HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2018 September 27.

Published in final edited form as:

Biometrics. 2018 September ; 74(3): 900–909. doi:10.1111/biom.12845.

Modeling Survival Distribution as a Function of Time to Treatment Discontinuation: a Dynamic Treatment Regime Approach

Shu Yang[1], **Anastasios A. Tsiatis**[2], and **Michael Blazing**[3]

[1]Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A

[2]Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A

[3]Duke University Medical Center

Summary

We consider estimating the effect that discontinuing a beneficial treatment will have on the distribution of a time to event clinical outcome, and in particular assessing whether there is a period of time over which the beneficial effect may continue after discontinuation. There are two major challenges. The first is to make a distinction between mandatory discontinuation, where by necessity treatment has to be terminated and optional discontinuation which is decided by the preference of the patient or physician. The innovation in this article is to cast the intervention in the form of a dynamic regime "terminate treatment optionally at time υ unless a mandatory treatment-terminating event occurs prior to υ" and consider estimating the distribution of time to event as a function of treatment regime υ. The second challenge arises from biases associated with the nonrandom assignment of treatment regimes, because, naturally, optional treatment discontinuation is left to the patient and physician, and so time to discontinuation may depend on the patient's disease status. To address this issue, we develop dynamic-regime Marginal Structural Models and use inverse probability of treatment weighting to estimate the impact of time to treatment discontinuation on a time to event outcome, compared to the effect of not discontinuing treatment. We illustrate our methods using the IMPROVE-IT data on cardiovascular disease.

Keywords

Dynamic treatment; Inverse weighting; Observational study; Survival distribution; Time-varying confounding; Treatment discontinuation

1. Introduction

The IMProved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE-IT; Cannon et al., 2008, 2015) was a clinical trial designed to examine whether the combination

Correspondence to: Shu Yang.

Supplementary Material

Web Appendices 1–4 referenced in Sections 3 and 5 and R code files are are available with this paper at the *Biometrics* website on Wiley Online Library.

of simvastatin (a statin drug) and ezetimibe translates into clinical benefit on cardiovascular disease (CVD) outcomes compared with simvastatin monotherapy in 18,144 patients stabilized and randomized within 10 days of an acute coronary syndrome (ACS). The primary endpoint was the first occurrence of cardiovascular death, nonfatal myocardial infarction, rehospitalization for unstable angina, coronary revascularization ≥ 30 days postrandomization, or stroke. Using the standard intent to treat (ITT) analysis, the trial results confirmed the superiority of the statin–ezetimibe therapy.

Because a substantial number of patients discontinued their assigned statin treatment, as a follow up to this study, we were asked to evaluate the effect that treatment discontinuation from statin drugs had on the clinical outcome. Specifically, whether there was a benefit to the use of statins that continued even after treatment was discontinued and, if so, how long that benefit would continue. That is, to evaluate whether there was a "legacy" effect of statin therapy. For some patients, discontinuation of treatment was mandatory for safety considerations; for example, when a serious side effect occurred. For some patients, discontinuation of treatment was optional; for example, the patient stopped statin drugs simply because s/he preferred to control cholesterol by healthy diet. It is important to distinguish between mandatory treatment discontinuation, where by necessity a patient must have their treatment terminated, and optional treatment discontinuation. Therefore, the appropriate clinical question should be "what are the consequences of having discontinued treatment for non-mandatory reasons on clinical outcome?" More precisely, we consider the dynamic treatment regime or treatment policy; namely, "discontinue treatment at time v as long as the patient did not have a treatment terminating event or a clinical outcome prior to time v". This allows the possibility that there are treatment-terminating events that would necessitate stopping treatment and therefore would respect clinical practice. The statistical problem would then be to consider the distribution of time to a clinical event as a function of time-to-discontinuation regime "v". The best and most direct way to answer this question is through a randomized study, where patients would be randomized to different time-to-discontinue regimes, and then the distribution of time to event could be compared directly for the different randomized groups using standard statistical methods. Unfortunately, such a time-to-discontinuation randomized trial is not available. We rely on the IMPROVE-IT data, where for each patient we observe a time-to-discontinuation (including ∞ if the patient stayed on treatment throughout the study) that was determined by the study protocol that dictates the situations for mandatory discontinuation or was decided by the patient or the physician for optional reasons, together with the patient's time to event possibly censored. Two complications arise in estimation and inference. Firstly, if the patient discontinued the treatment at time u for mandatory reasons or if a clinical outcome occurred, this patient is consistent with any regime "v" for $v \geq u$. Secondly, it is also difficult to draw correct causal conclusions for regime effects based on the IMPROVE-IT data due to confounding by indication or selection bias. For example, patients who are relatively unhealthy may decide to stop treatment earlier, so patients following different time-to-discontinuation regimes may differ systematically.

