Published in final edited form as:

Stat Med. 2017 October 15; 36(23): 3654–3669. doi:10.1002/sim.7377.

How to Control for Unmeasured Confounding in an Observational Time-To-Event Study With Exposure Incidence Information: The Treatment Choice Cox Model

James F. Troendle^{1,*}, Eric S. Leifer¹, Zhiwei Zhang², Song Yang¹, and Heather Tewes³

¹Office of Biostatistics Research, Division of Cardiovascular Sciences of the National Heart, Lung, and Blood Institute, NIH/DHHS, Bld RLM2 Room 9196, Bethesda, MD 20892, USA

²Department of Statistics, University of California Riverside, 1430 Olmsted Hall, 900 University Ave., Riverside, CA 92521, USA

³Nebraska Department of Health and Human Services, P. O. Box 95026, Lincoln, NE 68509-5026, USA

Abstract

In an observational study of the effect of a treatment on a time-to-event outcome, a major problem is accounting for confounding due to unknown or unmeasured factors. We propose including covariates in a Cox model that can partially account for an unknown time-independent frailty that is related to starting or stopping treatment as well as the outcome of interest. These covariates capture the times at which treatment is started or stopped and so are called *treatment choice* (TC) covariates. Three such models are developed. First, an interval TC model which assumes a very general form for the respective hazard functions of starting treatment, stopping treatment, and the outcome of interest. Second, a parametric TC model that assumes that the log hazard functions for starting treatment, stopping treatment, and the outcome event include frailty as an additive term. Finally, a hybrid TC model that combines attributes from the parametric and interval TC models. As compared to an ordinary Cox model, the TC models are shown to substantially reduce the bias of the estimated hazard ratio for treatment when data are simulated from a realistic Cox model with residual confounding due to the unobserved frailty. The simulations also indicate that the bias decreases or levels off as the sample size increases. A TC model is illustrated by analyzing the Women's Health Initiative observational study of hormone replacement for post-menopausal women.

Keywords

Bias; Frailty; Hazard; Matching; Propensity

*Correspondence: jt3t@nih.gov.

Supplementary Materials

Supplementary materials referenced in Section 4 and 5 are available with the online version of this article.

1 Introduction

For a nonrandomized observational study, consider the problem of studying the effect of a treatment on a time-to-event outcome where some of the study subjects start treatment sometime during the period of observation while the rest of the subjects never start treatment. The hazard function for the outcome event is modeled using a Cox model and the focus of the analysis is on estimation of the hazard ratio of being on treatment as compared to not being on treatment. A major problem with such an analysis is confounding due to unobserved factors that are related to both the outcome event and the initiation and termination of treatment.

Methods to analyze observational data for treatment effects are becoming more important as the availability of electronic health records data increases. Many analyses of health records are being published using statistical methods to adjust for confounding [1–3]. In the situation where treatment is initiated, if at all, at the start of the observation period, propensity score methods can be useful for taking into account factors that influence treatment selection [4]. In the situation where subjects start on study prior to initiating treatment, there are several ways one might use propensity scores. One propensity method matches subjects on the probability of initiating treatment during the observed follow-up period. However, this method may sacrifice sample size by discarding non-matched subjects or allow for suboptimal matching. Another drawback is that this type of analysis is based solely on the total number of events and person years of follow-up in the study and therefore does not fully utilize the information inherent in the event times. Another way to utilize the propensity score is to include the estimated propensity score as a covariate in a Cox model for the outcome event. As is also the case for propensity matching, the regression adjustment approach cannot account for unknown factors that are associated with both starting treatment and occurrence of the outcome.

A more general approach that does try to utilize the outcome event information more fully is the marginal structural Cox proportional hazard model (MSCPHM) that has been proposed to obtain the causal effects of possibly multiple treatments [5]. The major advantage of an MSCPHM is that it allows estimation of the causal effects in the presence of a very general time dependent treatment initiation process when there is a measured time dependent intermediary that is also a confounder for treatment. In this case, it has been shown [6] that ordinary regression models are unable to obtain the true causal effect of treatment. This is because if one conditions on the intermediary by including it in the model, then the treatment effect is attenuated while if one doesn't condition on the intermediary the analysis is confounded. Importantly, a MSCPHM assumes that one has measured all of the intermediaries that are also confounders. Moreover, a MSCPHM accounts for exposures by modeling at each discrete exposure time to obtain propensities of exposure, each of which must be correctly specified for the overall method to be valid. In contrast, the approach proposed here informs the outcome Cox model directly through the use of time varying covariates.

In any observational study, the effect of treatment may be biased by confounding due to selection of who gets treatment as well as the time at which treatment is started or stopped.

We refer to a subject's decisions to start and stop treatment and the corresponding times at which those decisions are made as the subject's *treatment choice* (TC) process. When the unmeasured confounders are related to the outcome event and the TC process, our method can potentially account for much of the treatment effect bias due to the unmeasured confounders. We first model the TC process in an empirical manner through a collection of time varying *interval TC covariates*, each of which corresponds to a specific short interval of time and as a collection can change values on a fine grid of times. The bias correction improves as the sample size increases and the grid can be made finer and longer. When the sample size is moderate, we combine the interval TC covariates as linear combinations into two different *cumulative TC covariates* that are appropriate under certain parametric assumptions on the confounding. We also discuss hybrid methods that use both the interval and cumulative TC covariates.

The paper proceeds as follows. We give notation and define the problem in Section 2. We next describe the proposed TC methods for the time-to-event analyses in Section 3. Simulations to explore the statistical properties of the TC methods are given in Section 4. We illustrate the TC methods by applying them to data from the Women's Health Initiative Observational Study in Section 5. Finally, our recommendations are given in Section 6.

2 Notation and Problem

We assume that there are n subjects in the study and each subject is initially untreated. In practice subjects may start the study on treatment, but we will restrict our analysis to the subset of subjects that start out untreated. Some of these n subjects will be observed to start treatment at some point during the study (treated subjects) and the remaining are not observed to start treatment (untreated subjects). We assume that the time-to-start treatment, time-to-stop treatment, and the time-to-event data are generated from a competing hazards model with frailty, where a subject's frailty will represent unmeasured or unknown (to the data analyst) covariates that are perhaps available to those in charge of deciding whether to treat the subject or not. For example a treating clinician may decide to initiate treatment for a subject for reasons that are not included in the data base. Our data generating model is specified by the cause specific hazards for four "events": initiation of treatment, termination of treatment, censoring, and outcome. For simplicity we assume that treatment is binary (treated or untreated). For each of the four types of events, we assume that the corresponding true hazard function is a function of treatment, age, and frailty. Age is the only measured confounder in the data generating model, but in general any measured confounder could be included. A subject's frailty is assumed to be a latent random effect. Our data generating model for subject l ($l = 1, 2, \dots, n$) at follow-up time t is as follows:

$$F_l = \alpha_f + Z_l, \quad (1)$$

$$\theta_l(t) = \{1 - IT_l(t)\} * \exp \{ \alpha_{tr} + \beta_{1tr} * A_l(t) + \beta_{2tr} * F_l \}, \quad (2)$$

$$S_I(t) = IT_I(t) * \exp \{ \alpha_s + \beta_{1s} * A_I(t) + \beta_{2s} * F_I \}, \quad (3)$$

$$G_I(t) = \exp(\alpha_g), \quad (4)$$

$$H_I(t) = \exp \{ \alpha_h + \beta_{1h} * A_I(t) + \beta_{2h} * F_I + \beta_{3h} * IT_I(t) \}, \quad (5)$$

where F is the frailty, A is the subject's age, θ is the hazard for initiation of treatment, S is the hazard for stopping treatment, G is the hazard for random censoring, H is the hazard for the outcome, Z is a random frailty component, IT is the indicator of being on treatment, and the α 's and β 's are fixed parameters. Note that the frailty term, F , is static, i.e., independent of time, for each subject. One might suppose that as a subject ages certain risk factors would change resulting in an overall dynamic frailty. Even though frailty as specified above is static for each subject, our models statistically capture changing risk through age, which is assumed dynamic. We use a static frailty to capture those features of individuals which lead them to have different inherent risks independent of their age.

The hazard ratio for treatment is given in (5) as β_{3h} . The basic problem is to draw inference on β_{3h} , i.e. to determine if treatment is helpful or harmful. Note however in (5) that F_I is unobserved.

