditional methods apply.

The biases of the approaches featuring either baseline-covariate-only or fully time-dependent covariate adjustment is demonstrated in Section 5 in the context of the development of post-liver transplant ESRD in the A2ALL study. While all three methods produced significant results, the fully time-dependent model, which included a time-dependent indicator for development of ESRD as well as time-dependent lab values and previous number of days hospitalized (each associated with renal failure progression) underestimated the effect of ESRD on the rate of days hospitalized by almost half. In contrast, the baseline-covariate-only approach over-estimated the effect of ESRD development. While in this application the two traditional models produced bias in opposite directions, this may not always be the case, and the direction of the bias may depend on the magnitude and direction of association between the time-dependent covariate(s) and the probability of treatment and the recurrent event rate, the degree of model misspecification, and the presence of non-linear link functions. In addition, the comparison groups for these models are not constructed in a way that gives the desired interpretation; i.e. a comparison of the event rate in the time period following treatment in the presence and absence of treatment.

Note that the outcome chosen in the application was days hospitalized instead of hospital admissions. Analyses of hospital admissions often ignore the fact that patients are not at risk for hospitalization during the period in which they are

in the hospital. This can be accounted for by removing the duration of hospitalization from the risk set, however, this step is often ignored. Modeling days hospitalized instead of hospital admissions automatically removes this potential for error, however, in some situations, hospital admissions may be a more relevant outcome.

As mentioned previously, it follows from the dependence of treatment initiation on the event history that censoring of matched controls due to treatment would constitute dependent censoring. We have shown that when the treatment is rare, bias is not substantial, and therefore IPCW to correct for dependent censoring is not necessary. However, for more common treatments, bias will be induced and therefore some sort of weighting must be done to preserve the unbiased properties of β_x .

The proposed method makes use of the prognostic score in order to match as yet untreated patients into strata. Another viable alternative would be propensity matching [22], i.e., matching on the probability of receiving treatment. A time-dependent propensity score has been proposed by Lu [29], and could be used in this setting. Our goal, however, was to create a comparison group that mirrored the treatment-free experience of a subject treated at time s . It was therefore necessary to ensure that the event trajectories up until s were the same between treated and control subjects, a property that the propensity score does not preserve. Some combination of prognostic and propensity scores could also be used. Another potential option would be to hard match on the event counter at the time of treatment initiation, given that it is the strongest predictor of the pre-treatment event rate. However, we previously investigated this approach and found that prognostic score matching performed better ([27]).

One limitation of the proposed method is that it does not account for terminating events that halt all further occurrences of the recurrent event. The terminal event is often correlated with the recurrent event, and when treated as a censoring mechanism, can result in dependent censoring. Numerous methods for the simultaneous modeling of recurrent and terminal events are currently available. Marginal models of recurrent events while subjects are alive have been described by Cook and Lawless [23] and Ghosh and Lin [24]. Joint modeling methods where the recurrent and terminal event processes are linked through a subject-level random effect have been proposed by Liu et al. [25], Ye et al. [26], Kalbfleisch et al. [28], and others. Development of these methods is currently underway.

Acknowledgement

This research was supported in part by National Institutes of Health Grant R01-DK070869. The authors wish to thank the Reviewer and Associate Editor for their thoughtful suggestions, which resulted in a considerably improved manuscript. They also thank the Scientific and Data Coordinating Center project area team at Arbor Research Collaborative for Health, and the investigators, coordinators, and other staff that participate in the A2ALL Study. Data reported in this publication were provided by the Adult-to-Adult Living Donor Liver Transplantation Cohort Study (A2ALL) study and research was supported by the National Institute of Diabetes & Digestive & Kidney Diseases cooperative agreement U01-DK62498. Some data reported here were supplied by the Minneapolis Medical Research Foundation as the contractor for the Scientific Registry of Transplant Recipients. The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the Scientific Registry of Transplant Recipients or the US Government.

References

1. Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. *Annals of Statistics* 1982; **10**:129-143.
2. Aalen, O.O. A linear regression model for the analysis of life times. *Statistics in Medicine* 1989; **12**: 1569–1588.
3. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *Journal of the American Statistical Association* 1993; **88**:811-820.
4. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995; **37**: 158-168.
5. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society* 2000; Series B **62**: 711-730.
6. Cook RJ, Lawless JF, Lakhal-Chaieba 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. *Journal of the American Statistical Association* 2009; **104**: 60-75.
7. Schaubel DE, Zhang M. Estimating treatment effects on the marginal recurrent event mean in the presence of a terminating event. *Lifetime Data Analysis* 2010; **16**: 451-477.
8. Chen Q, Zeng D, Ibrahim JG, Akacha M, Schmidli H. Estimating time-varying effects for overdispersed recurrent events data with treatment switching. *Biometrika* 2013; **100**: 339- 354.
9. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. Wiley: New York (2nd Edition), 2002.
10. Schaubel DE, Wolfe RA, Sima CS, Merion RM. Estimating the effect of a time-dependent treatment by levels of an internal time-dependent covariate. *Journal of the American Statistical Association* 2009; **104**: 49-59.

11. Pepe MS, Couper D. Modeling partly conditional means with longitudinal data. *Journal of the American Statistical Association* 1997; **92**: 991-998.
12. Zheng YY, Heagerty PJ. Partly conditional survival models for longitudinal data. *Biometrics* 2005; **61**: 379-391.
13. Gong Q, Schaubel DE. Partly conditional estimation of the effect of a time-dependent factor in the presence of dependent censoring. *Biometrics* 2013; **69**: 338-347.
14. Hansen BB. The prognostic analogue of the propensity score. *Biometrika* 2008; **95**: 481-488.
15. Cox DR. Partial likelihood. *Biometrika* 1975; **62**: 269-276.
16. Rubin DB, Thomas N. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association* 2000; **95**: 573-585.
17. Stuart EA, Lee BK, Leacy FP. Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology* 2013; **66**: S84-S90.
18. Leacy FP, Stuart EA. On the joint use of propensity and prognostic scores in estimation of the average treatment effect on the treated: a simulation study. *Statistics in Medicine* 2014; **33**: 3488-3508.
19. Li Y, Schaubel DE, He K. Matching methods for obtaining survival functions to estimate the effect of a time-dependent treatment. *Statistics in Biosciences* 2014; **6**: 105-126.
20. Robins JM, Finkelstein DM. Correcting for non-compliance and dependent censoring in an AIDS clinical trial with Inverse Probability of Censoring Weighted (IPCW) logrank tests. *Biometrics* 2000; **56**: 779-788.
21. Miloslavsky M, Keles S, Van Der Laan MJ, Butler S. Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *Journal of the Royal Statistical Society* 2004; Series B **66**: 239-257.
22. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41-55.
23. Cook RJ, Lawless JF. Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* 1997; **16**: 911-924.
24. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics* 2000; **56**: 554-562.
25. Liu L, Wolfe RA, Huang X. Shared frailty model for recurrent events and a terminal event. *Biometrics* 2004; **60**: 747-756.
26. Ye Y, Kalbfleisch JD, Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* 2007; **63**: 78-87.
27. Smith AR, Schaubel DE. Time-dependent prognostic score matching for recurrent event analysis to evaluate a treatment assigned during follow-up. *Biometrics* 2015; **71**: 950-959.
28. Kalbfleisch JD, Schaubel DE, Ye Y, Gong Q. An estimating function approach to the analysis of recurrent and terminal events. *Biometrics* 2013; **69**: 366-374.
29. Lu B. Propensity score matching with time-dependent covariates. *Biometrics* 2005; **61**: 721-728.