To answer the scientific question of interest, we characterize the population hazard rate (or equivalently the survival distribution) of having a clinical event as a function of time-to-discontinuation regime "v", which leads to a dynamic-regime Marginal Structural Model

(dr-MSM, e.g. Orellana et al., 2010). In this formulation, the dr-MSM for regime "υ" is conceived as the hazard rate of having a clinical event if all patients in the population had followed regime "υ". This idea of the dynamic treatment regime approach for discontinuation of treatment has been proposed to compare two competing treatments if no one was optionally discontinued, see Zhang et al. (2010), or using binary outcomes, see Johnson and Tsiatis (2004, 2005). The problem we are dealing with is considerably more complicated because the time-to-discontinuation regime of interest is continuous in time and the outcome of interest is time to event possibly censored, which adds another level of difficulty in modeling of regime effects and estimation.

Because treatment discontinuation can be affected by indications that occur over time, we must be careful with regards to time-dependent confounding. In such a case, even the time-dependent Cox model adjusting for time-dependent covariates is fallible (Robins et al., 2000; Daniel et al., 2013). This is because if decisions to stop treatment were influenced by a patient's state in the past, and additionally a patient's state may be affected by previous decision to treatment discontinuation, then simply conditioning on a patient's state may block the effect of past decision to treatment discontinuation, and such analyses may distort the true causal effect of the time-to-discontinuation regime.

In this article, we follow the idea of Inverse-Probability-of-Risk-Set-Weighting method in Robins (2002); Hernán et al. (2006) to adjust for time-varying confounders and derive consistent estimators of the parameters in dr-MSMs from the observed data, which involves finding individuals who are consistent with each regime and weighting them properly by the inverse probability of the treatment regime they actually followed. Like classical causal inference methods, our identifying assumption is that there is no unmeasured confounding (Robins et al., 1992), which implies that the set of time-varying confounders contains all prognostic factors that also affect decision to discontinue treatment over time.

This article is organized as follows. Section 2 introduces the conceptualization, models and assumptions for this article. Section 3 proposes our estimators as solutions to unbiased estimating equations. Section 4 reports a simulation study to evaluate finite sample properties of our estimator. Section 5 applies our methods to the IMPROVE-IT data to investigate the legacy effect of the statin–ezetimibe therapy on CVD outcomes. Section 6 concludes the article with a discussion.

2. Conceptualization, Models and Assumptions

2.1 Potential outcomes

We use the potential outcomes framework (Rubin, 1974) to formulate the problem in this article. We assume that the individuals are a random sample from a larger population. Let D denote the potential (not always observed) time to a clinical event ("failure" or "survival") or a treatment-terminating event if the individual never had treatment withdrawn. Let $T^{(D)}$ denote the potential time to a clinical event had treatment never been withdrawn except for a treatment terminating event. Note that $T^{(D)} = D$ if the individual has a clinical event without having a treatment terminating event. For $\upsilon < D$, we define $T^{(\upsilon)}$ to denote the potential (not always observed) time to a clinical event had treatment been withdrawn at time υ. It can then