The observed data for subject I consists of a follow-up time, U_I (the subject was under observation for $t \in (0, U_I]$), along with r_I event times, $V_{I1}, \dots, V_{Ir_I}$ and type of each event, $W_{I1}, \dots, W_{Ir_I}$ ($W = 0$ if censoring, $W = 1$ if outcome, $W = 2$ if treatment initiation, $W = 3$ if treatment stop), along with other measured risk factors such as baseline age. It is assumed that a subject's follow-up ends once a censoring or outcome event occurs.

3 Treatment Choice Models

Our goal is to estimate the treatment effect β_{3h} in (5) using the observable treatment initiation and stopping times in place of the unobservable frailty F_I . To motivate TC models, consider the simpler situation in which once a subject starts treatment, the subject stays on treatment for the remainder of his/her follow-up. That is, there are only treatment initiation times with no treatment stopping times.

For the purpose of our motivation, suppose frailer subjects (i.e., subjects with larger F_I values) are more likely to start treatment earlier than less frail subjects and are also more likely to have an outcome event sooner. In our data generating model, this corresponds to $\beta_{2tr} > 0$ in (2) and $\beta_{2h} > 0$ in (5). As presently explained, a reasonable *analysis Cox model* would model subject I 's log hazard function as

$$\log\{\lambda_l(t)\} = \log\{\lambda_0(t)\} + \beta_1 * BA_l + \beta_2 * TC_l(t) + \beta_3 * IT_l(t) \quad (6)$$

where $\lambda_0(t)$ is the nuisance hazard function, BA_l is baseline age, $TC_l(t)$ is a time-varying covariate that captures the treatment initiation time, and $IT_l(t)$ is the on-treatment indicator. The usual Cox treatment effect estimate $\hat{\beta}_3$ for β_3 in (6) will be our estimate for β_{3h} in (5). A particular choice for $TC_l(t)$, called the *cumulative constant TC covariate*, is defined below in Section 3.1. It has the property that it is largest for subjects who start treatment earlier and smallest for subjects who never start treatment. That is, for three arbitrary subjects under observation at time t with subject 1 starting treatment at time $t_1 < t$, subject 2 starting treatment at time $t_2 < t$ with $t_2 > t_1$, and subject 3 not having started treatment by t , we have $TC_1(t) > TC_2(t) > TC_3(t)$. Since frailer subjects are more likely to start treatment sooner by virtue of $\beta_{2tr} > 0$ in (2), frailer subjects will tend to have larger $TC_l(t)$ values. Also, since frailer subjects are more likely to have an outcome event sooner by virtue of $\beta_{2h} > 0$ in (5), we would expect that the estimate $\hat{\beta}_2$ for (6) to partially capture the relationship between frailty and outcome by having a positive value. This intuition suggests that the estimate $\hat{\beta}_3$ for (6) will be closer to the parameter of interest, β_{3h} , than if we did not include the $TC_l(t)$ term in (6). Similar intuition applies to the other three sign possibilities for $(\beta_{2tr}, \beta_{2h})$: (i) $\beta_{2tr} > 0$ and $\beta_{2h} < 0$; (ii) $\beta_{2tr} < 0$ and $\beta_{2h} > 0$; (iii) $\beta_{2tr} < 0$ and $\beta_{2h} < 0$.

It is important to note in all of these cases that F_l is a baseline, albeit unobservable, quantity in the data generating models (2) and (5). In contrast, $TC_l(t)$ is a post-baseline observable quantity that is stochastically related to F_l through (2). Thus, there is no expectation that (6) will provide an asymptotically unbiased estimate $\hat{\beta}_{3h}$ of β_{3h} in (5) since $TC_l(t)$ is an imperfect surrogate for F_l . However, the numerical studies described in Section 4 confirm the intuition discussed above that including TC terms in the analysis Cox model give better estimates of β_{3h} than if TC is ignored.

Our motivation for including $TC_l(t)$ as a single additive term in the analysis model (6) was because F_l is an additive term in (2) and (5). Since there could be a more complicated relationship between frailty and treatment initiation and outcome, it might be important to allow more flexibility in modeling the TC process including having more than a single TC covariate. To begin our general construction of TC covariates, we fix a time I_{max} during the study's follow-up period such that the TC covariates will only be constructed to capture TC information on $(0, I_{max}]$. We next choose a fixed time increment such that the intervals $(0,], (, 2], \dots, (I_{max} - , I_{max}]$ partition $(0, I_{max}]$. These intervals are called *TC intervals* since the TC covariates will model the subjects' treatment choices as step functions on these intervals.

In Section 3.1 we describe TC covariates in the case where as in the above motivating discussion, we assume that once a subject starts treatment, that subject never stops treatment. In Section 3.2, we will consider the more general case where some subjects stop treatment.

3.1 Without Stopping of Treatment

The first type of TC covariate we describe is called an *interval* TC covariate, or ITC covariate. For the current scenario in which a subject does not stop treatment once he/she has started, an ITC covariate records whether a subject starts in a particular TC interval $(t_1, t_1 + \Delta]$. We denote the corresponding ITC covariate as $TC_{(t_1, t_1 + \Delta]}(t)$.

To define $TC_{(t_1, t_1 + \Delta]}(t)$, we categorize the subjects into two groups: the subjects who have not started treatment by t_1 and have follow-up past $t_1 + \Delta$ comprise the *at risk to start treatment* in $(t_1, t_1 + \Delta]$ group. All other subjects comprise the *not at risk to start treatment* group. We define $TC_{(t_1, t_1 + \Delta]}(t) = 0$ for all t for all subjects not at risk to start treatment in $(t_1, t_1 + \Delta]$. For the at risk to start subjects, we define $TC_{(t_1, t_1 + \Delta]}(t) = 0$ for all $t < t_1 + \Delta$. However, $TC_{(t_1, t_1 + \Delta]}(t)$ will make a single jump at $t_1 + \Delta$ for those at risk to start, as presently described. Let n_1 be the number of subjects at risk to start in $(t_1, t_1 + \Delta]$, so $n - n_1$ subjects were not at risk to start. Suppose n_{11} subjects at risk to start treatment during the interval did start and n_{10} subjects did not start treatment in the interval so $n_1 = n_{11} + n_{10}$. Let $f = n_{11}/n_1$. For the n_{11} subjects at risk to start treatment who did start, we define $TC_{(t_1, t_1 + \Delta]}(t) = 1 - f$ for all $t \geq t_1 + \Delta$. For the n_{10} subjects at risk to start treatment who did not start, we define $TC_{(t_1, t_1 + \Delta]}(t) = -f$ for all $t \geq t_1 + \Delta$. This definition implies that on $(t_1 + \Delta, t_1 + 2\Delta]$ which is the first TC interval impacted by the TC information from $(t_1, t_1 + \Delta]$, $TC_{(t_1, t_1 + \Delta]}(t)$ has average value equal to zero among the subjects at risk to start treatment in $(t_1, t_1 + \Delta]$ as well as average value zero for the subjects not at risk to start treatment in $(t_1, t_1 + \Delta]$. Thus, in the analysis Cox model the covariate $TC_{(t_1, t_1 + \Delta]}(t)$ will only distinguish between subjects who started in $(t_1, t_1 + \Delta]$ from those who could have started in $(t_1, t_1 + \Delta]$, but did not. It will not imply anything about the difference in the average log hazard for the outcome between those at risk to start in $(t_1, t_1 + \Delta]$ and those not at risk to start. In this sense, $TC_{(t_1, t_1 + \Delta]}(t)$ purely captures the treatment choice in $(t_1, t_1 + \Delta]$ of subjects who could choose whether or not to start treatment in $(t_1, t_1 + \Delta]$.

We note that without this special centering of $TC_{(t_1, t_1 + \Delta]}(t)$ the resulting models would perform poorly compared to the centered version we propose here, and would not have the important characteristic (shown via simulation in Section 4) that the absolute bias is always less than that of the ordinary Cox model. Simulation results of models using un-centered TC covariates are included in the Supplementary Materials.

To gain some more understanding about ITC covariates, we now give an example of them for a small, artificial data set. For the sake of illustration, the example data consist of six subjects with follow-up histories shown in Figure 1. While follow-up continues to year 4, we take $I_{max} = 3$ years and $\Delta = 1$ year so the TC intervals are $(0, 1]$, $(1, 2]$, and $(2, 3]$. To simplify our discussion, treatment start times, outcomes, and censoring events only occur at integer year times; we compute $TC_{(0,1]}(t)$, $TC_{(1,2]}(t)$, and $TC_{(2,3]}(t)$.

Notice that all six subjects start the study untreated, but subject #1 reaches an endpoint at 1 year and therefore does not have follow-up past 1 year. Thus there are $n_1 = 5$ subjects who are at risk to start treatment in $(0, 1]$. Of those five subjects, only subject #3 starts treatment in $(0, 1]$. Thus $f = 1/5$ for this interval and the values for $TC_{(0,1]}(t)$ are as given in Table 1 as well as the values for the usual on-treatment indicator $IT(t)$. To compute $TC_{(1,2]}(t)$, we see in

Figure 1 that subjects #2, #4, and #6 are untreated at the start of (1,2] with follow-up past 2 years. Thus there are $n_1 = 3$ subjects who are at risk to start treatment in (1, 2]. Of these three subjects, only subject #2 starts treatment in (1, 2]. Thus $f = 1/3$ for this interval and the values for $TC_{(1,2]}(t)$ are as given Table 1. To compute $TC_{(2,3]}(t)$, we see in Figure 1 that subjects #4 and #6 are untreated at the start of (2,3] with follow-up past 3 years. Thus there are $n_1 = 2$ subjects who are at risk to start treatment in (2, 3]. Of those two subjects, only subject #4 starts treatment in (2, 3]. Thus $f = 1/2$ for (2, 3] and the values for $TC_{(2,3]}(t)$ are as given Table 1. Notice that the average value is zero on $(t_1 + 1, t_1 + 2]$ for each non-zero ITC covariate $TC_{(t_1, t_1+1]}$ for the at-risk to start subjects on $(t_1, t_1 + 1]$ as well as for the not at risk to start subjects. This ensures that the ITC covariate makes no implication about outcome risk based on whether a subject is at risk to start treatment or not. If such an implication were made, the ITC covariate would be capturing information beyond newly determined treatment choice. However, at future times $t > t_1 + 2$, the at risk to start in $(t_1, t_1 + 1]$ group is reduced due to outcome events and censoring so the average value of $TC_{(t_1, t_1+1]}(t)$ is no longer zero for those subjects that remain in the analysis Cox model risk set. Subject I 's log hazard function for the ITC analysis Cox model for Table 1 is

$$\log\{\lambda_I(t)\} = \log\{\lambda_0(t)\} + \beta_1 * BA_I + \beta_3 * IT_I(t) + \gamma_1 * TC_{(0,1],I}(t) + \gamma_2 * TC_{(1,2],I}(t) + \gamma_3 * TC_{(2,3],I}(t) \quad (7)$$

In general, we can compute an ITC covariate for any TC interval. However, to maintain the stability of the treatment effect estimate, the number of ITC covariates that are included in the analysis model is limited by the number of observed outcomes. If a particular ITC covariate is included in the analysis model, we will refer to the corresponding TC interval as an *ITC interval*. The recommendations for a TC interval $(t_1, t_1 +]$ to be an ITC interval will be discussed in greater detail in Section 4 and will involve the number of expected outcome events required from the n_{11} subjects who start treatment in $(t_1, t_1 +]$ as well as the number of expected outcome events required from the n_{10} subjects who were at risk to start treatment in $(t_1, t_1 +]$, but did not start.

When we are limited by sample size, there are other types of TC covariates that can be fit. As discussed in the motivation at the beginning of Section 3, the data generating model (2) and (5) assumes that the unobserved frailty F_I is an additive term in the log hazard functions for starting treatment $\log[\theta_I(t)]$ and outcome $\log[H_I(t)]$. As described in the motivation, it could be reasonable for such a situation to include a single TC covariate in the analysis Cox model which is parametrized so that it is largest for subjects who start treatment earlier and smallest for subjects who never start treatment. A single TC covariate that has this property can be obtained by cumulatively summing $TC_{(t_1, t_1+]}(t)$ over the TC intervals. We will denote this covariate as $CTC_{0I}(t)$ and call it the *cumulative constant TC covariate*. The resulting analysis Cox model is

$$\log\{\lambda_I(t)\} = \log\{\lambda_0(t)\} + \beta_1 * BA_I + \beta_3 * IT_I(t) + \gamma_0 * CTC_{0I}(t) \quad (8)$$

Table 1 can be checked to see that $CTC_{0l}(t)$ has the desired ordering property for the subjects on study at time t : (a) $CTC_{0l}(t)$ is larger for subjects who have started treatment by time t than for subjects who have not started treatment by time t ; (b) Among those subjects who have started treatment by time t , $CTC_{0l}(t)$ is larger for subjects who started treatment earlier. We formally state this ordering property in the following lemma whose proof is given in the Appendix:

Lemma 1—Ordering property of the cumulative constant covariate. Suppose we have the TC intervals $I_1 = (0, j_1]$, $I_2 = (j_1, j_2]$, ..., $I_K = (j_{K-1}, j_K]$. Suppose for 3 subjects on study that subject 1 starts treatment before subject 2 while subject 3 never starts treatment. Suppose subject 1 starts treatment in I_{j_1} and subject 2 starts treatment in I_{j_2} . Then for subjects 1, 2, and 3's respective cumulative constant covariates $CTC_{01}(t)$, $CTC_{02}(t)$, and $CTC_{03}(t)$, we have

$$CTC_{01}(t) \geq CTC_{02}(t) \geq CTC_{03}(t) \quad \forall t \quad (9)$$

and

$$CTC_{02}(t) > CTC_{03}(t) \quad \forall t > j_2 \Delta. \quad (10)$$

Moreover, if $j_1 < j_2$, then

$$CTC_{01}(t) > CTC_{02}(t) \quad \forall t > j_1 \Delta. \quad (11)$$

We can add more flexibility to (8) by including a TC covariate that is a linear function of the ITC covariates. This is accomplished by first multiplying each ITC covariate by the right hand endpoint of the corresponding TC interval and then taking the cumulative sum of the resulting covariates. We denote the summed covariate as $CTC_{1l}(t)$ and call it the *cumulative linear TC covariate*. Section 4's numerical studies include analysis models which we call *Parametric Treatment Choice* (PTC) models whose only TC covariates are the cumulative constant and linear TC covariates:

$$\log\{\lambda_l(t)\} = \log\{\lambda_0(t)\} + \beta_1 * BA_l + \beta_3 * IT_l(t) + \gamma_0 * CTC_{0l}(t) + \gamma_1 * CTC_{1l}(t) \quad (12)$$

Section 4's studies show that PTC models reduce bias more than ITC models for small or moderate sample sizes when the data are generated from (1)–(5). For the example dataset depicted in Figure 1 with TC intervals (0, 1], (1, 2], and (2, 3], the cumulative constant and cumulative linear TC covariates are given in Table 1. As seen in Table 1, the ITC covariates are step functions with at most one jump. In contrast, the cumulative constant and linear covariates are step functions that can have several jumps.

We can have additional modeling flexibility beyond (12) if we create a model that includes ITC covariates for intervals for which there is sufficient sample size to do so. Additionally,

we include a cumulative constant covariate which sums the $TC_{(t_1, t_1 + j]}(t)$ terms over intervals not corresponding to ITC covariates. We call the resulting model a *Hybrid Treatment Choice* (HTC) model. Collectively we will refer to the PTC, HTC, and ITC models as TC models.

For an example of an HTC model we again consider the example dataset depicted in Figure 1. Recall that the TC intervals are $(0, 1]$, $(1, 2]$, and $(2, 3]$ since we set $I_{max} = 3$. Thus, an HTC model will be determined by deciding which of the potential ITC covariates to include. Suppose we only include the ITC covariate $TC_{(0,1]}(t)$. Then the corresponding HTC model will include $TC_{(0,1]}(t)$ as well as a cumulative constant covariate corresponding to the intervals $(1, 2]$ and $(2, 3]$, that is, $TC_{(1,2]}(t) + TC_{(2,3]}(t)$.