be noted that each patient has associated potential outcomes $T^{(\upsilon)}$'s for all $\upsilon < D$. These are potential outcomes because in actuality, even without censoring, only one value $T^{(\upsilon)}$, $\upsilon \leq D$, can be observed for a particular patient. Thus, if we consider regime "υ", the corresponding potential clinical event time for an individual treated accordingly is given by $T^{(\upsilon)}$ if $\upsilon < D$, and $T^{(D)}$ if $\upsilon \geq D$. In addition to the potential outcomes $\{D, (T^{(\upsilon)} : \upsilon \leq D)\}$, we have a vector of baseline covariates X and the potential history of time-dependent covariates $\bar{Q}(u) = \{Q(\omega) : \omega \leq u\}$ observed for $u \leq D$, where $Q(\omega)$ denotes a vector of time-dependent covariates available at time ω, assuming that treatment had not been previously discontinued. Because of the possibility that the time to an event may be censored, we also define the time to censoring by C. We make the assumption of noninformative censoring under which the censoring time and the potential variables are independent, namely $C \perp\!\!\!\perp \{D, X, \bar{Q}(D), (T^{(\upsilon)} : \upsilon \leq D)\}$, where $\perp\!\!\!\perp$ means "independent of". Because the censored event times in IMPROV-IT were censored at the end of study follow-up (administrative censoring), the censoring time C is the time from the date of patient entry into the study until the date the study was terminated. In such cases, noninformative censoring can be reasonably assumed. In some studies however patients may also be censored due to drop out. In that case the assumption that C is independent of $\{D, X, \bar{Q}(D), (T^{(\upsilon)} : \upsilon \leq D)\}$ may not be reasonable. In Section 6 we have a short discussion on modifications that can be made to weaken the assumption of noninformative censoring. In any case, in our setting we consider the full set of potential outcomes to be given by $F = \{D, X, \bar{Q}(D), (T^{(\upsilon)} : \upsilon \leq D), C\}$.

2.2 Dynamic-regime marginal structural model approach to the legacy effect of treatment discontinuation

For comparing time-to-discontinuation regimes, we consider the ideal case where we can observe all potential outcomes in F for each individual. We define the causal parameter

$$\lambda_{\upsilon}(t) = \lim_{h \to 0} h^{-1} P(t \leq T^{(\upsilon)} < t + h | T^{(\upsilon)} \geq t)$$

as the hazard rate of failing at time t for a population of patients had they been subjected to the policy of treatment discontinuation at time υ. Our interest is to evaluate $\lambda_{\upsilon}(t)$ as a function of υ. We note that by the natural restrictions of the policy, $\lambda_{\upsilon}(t) = \lambda_t(t)$ for $\upsilon \geq t$, because if treatment were to continue until time υ for $\upsilon > t$, then at time t, they would have been in the same status as if treatment were to have continued until time t. We also define $U^{(\upsilon)} = \min(T^{(\upsilon)}, C)$ and $\Delta^{(\upsilon)} = I(T^{(\upsilon)} \leq C)$ to denote the time to a clinical event or censoring and the clinical event indicator, respectively, for an individual subjected to regime "υ".

Under noninformative censoring, we have

$$\lambda_{\upsilon}(t) = \lim_{h \to 0} h^{-1} P(t \leq U^{(\upsilon)} < t + h, \Delta^{(\upsilon)} = 1 | U^{(\upsilon)} \geq t).$$

We also define the counting process $N^{(\upsilon)}(t) = I(U^{(\upsilon)} \leq t, \Delta^{(\upsilon)} = 1)$ and the at-risk process $Y^{(\upsilon)}(t) = I(U^{(\upsilon)} \geq t)$, with the following natural constraints:

$$dN^{(\upsilon)}(t) = dN^{(t)}(t) \text{ for } \upsilon \geq t, \quad (1)$$

$$Y^{(\upsilon)}(t) = Y^{(t)}(t) \text{ for } \upsilon \geq t\,. \quad (2)$$

We consider semiparametric models that are aimed at modeling the legacy effect after treatment discontinuation. We break up "time after treatment discontinuation" into intervals $[0, l_1), [l_1, l_2), \ldots, [l_{p-1}, l_p)$. At any time t, we define the following time-dependent covariates

$$z_\upsilon(t) = I_j \text{ if } l_{j-1} \leq t - \upsilon < l_j, j = 1, \ldots, p, \quad (3)$$

$$z_\upsilon(t) = I_0 \text{ if } \upsilon > t, \quad (4)$$

where $I_j = (0, \ldots, 1, \ldots, 0)^T$ is the p-dimensional vector with 1 as the jth element, for $j = 1, \ldots p$, and $I_0 = (0, \ldots, 0)^T$. Note that $z_\upsilon(t)$ indicates how long an individual subjected to regime "υ" had stopped treatment at a specific time t, had he or she not had a treatment terminating event or clinical event. To evaluate the legacy effect of treatment after discontinuation, we consider the following model.