Using the notation from the beginning of Section 3 to describe the general case, suppose the TC intervals are $(0, j_1]$, $(j_1, j_2]$, ..., $(I_{max} - j_{K-1}, I_{max}]$. Suppose K of these intervals have sufficient sample size for an ITC covariate to be included in the analysis Cox model. Then the ITC model for subject l 's log hazard function is:

$$\log \lambda_l(t) = \log \lambda_0(t) + \beta_1 * BA_l + \beta_3 * IT_l(t) + \gamma_1 * TC_{1l}(t) + \dots + \gamma_K * TC_{Kl}(t) \quad (13)$$

where BA_l is subject l 's baseline age, $IT_l(t)$ is the on-treatment indicator, and $TC_{il}(t)$ is the ITC covariate for the i^{th} interval and γ_i is the corresponding parameter for $i = 1, \dots, K$. The estimate of the log hazard ratio for treatment is $\hat{\beta}_3$, the fitted value of β_3 from (13). Tests for H_0 are based on $TS_{ITC} = \hat{\beta}_3 / \widehat{SE}(\hat{\beta}_3)$ which has approximately a standard normal distribution under H_0 . A two-sided 5% level test of H_0 rejects if $|TS_{ITC}| > 1.96$.

The HTC model adds a single cumulative TC covariate to the ITC model for subject l 's log hazard function:

$$\log \{\lambda_l(t)\} = \log \{\lambda_0(t)\} + \beta_1 * BA_l + \beta_3 * IT_l(t) + \gamma_1 * TC_{1l}(t) + \dots + \gamma_K * TC_{Kl}(t) + \gamma_{K+1} * CTC_{0l}(t) \quad (14)$$

where $CTC_{0l}(t)$ is the cumulative constant TC covariate for the TC intervals not included in the ITC model and γ_{K+1} is the corresponding parameter. The estimate of the log hazard ratio for treatment is $\hat{\beta}_3$, the fitted value of β_3 from (14). Tests for H_0 are based on

$TS_{HTC} = \hat{\beta}_3 / \widehat{SE}(\hat{\beta}_3)$ which has approximately a standard normal distribution under H_0 . A two-sided 5% level test of H_0 rejects if $|TS_{HTC}| > 1.96$.

The PTC model has a cumulative constant TC covariate and a cumulative linear TC covariate, but no ITC covariates, for subject l 's log hazard function:

$$\log \{\lambda_l(t)\} = \log \{\lambda_0(t)\} + \beta_1 * BA_l + \beta_3 * IT_l(t) + \gamma_0 * CTC_{0l}(t) + \gamma_1 * CTC_{1l}(t) \quad (15)$$

where $CTC_0(t)$ is the cumulative constant TC covariate and γ_0 is its corresponding parameter, and $CTC_1(t)$ is the cumulative linear TC covariate and γ_1 is its corresponding parameter. The estimate of the log hazard ratio for treatment is $\hat{\beta}_3$, the fitted value of β_3 from (15). Tests for H_0 are based on $TS_{PTC} = \hat{\beta}_3 / \widehat{SE}(\hat{\beta}_3)$ which has approximately a standard normal distribution under H_0 . A two-sided 5% level test of H_0 rejects if $|TS_{PTC}| > 1.96$.

3.2 Allowing for Stopping of Treatment

When stopping of treatment is allowed the definition of at risk to start treatment is slightly adjusted to only include those subjects untreated at time t_1 , never previously treated, and with follow-up past $t_1 + \Delta$. This slight modification of the definition does not change the models described in Section 3.1 since in that section no subjects stop treatment, but reflects a decision to model only first starts and first stops of treatment. Stopping treatment is handled in the analysis model in a very similar manner as starting treatment. We define the ITC covariate corresponding to stopping treatment in the time interval $(t_1, t_1 + \Delta]$ in an entirely analogous manner to the ITC covariate for starting treatment. However for stopping treatment, we define the *at risk to stop* set for $(t_1, t_1 + \Delta]$ as the subjects who are treated at time t_1 , never previously stopped, and with follow-up past $t_1 + \Delta$. Similar definitions to those given in Section 3.1 are then used to distinguish between those at risk to stop who do stop from those at risk to stop who do not stop. In the Supplementary Materials, we give an example of starting and stopping ITC covariates for a small, artificial data set.

One question about capturing the effects of starting and stopping treatment is whether one wants to model the second and subsequent starts or stops of treatment for a given subject if such events do occur. Although one could try to capture such events, we do not do so in this paper. Thus our definition of at risk to start (stop) only includes those without a prior start (stop). If one did want to include second and subsequent starts or stops, one would have to decide whether to treat such events differently than first starts and stops. Without introducing additional assumptions, one would then need to treat subsequent starts and stops as different than first starts and stops. The simulations presented in Section 4 show that one does not need to include second and subsequent starts or stops to substantially correct for the unknown frailty. Thus the gain from incorporating second and subsequent starts or stops would likely be small, and the sample sizes required for reliable model fitting would be very large.

Once the starting and stopping TC covariates are constructed, the ITC method then fits a Cox model exactly as given in (13) above, except that now the K ITC covariates include those for starting as well as those for stopping. Similarly the HTC method adds two cumulative constant TC covariates (one for starting and one for stopping) to the ITC model. Finally, the PTC model has two cumulative constant TC covariates plus two cumulative linear TC covariates (one of each type for starting and for stopping).

4 Simulations

We use simulated data to study the statistical properties of the proposed TC methods and to compare them to the ordinary Cox model and propensity methods which are described in

Supplementary Materials. Data are simulated from the competing hazard model given in (1)–(5) from Section 2. Maximum length of follow-up is set at $U_{max} = 10$ years. Baseline age in years is generated randomly from the Uniform(50,70) distribution. The random component Z_I of the frailty term $F_I = \alpha_f + Z_I$ is generated as Uniform(0, ψ). Using an approach developed for competing risk models [7], we determine sequentially (using a time increment of $t_{inc} = .001$ years) if an event has occurred (of any type) in the current follow-up time interval from the survival distribution

$$P(V > t + t_{inc} | V > t) = \exp \left[- \int_t^{t+t_{inc}} \{ \theta(u) + S(u) + G(u) + H(u) \} du \right].$$

Given an event at time $V < U_{max}$, we determine the cause according to the conditional probabilities

$$P(\text{treatment initiation}) = \frac{\theta(V)}{\theta(V) + S(V) + G(V) + H(V)},$$

$$P(\text{treatment stopping}) = \frac{S(V)}{\theta(V) + S(V) + G(V) + H(V)},$$

$$P(\text{censoring}) = \frac{G(V)}{\theta(V) + S(V) + G(V) + H(V)},$$

$$P(\text{outcome event}) = \frac{H(V)}{\theta(V) + S(V) + G(V) + H(V)}.$$

This continues until a censoring, an outcome event, or until we reach time U_{max} , a point at which administrative censoring is assumed to occur.

To maintain the stability of the treatment effect estimate, the upper bound I_{max} for the TC intervals was chosen as the largest value such that there were at least 50 expected outcome events among subjects who started treatment after I_{max} . The expected number of events from subjects who start after I_{max} is calculated by multiplying the number of such subjects by the overall event proportion in the entire dataset ($\#events/n$). We used a fixed width of $\Delta = 0.1$ years for the TC intervals.

Notice that as I_{max} becomes larger, the treatment initiation model (2) implies that subjects who have not started treatment by I_{max} become more homogeneous with respect to frailty. Indeed, suppose $\beta_{2tr} > 0$ in (2). Then those subjects with larger frailty values are more likely to start treatment earlier. Thus, subjects who are still untreated at I_{max} will likely have lower frailty values. An analogous argument applies when $\beta_{2tr} < 0$.

It is important to have homogeneity of frailty for subjects who have not started treatment by I_{max} since we cannot adjust the treatment effect for frailty using TC covariates for such subjects since TC covariates are not constructed for treatment initiation times after I_{max} . However, this becomes less of an issue if such subjects are relatively homogeneous with respect to frailty. This is because if the variability in frailty is small for such subjects, nearly all of the variability in the outcome model (5) among such subjects will be through the relevant measured covariates. Recall that for simplicity of our presentation, (5) assumes that age is the only relevant measured covariate for outcome. In the extreme, but unrealistic, case that all subjects who have not initiated treatment by I_{max} have the same frailty, then frailty is not confounded with treatment for those subjects.

Recall from Section 3.1 that the ITC intervals are those TC intervals for which the corresponding ITC covariate is included in the analysis model. Our numerical work suggests that to maintain stability of the treatment effect estimate, a TC interval can be an ITC interval if that interval has 50 expected events per cell of subjects. A particular TC interval $I_j = (j, (j+1)]$ has two cells of subjects: those subjects, at risk to start, who started treatment during I_j and those subjects, at risk to start, who did not start treatment during I_j . Recall from the beginning of Section 3.1 that those two cells of subjects comprise the at risk to start treatment in I_j group.

Table 2 shows the bias (estimated from 500 replications) for the estimation procedures for selected values of ψ , β_{2tr} , and β_{3tr} , $n = 1000$ subjects, and where a subject stays on treatment once he/she has started. Since the parameter ψ controls the contribution of the subject specific random component $Z_I \sim \text{Uniform}(0, \psi)$ of the frailty $F_I = \alpha_I + Z_I$, the case of $\psi = 0$ corresponds to the case of no confounding due to frailty. In this case, the ordinary Cox model provides an unbiased estimate of the treatment effect β_{3tr} . When $\psi = 0$, the propensity modeling method, ordinary Cox model and the HTC and PTC models (14)–(15) have little bias with the standard deviations of the simulation estimates varying from .006 to .010. In contrast and as expected, when $\psi \neq 0$ so there is residual confounding due to the unobserved frailties, none of the methods completely control for such confounding. However, the magnitude of bias varies dramatically between the methods. The propensity matching method is greatly biased in certain cases, including when $\psi = 0$, and is not recommended or included in further comparisons. The propensity model method is somewhat better than the matching version, but is still more biased than the PTC and HTC models. Note that the ITC model is not included in Table 2 since there were not enough events among the 1000 subjects to meet the criteria described in the previous paragraph for a TC interval to be an ITC interval.

Table 3 uses the same parameter settings as Table 2, but for sample sizes of 10000 and 20000, respectively. These sample sizes might be appropriate for an electronic medical record registry. One can see that there is very little bias in the TC estimation procedures whenever the unobserved frailty has no effect ($\psi = 0$). Moreover, when $\psi \neq 0$, the TC bias is smaller for $n = 10000$ than for $n = 1000$ in Table 2, and decreases or stays about the same as n increases from 10000 to 20000. Tables 2 and 3 use a fixed TC interval width of $\Delta = 0.1$ years. As shown in the Supplementary Materials, when we used a smaller interval width of $\Delta = 0.05$ years, further bias reduction was observed.

Four other points are worth making about Table 3's results when $n = 10000$, $\psi = 0.8$, and $\beta_{2tr} = \pm 2.0$. First, when $\beta_{2tr} = -2.0$ the ITC bias is almost as large as the ordinary Cox bias and is 100–200% larger than the ITC bias when $\beta_{2tr} = 2.0$. This is partly because when $\beta_{2tr} = -2.0$, fewer subjects start treatment than when $\beta_{2tr} = 2.0$. Consequently, typically only 12 ITC covariates are included in the ITC model corresponding to the intervals (0.0, 0.1]yrs, (0.1, 0.2]yrs, ..., (1.1, 1.2]yrs when $\beta_{2tr} = -2.0$. These ITC intervals contain approximately 57% of the treatment start times when $\beta_{2tr} = -2.0$. In contrast when $\beta_{2tr} = 2.0$, 15 ITC covariates corresponding to the intervals (0.0, 0.1]yrs, (0.1, 0.2]yrs, ..., (1.4, 1.5]yrs are typically included in the ITC model and these intervals contain approximately 87% of the treatment start times. The second point is that the ITC bias when $\beta_{2tr} = -2.0$ is about 60–100% larger than the HTC bias. This is because the HTC model includes the ITC covariates for (0.0, 0.1]yrs, (0.1, 0.2]yrs, ..., (1.1, 1.2]yrs as well as the cumulative constant covariate that typically corresponds to (1.2, I_{max}]yrs where $I_{max} = 4.9$ yrs. Thus the HTC model typically includes information from intervals containing approximately 97% of the treatment start times while the ITC model includes approximately 57% of such times in its informative intervals. The third point is that the PTC and HTC biases are quite comparable. This is because frailty is an additive term in the true treatment initiation log hazard (2) and outcome event log hazard (5) models. This supports the intuitive argument made at the beginning of Section 3 that when frailty is an additive term in the treatment initiation and outcome log hazard functions, the cumulative constant TC covariate included in the PTC model (12) and HTC model (14) should account for much of the unobserved frailty. The fourth point is that when $\beta_{2tr} = -2.0$, the standard deviation of the PTC and HTC treatment effect estimates are about 1.5 times as large as that of the ITC estimate. Again, this is because the ITC model typically fits TC covariates for the intervals (0.0, 0.1]yrs, (0.1, 0.2]yrs, ..., (1.1, 1.2]yrs while the PTC and HTC models fit TC covariates over (0.0, I_{max}]yrs where I_{max} is typically 4.9yrs. This is an example of a bias-variance trade-off.

To see how well an ITC model can come to an unbiased treatment effect estimate, we next consider in Table 4 a nearly ideal, albeit unrealistic, case for capturing TC information through an ITC model. For this case we modify the simulation settings in (4) and (5) so that the hazard is zero for censoring and the outcome event during the first year; then, after the first year the hazards are as specified according to those equations. Additionally, after the first year, we remove the frailty term from the hazard for starting treatment in (2) so that starting treatment after the first year of the study is uninformative for frailty. Thus, frailty only impacts when treatment is initiated during (0, 1]yrs. Frailty has no impact on treatment initiation after 1 year. However, frailty does have an effect on outcome after 1 year. Finally, we force 10 ITC covariates into the analysis model corresponding to the intervals (0.0, 0.1]yrs, (0.1, 0.2]yrs, ..., (0.9, 1.0]yrs.

Since the PTC model includes cumulative constant and linear TC covariates up to I_{max} which in these simulations is typically 5.0 years, it incorrectly takes into account when subjects start treatment after $t = 1$ year which is when starting treatment is not informative for frailty. Similar comments apply to the HTC model which includes the same ITC covariates as the ITC model on (0.0, 1.0]yrs and a cumulative constant TC covariate on (1.0, I_{max}] years. Table 4 shows the bias of the TC models (13)–(15) for several configurations of simulation parameters compared to the bias of the ordinary Cox model. Once again there is

very little bias in any of the models whenever frailty has no effect ($\psi = 0$). When $\psi = 0.8$ and $\beta_{2h} = 1.0$, the ITC model typically has smaller absolute bias than the PTC and HTC models since the PTC and HTC models incorrectly model treatment initiation on $(1, I_{max}]$ years. Nevertheless, all three TC models have less absolute bias than the ordinary Cox model. Some bias remains in the ITC model since while starting treatment is uninformative for frailty after the first year, frailty does affect outcome after the first year through the $U_{max} = 10$ years of follow-up. Moreover, the TC covariates are only stochastically, not deterministically, related to F_j through (2). Thus, even in this nearly ideal scenario, we still expect the ITC model to provide a slightly biased treatment effect estimate. When $\beta_{2h} = 0.5$, thus reducing frailty's impact on the outcome hazard, bias for all three TC models improves with the ITC model typically remaining the best.

We simulated data with both starting and stopping of treatment. The details of that analysis can be found in the Supplementary Materials, but we note here that the results are similar to those seen in Table 3. That is, for fixed TC interval width $= 0.1$ years, the bias decreases or stays about the same as the sample size increases from 10000 to 20000.

To see how well the TC models can come to unbiased treatment effect estimates with stopping present, we also considered an ideal case for capturing TC information from starting and stopping of treatment. The details of that analysis can be found in the Supplementary Materials, but we note here that the results are similar to those seen in Table 4.

5 Women's Health Initiative Observational Study

The Women's Health Initiative Observational Study (WHI-OS) was a companion study to the Women's Health Initiative Randomized Trials (WHI-RCT) sponsored by the National Heart, Lung, and Blood Institute of the National Institutes of Health [8]. The WHI-RCT consisted of a pair of trials differentiated by whether the woman had an intact uterus at enrollment or not. Women with a uterus were randomized to either 0.625 mg/day conjugated equine estrogen (CEE) plus 2.5 mg/day medroxyprogesterone acetate (MPA) or placebo. Women without a uterus were randomized to either 0.625 mg/day CEE or placebo. Time to coronary heart disease (CHD) was the primary efficacy outcome and time to breast cancer was the primary safety outcome for both trials. Other outcomes included time to stroke, colorectal cancer, and hip fracture. A total of 16,608 women with a uterus were enrolled in the CEE/MPA trial while 10,739 women without a uterus were enrolled in the CEE trial.

Women either ineligible for the clinical trial or unwilling to join were invited to join the observational study ($n=93,676$). Recruitment was from September 1993 through December 1998. Post-menopausal women aged 50 to 79 were eligible for the WHI-OS. Baseline information was obtained by measurement on a series of physical measurements and blood chemistry. Medication, medical history, lifestyle, and demographic information was ascertained by questionnaires. Follow-up for events and treatment exposure was collected yearly up to 8 years post study entry. A major limitation of the observational study is the lack of detailed information on the timing of exposure. The questionnaires asked if the participant used hormones during the interval between the last questionnaire and the current

questionnaire, but not when hormones were started. For a subject who reported no hormone use on the previous questionnaire and reported hormone use on the current questionnaire, we imputed the date for starting hormones as the midpoint of the dates of the current questionnaire and the previous questionnaire. Similar methods were used to impute dates of stopping hormone use.

Prentice [9] describes a series of analyses aimed at comparing results from the observational and randomized studies. In that paper Prentice lists seven factors that he found important in controlling for confounding for risk of CHD: age, ethnicity, body mass index, education, smoking, age at menopause, and a physical function construct. Each of the ordinary Cox and TC models applied here include these factors as covariates. Prentice also found other factors that were important in aligning the WHI-OS and WHI-RCT results: namely that the hazard ratio (HR) for hormone therapy depended on the time since initiation of treatment. Thus the proportional hazard assumption is suspect. Prentice jointly analyzes the WHI-OS and WHI-RCT using product terms to allow the HR to vary according to the time since initiation of treatment. Yang and Prentice [10] analyzed the WHI-OS and WHI-RCT data using a semi-parametric model which allows the treatment effect to be time dependent while keeping the proportional hazards model as a sub-model. Since we will not be analyzing the WHI-RCT directly, we do not attempt to model the HR in any way other than by assuming it is constant. Results from Yang and Prentice [10] indicate that the proportional hazard assumption is reasonable for the WHI-OS with CHD or stroke outcomes. With extensive post-hoc modeling, Prentice [9] is able to achieve fairly close agreement between the WHI-OS and WHI-RCT for CHD and breast cancer outcomes, but unable to achieve comparable results for stroke, colorectal cancer, or hip fracture.

To use the methods described in Section 3, we restricted our analysis of the WHI-OS to those subjects who reported no current hormone use at baseline. This left 28,288 women with a uterus (CEE/MPA study) and 11,184 women without a uterus (CEE study). The TC interval width was set equal to 0.1 years. For choosing I_{max} as well as the ITC intervals, we used the 50 expected event criteria described in Section 4.

We report here the results of applying the PTC model since the limited number of events resulted in few or no ITC intervals. Unlike cancers, the CHD and stroke events are immediately evident without a latency period. This makes them the best candidates for analysis with a time varying treatment indicator. CHD was analyzed separately in the CEE/MPA and CEE studies since the clinical trial HR was substantially different for those two subject groups. Figures summarizing the results we now describe are contained in the Supplementary Materials. For the CEE study with CHD outcome (257 events) a total of 13 starting intervals (from (0.4,0.5]yrs to (1.7,1.8]yrs) and 27 stopping intervals (from (1.8,1.9]yrs to (4.4,4.5]yrs) are utilized, resulting in an HR estimate of 0.82 as compared to the ordinary Cox model estimated HR of 0.82 and the RCT estimated HR of 0.95. For the CEE/MPA study with CHD outcome (491 events) a total of 18 starting intervals (from (0.4,0.5]yrs to (2.1,2.2]yrs) and 38 stopping intervals (from (1.7,1.8]yrs to (5.4,5.5]yrs) are utilized by the PTC method, resulting in an HR estimate of 1.08 as compared to the ordinary Cox model estimated HR of 0.98 and the RCT estimated HR of 1.24. For the stroke outcome in the combined CEE/MPA and CEE studies (543 events), there are 18 starting intervals

(from (0.4,0.5]yrs to (2.1,2.2]yrs) and 39 stopping intervals (from (1.7,1.8]yrs to (5.5,5.6]yrs) utilized by the PTC method, resulting in an HR estimate of 1.23 as compared to the ordinary Cox model estimated HR of 1.04 and the RCT estimated HR of 1.34. Although the confidence intervals are wide and no firm conclusions can be drawn, the HR estimate for the CHD and stroke outcomes from the PTC model are closer to the corresponding RCT estimate than the ordinary Cox model estimate. This is especially the case for the analyses with larger numbers of events allowing more information on the TC process to be used through more TC intervals. The lack of precise information about the timing of exposure to hormones in the WHI-OS represents a severe limitation to full utilization of TC covariates. Further analyses need to be conducted with more modern electronic health record data bases that more accurately capture exposure information and have larger numbers of events.

6 Discussion

We considered the problem of analyzing time-to-event data from an observational study of a treatment where all subjects begin follow-up untreated with some subjects starting treatment during follow-up. Our goal was to obtain an estimate for the treatment effect through a Cox model. For the simplicity of our discussion, we assumed that the subject's age was the only measured confounder available, but in practice there would typically be additional measured confounders used in the analysis Cox model, such as the seven risk factors in our Women's Health Initiative analysis. We assumed that the true log hazard functions for starting treatment, stopping treatment, and the outcome of interest, each included age as well as an unobserved subject specific frailty as an additive term that is not dependent on the subject's time on study. We described the interval, parametric, and hybrid treatment choice (ITC, PTC, HTC) Cox models with time varying treatment choice (TC) covariates that try to recover information about frailty by modeling when subjects start and stop treatment. These models are created by designating an interval $[0, I_{max}]$ within the follow-up period and partitioning $[0, I_{max}]$ into fixed-length TC subintervals on which the TC covariates are constructed. The choice of I_{max} was based on an expected number of events condition.

The ITC model includes ITC covariates on a subset of the TC intervals, called ITC intervals, which meet an expected number of events condition. The ITC covariates are step functions that take at most one step for each subject. These steps are taken only at the beginning of an ITC interval. However, outside of those ITC intervals, the ITC model does not model starting or stopping treatment. In contrast, the PTC model only fits two TC covariates for starting treatment and two TC covariates for stopping treatment, on all of $[0, I_{max}]$: a cumulative constant TC covariate and a cumulative linear TC covariate. The cumulative constant TC covariate for starting treatment has the ordering property that it is largest for subjects who start treatment earlier and smallest for subjects for who never start treatment. The PTC model is most appropriate for when the treatment starting, stopping, and outcome event log hazard functions include frailty as an additive term. The HTC model is a hybrid approach between the ITC and PTC models. The HTC fits ITC covariates on the ITC intervals. Outside of the ITC intervals in $[0, I_{max}]$, it fits a single cumulative constant TC covariate for starting treatment and a single cumulative constant TC covariate for stopping treatment.

For most of our simulations, we used a fixed TC interval width = 0.1 years and found that as the sample size increased, the TC models' treatment effect bias decreased and sometimes leveled off at large sample sizes. However, further bias reduction was observed when the interval width was reduced to = 0.05 years. The HTC model was typically preferable to the ITC and PTC models for modest sample sizes and is recommended for most applications. The treatment effect estimate from the HTC model typically had slightly higher variability than the estimate from the ordinary Cox model, as would be expected from fitting the additional TC covariates. Nevertheless, we believe that the reduced bias justified the inclusion of the TC covariates and that our approach should be considered when an unobserved, time-independent frailty is suspected.

Including TC covariates in the analysis model requires a sufficient number of outcome events for the stability of the treatment effect estimate. Thus, we see our method as well suited for analyzing large electronic health record data bases which have a sufficient number of events. The current work does require that the true treatment effect has no carryover when the subject stops treatment. The degree to which this is true is typically unknown and the usual Cox model that fits a time-varying treatment indicator has the same shortcoming. One could perform a sensitivity analysis to see the degree to which different carryover effect patterns affect the conclusions by adjusting the time varying treatment indicator accordingly. However, assuming no carryover, our current work suggests that incorporating TC covariates can substantially reduce the bias of the treatment effect estimate as compared to the usual Cox model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Michael Sklar for thoughtful contributions to the work. We also thank an anonymous reviewer and associate editor for many helpful questions and suggestions that improved this work.

This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (<http://biowulf.nih.gov>).

The views expressed in this article are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; National Institutes of Health; or the United States Department of Health and Human Services.

References

1. Go AS, Yang J, Gurwitz JH, Hsu J, Lane K, Platt R. Comparative effectiveness of different β -adrenergic antagonists on mortality among adults with heart failure in clinical practice. *Archives of Internal Medicine*. 2008; 168:2415–2421.
2. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomized controlled trial findings. *British Medical Journal*. 2009; 338:b81. [PubMed: 19174434]
3. Yeh RW, Chandra M, McCulloch CE, Go AS. Accounting for the mortality benefit of drug-eluting stents in percutaneous coronary intervention: a comparison of methods in a retrospective cohort study. *BioMed Central Medicine*. 2011; 9:78.

4. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55.
5. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*. 2001; 96:440–448.
6. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
7. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risk data in survival analysis. *Statistics in Medicine*. 2009; 28:956–971. [PubMed: 19125387]
8. WHI Study Group. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Controlled Clinical Trials*. 1998; 19:61–109. [PubMed: 9492970]
9. Prentice RL. Chronic disease prevention research methods and their reliability, with illustrations from the Women's Health Initiative. *Journal of the American Statistical Association*. 2010; 105:1431–1443.
10. Yang S, Prentice RL. Assessing potentially time-dependent treatment effect from clinical trials and observational studies for survival data, with applications to the Women's Health Initiative Combined Hormone Therapy Trial. *Statistics in Medicine*. 2015; 34:1801–1817. [PubMed: 25689356]

APPENDIX

Lemma 1

Ordering property of the cumulative constant covariate. Suppose we have for TC intervals $I_1 = (0, \quad], I_2 = (\quad, 2 \quad], \dots, I_K = (I_{max} - \quad, I_{max}]$, and denote the corresponding TC terms as $TC_1(t), \dots, TC_K(t)$. Suppose for 3 subjects on study that subject 1 starts treatment before subject 2 while subject 3 never starts treatment. Suppose subject 1 starts treatment in I_{j_1} and subject 2 starts treatment in I_{j_2} . Then for subjects 1, 2, and 3's respective cumulative constant covariates $CTC_{01}(t)$, $CTC_{02}(t)$, and $CTC_{03}(t)$, we have

$$CTC_{01}(t) \geq CTC_{02}(t) \geq CTC_{03}(t) \quad \forall t \quad (16)$$

and

$$CTC_{02}(t) > CTC_{03}(t) \quad \forall t > j_2 \Delta. \quad (17)$$

Moreover, if $j_1 < j_2$, then

$$CTC_{01}(t) > CTC_{02}(t) \quad \forall t > j_1 \Delta. \quad (18)$$

Proof

The basic idea is that the cumulative constant covariate of a subject who starts treatment becomes one unit larger at the beginning of the next TC interval than the cumulative constant covariate of the subjects that did not yet start treatment. Since the subject who started treatment's cumulative constant covariate does not subsequently change, subjects who start in a later TC interval or never start at all never make up for that earlier one unit difference. Now we proceed more carefully through the various scenarios. Throughout the

proof, we have $TC_j(t) = 0 \forall t > j$ since $TC_j(t)$ only captures TC information for the interval $I_j = ((j-1)\Delta, j\Delta)$. Also, if a subject starts treatment in I_j , we have $TC_{j+1}(t) = \dots = TC_K(t) = 0 \forall t$ because that subject is not at risk to start treatment after I_j .

Since subject 1 starts treatment in $I_{j_1} = ((j_1-1)\Delta, j_1\Delta)$ and subject 2 starts treatment after subject 1,

$$TC_{i1}(t) = TC_{i2}(t) \text{ for } i=1, \dots, j_1-1 \text{ and } \forall t \quad (19)$$

If subject 2 also starts treatment in I_{j_1} albeit later than subject 1 so that $j_1 = j_2$, we also have

$$TC_{i1}(t) = TC_{i2}(t) \text{ for } i=j_1, \dots, K \text{ and } \forall t \quad (20)$$

Combining (19) and (20) gives us that if subjects 1 and 2 both start treatment in I_{j_1} , their cumulative constant functions are identical:

$$CTC_{01}(t) = TC_{11}(t) + \dots + TC_{K1}(t) = TC_{12}(t) + \dots + TC_{K2}(t) = CTC_{02}(t) \quad \forall t \quad (21)$$

Next suppose subject 2 starts treatment in a later TC interval than subject 1 so that $j_1 < j_2$. Then

$$TC_{j_1+1,1}(t) - TC_{j_1+1,2}(t) = 1 \quad \forall t > j_1\Delta \quad (22)$$

and

$$TC_{j_1+k,1}(t) = 0 \geq TC_{j_1+k,2}(t) \text{ for } k=1, j_2-j_1-1 \text{ and } \forall t > (j_1+k)\Delta \quad (23)$$

and

$$0 = TC_{j_2,1}(t) < TC_{j_2,2}(t) < 1 \text{ and } 0 = TC_{k,1}(t) = TC_{k,2}(t) \text{ for } k=j_2+1, \dots, K, \forall t > j_2\Delta \quad (24)$$

Combining (19) and (21)–(24) gives

$$CTC_{01}(t) = CTC_{02}(t) \quad \forall t \leq j_1\Delta \text{ and } CTC_{01}(t) > CTC_{02}(t) \quad \forall t > j_1\Delta \quad (25)$$

We have thus established the left hand inequality in (16) and (18).

As for (17) and the right hand inequality in (16), subject 2 starts treatment in $I_{j_2} = ((j_2-1)\Delta, j_2\Delta)$ and subject 3 never starts treatment. Thus,

$$TC_{i2}(t)=TC_{i3}(t) \text{ for } i=1,\dots,j_2-1 \text{ and } \forall t \quad (26)$$

and

$$TC_{j_2,2}(t) - TC_{j_2,3}(t)=1 \quad \forall t>j_2\Delta \quad (27)$$

and

$$TC_{k,2}(t)=0 \geq TC_{k,3}(t) \text{ for } k=j_2+1,\dots,K \text{ and } \forall t>k\Delta \quad (28)$$

Combining (26)–(28) establishes (17) and the right hand inequality of (16).

Figure 1.

Table 1

TC covariate and treatment indicator values for example data.

Subject	Model Term	Time Interval			
		(0, 1]	(1, 2]	(2, 3]	(3, 4]
# 1	$IT(t)$	0			
	$TC_{(0,1]}(t)$	0			
	$TC_{(1,2]}(t)$	0			
	$TC_{(2,3]}(t)$	0			
	cum. constant	0			
# 2	cum. linear	0			
	$IT(t)$	0	0	1	
	$TC_{(0,1]}(t)$	0	-1/5	-1/5	
	$TC_{(1,2]}(t)$	0	0	2/3	
	$TC_{(2,3]}(t)$	0	0	0	
# 3	cum. constant	0	-1/5	7/15	
	cum. linear	0	-1/5	17/15	
	$IT(t)$	0	1	1	
	$TC_{(0,1]}(t)$	0	4/5	4/5	
	$TC_{(1,2]}(t)$	0	0	0	
# 4	$TC_{(2,3]}(t)$	0	0	0	
	cum. constant	0	4/5	4/5	
	cum. linear	0	4/5	4/5	
	$IT(t)$	0	0	0	1
	$TC_{(0,1]}(t)$	0	-1/5	-1/5	-1/5
# 5	$TC_{(1,2]}(t)$	0	0	-1/3	-1/3
	$TC_{(2,3]}(t)$	0	0	0	1/2
	cum. constant	0	-1/5	-8/15	-1/30
	cum. linear	0	-1/5	-13/15	19/30

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Subject	Model Term	Time Interval			
		(0,1]	(1,2]	(2,3]	(3,4]
# 5	$IT(t)$	0	0		
	$TC_{(0,1]}(t)$	0	-1/5		
	$TC_{(1,2]}(t)$	0	0		
	$TC_{(2,3]}(t)$	0	0		
	cum. constant	0	-1/5		
	cum. linear	0	-1/5		
# 6	$IT(t)$	0	0	0	0
	$TC_{(0,1]}(t)$	0	-1/5	-1/5	-1/5
	$TC_{(1,2]}(t)$	0	0	-1/3	-1/3
	$TC_{(2,3]}(t)$	0	0	0	-1/2
	cum. constant	0	-1/5	-8/15	-31/30
	cum. linear	0	-1/5	-13/15	-71/30

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Bias^{*J*} of the estimation procedures for β_{3H} without stopping of treatment. ($a_f = 0.0$, $a_H = -2.0$, $\beta_{1H} = 0.02$, $a_S = -99.0$, $\beta_{1S} = 0.02$, $a_g = -1.5$, $a_H = -3.0$, $\beta_{1H} = 0.02$, $\beta_{2H} = 1.0$)

ψ	β_{2r}	β_{3r}	Propensity		Cox Model ³	PTC Model ⁴	HTC Model ⁵
			Match ²	Model ³			
0.0	2.0	0.0	+0.08	+0.03	+0.03	-0.013	-0.012
		-1.0	+0.329	-0.001	-0.00	-0.007	-0.004
0.8		1.0	-0.554	+0.002	+0.005	-0.005	-0.004
	2.0	0.0	+0.076	+0.110	+0.164	+0.052	+0.051
		-1.0	+0.505	+0.126	+0.176	+0.054	+0.056
		1.0	-0.604	+0.084	+0.142	+0.072	+0.068
	-2.0	0.0	-0.056	-0.130	-0.138	-0.126	-0.124
		-1.0	+0.345	-0.123	-0.133	-0.121	-0.119
		1.0	-0.694	-0.142	-0.144	-0.117	-0.114

¹Based on 500 replications each with $n = 1000$

²Including matching on follow-up time; model given in Supplementary Material

³Model given in Supplementary Material

⁴PTC model (15)

⁵HTC model (14)

Table 3

Bias $\pm$ 2 SSE^{*l*} of the TC models (13)–(15) for $\beta_{1H}^H = 0.02$, $\beta_{1S} = 0.02$, $\alpha_S = -99.0$, $\beta_{1H} = -2.0$, $\alpha_H = 0.0$, $\beta_{1I} = -1.5$, $\alpha_I = -3.0$, $\beta_{1H} = 0.02$, $\beta_{2H} = 1.0$)

ψ	β_{2H}	β_{1H}	Propensity Model ²	Cox Model ²	PTC Model ³	HTC Model ⁴	ITC Model ⁵
$n = 10000$							
0.0	2.0	0.0	-0.000 $\pm$.003	+0.001 $\pm$.003	+0.001 $\pm$.006	-0.004 $\pm$.006	-0.002 $\pm$.005
		-1.0	+0.001 $\pm$.004	+0.002 $\pm$.004	-0.004 $\pm$.008	-0.007 $\pm$.008	-0.004 $\pm$.006
		1.0	+0.000 $\pm$.003	+0.001 $\pm$.003	-0.001 $\pm$.005	-0.006 $\pm$.005	-0.004 $\pm$.004
0.8	2.0	0.0	+0.107 $\pm$.003	+0.159 $\pm$.003	+0.031 $\pm$.005	+0.028 $\pm$.005	+0.045 $\pm$.004
		-1.0	+0.127 $\pm$.004	+0.177 $\pm$.004	+0.019 $\pm$.007	+0.016 $\pm$.007	+0.052 $\pm$.006
		1.0	+0.079 $\pm$.002	+0.133 $\pm$.003	+0.058 $\pm$.004	+0.056 $\pm$.004	+0.062 $\pm$.004
	-2.0	0.0	-0.126 $\pm$.003	-0.132 $\pm$.003	-0.057 $\pm$.007	-0.064 $\pm$.007	-0.120 $\pm$.004
		-1.0	-0.116 $\pm$.004	-0.123 $\pm$.004	-0.064 $\pm$.010	-0.070 $\pm$.010	-0.117 $\pm$.006
		1.0	-0.138 $\pm$.003	-0.141 $\pm$.003	-0.052 $\pm$.006	-0.059 $\pm$.006	-0.123 $\pm$.004
$n = 20000$							
0.0	2.0	0.0	-0.000 $\pm$.002	+0.000 $\pm$.002	+0.003 $\pm$.004	-0.000 $\pm$.004	+0.000 $\pm$.004
		-1.0	-0.000 $\pm$.003	+0.000 $\pm$.003	-0.001 $\pm$.006	-0.005 $\pm$.006	-0.003 $\pm$.005
		1.0	-0.000 $\pm$.002	-0.000 $\pm$.002	-0.000 $\pm$.003	-0.004 $\pm$.003	-0.002 $\pm$.003
0.8	2.0	0.0	+0.108 $\pm$.002	+0.160 $\pm$.002	+0.033 $\pm$.003	+0.032 $\pm$.003	+0.039 $\pm$.003
		-1.0	+0.128 $\pm$.002	+0.179 $\pm$.003	+0.018 $\pm$.005	+0.016 $\pm$.005	+0.033 $\pm$.005
		1.0	+0.080 $\pm$.002	+0.134 $\pm$.002	+0.058 $\pm$.003	+0.057 $\pm$.003	+0.058 $\pm$.003
	-2.0	0.0	-0.126 $\pm$.002	-0.132 $\pm$.002	-0.049 $\pm$.006	-0.055 $\pm$.006	-0.102 $\pm$.004

ψ	β_{2tr}	β_{3h}	Propensity Model ²	Cox Model ²	PTC Model ³	HTC Model ⁴	ITC Model ⁵
		-1.0	-1.18±.003	-1.25±.003	-.062±.008	-.068±.008	-.105±.005
		1.0	-.139±.002	-.141±.002	-.046±.004	-.052±.004	-.102±.003

¹Based on 500 replications; SSE is the empirical standard deviation of the simulation estimate

²Model given in Supplementary Material

³PTC model (15)

⁴HTC model (14)

⁵ITC model (13)

Bias±2 SSE^I of the TC models (13)–(15) for β_{1H} without stopping of treatment and when all informative treatment decisions are fully captured (no outcome events or censoring during first year and no informative starting of treatment after first year). ($\alpha_f = 0.0$, $\alpha_{tr} = -2.0$, $\beta_{1tr} = 0.02$, $\alpha_s = -99.0$, $\beta_{1s} = 0.02$, $\alpha_g = -1.5$, $\alpha_h = -3.0$, $\beta_{1h} = 0.02$)

Table 4

ψ	β_{2r}	β_{1h}	Cox Model ²	PTC Model ³	HTC Model ⁴	ITC Model ⁵
$\beta_{2H} = 1.0$						
0.0	2	0.0	+0.001±.002	+0.001±.004	+0.000±.004	+0.000±.003
		-1.0	+0.001±.003	-0.002±.006	-0.003±.006	-0.001±.004
		1.0	+0.002±.002	+0.003±.004	+0.002±.004	+0.002±.002
0.8	2	0.0	+0.134±.002	-0.002±.004	+0.009±.004	+0.003±.003
		-1.0	+0.144±.003	-0.015±.006	-0.008±.006	+0.003±.004
		1.0	+0.111±.002	+0.026±.004	+0.045±.004	+0.003±.003
	-2	0.0	-0.048±.002	+0.005±.004	-0.002±.004	-0.001±.002
		-1.0	-0.035±.003	-0.007±.005	-0.017±.005	+0.008±.003
		1.0	-0.066±.002	+0.027±.003	+0.020±.003	-0.016±.002
$\beta_{2H} = 0.5$						
0.0	4	0.0	+0.001±.002	+0.001±.004	+0.000±.004	+0.000±.003
		-1.0	+0.001±.003	-0.002±.006	-0.003±.006	-0.001±.004
		1.0	+0.002±.002	+0.003±.004	+0.002±.004	+0.002±.002
0.8	4	0.0	+0.137±.003	-0.022±.006	+0.005±.006	+0.002±.004
		-1.0	+0.138±.004	-0.025±.007	-0.004±.008	-0.003±.005
		1.0	+0.130±.003	-0.018±.005	+0.013±.005	+0.006±.004
	-4	0.0	-0.028±.002	+0.005±.004	+0.001±.004	-0.001±.002

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

ψ	β_{irr}	β_{3h}	Cox Model ²	PTC Model ³	HTC Model ⁴	ITC Model ⁵
		-1.0	-.024±.003	+.000±.005	-.006±.006	+.002±.003
		1.0	-.032±.002	+.008±.003	+.004±.003	-.004±.002

¹Based on 500 replications each with $n = 20000$; SSE is the empirical standard deviation of the simulation estimate

²Model given in Supplementary Material

³PTC model (15)

⁴HTC model (14)

⁵ITC model (13)