Definition 1 (Dynamic-Regime Marginal Structural Model, dr-MSM)—The marginal structural model for comparing all dynamic regimes "υ"'s is the proportional hazards model

$$\lambda_\upsilon(t) = \lambda_0(t) \exp\{\beta^T z_\upsilon(t)\}, \quad (5)$$

where $\lambda_0(t)$ is the population hazard rate if all individuals had treatment never withdrawn, $\beta = (\beta_1, \ldots, \beta_p)^T$, and $z_\upsilon(t)$ takes on p-dimensional vector values, defined in (3) and (4).

This is a model for the regime effects, because it compares the outcome under different regimes in the same group of individuals (namely, the population of all individuals). The parameter β_j describes the relative hazard of having a clinical event for individuals that have stopped treatment for a period of time $[l_{j-1}, l_j)$ as compared to if all individuals stayed on treatment. Thus, if we expect the hazard of having a clinical event to increase the longer that a patient has gone off treatment, then we would expect $0 < \beta_1 < \beta_2 < \cdots < \beta_p$. However, if there is a legacy effect, then we might expect some of the β_j's (for j small) to be close to zero.

Model (5), like any parametric or semiparametric model, is generally not correct but may be useful for three reasons. First, our model preserves the null hypothesis. If going off treatment

optionally has no effect on the clinical outcome; that is, $\lambda_v(t) = \lambda_0(t)$ for all "v", then our model would be correctly specified in that all the β_j's would be zero. Secondly, if there is a legacy effect, we expect such a model to be able to broadly capture such effects, and moreover the model parameter β is easy to interpret. As pointed out by a reviewer, other models, such as Aalen's additive hazard model (Aalen, 1989) and sequential Cox models (Gran et al., 2010), might be interesting alternatives. Extension of our methods to these models however is beyond the scope of the current paper and can be an interesting avenue for future research. Thirdly, we shall show in Section 3.1 that the proposed estimator can be understood from a partial likelihood point of view, and that it can often be calculated using standard software. The possibility for our proposed method to use standard software is attractive, because it can be easily implemented.

We have cast the problem through potential outcomes and developed a marginal structural model for the distribution of these potential outcomes that is useful in answering the clinical question of interest. In order that this be of use to us, we must be able to estimate the parameters in the dr-MSM using the observed data. We now discuss the observed data we get to see and the assumptions necessary to use these data for our purpose.

2.3 Observed data and the no unmeasured confounding assumption

In contrast to the potential outcomes, the observed data can be summarized as $O_i = \{U_i = \min(T_i, C_i), \Delta_i = I(T_i \leq C_i), V_i, \Gamma_i, X_i, \bar{Q}_i(U_i)\}$, for $i = 1, \ldots, n$, where T_i is the time to a clinical event, C_i is the observed censoring time, U_i is the time to a clinical event or censoring, Δ_i is the clinical event indicator, V_i is the time to optional treatment discontinuation, a clinical event, censoring, or a treatment-terminating event, and Γ_i is the indicator of optional treatment discontinuation; that is, $\Gamma_i = 1$ if treatment discontinuation was optional and 0 if treatment discontinuation was due to a clinical event, censoring, or a treatment-terminating event. If $\Gamma_i = 0$ and treatment discontinuation was due to a clinical event or censoring, then $U_i = V_i$, otherwise, $U_i > V_i$. Also if $\Gamma_i = 1$, the regime for individual i actually followed is regime "V_i"; however, if $\Gamma_i = 0$, the regime intended for individual i is censored. Therefore, the observed possibly censored outcome U_i and the observed possibly censored discontinuation time V_i are both continuous in time and are subject to censoring. X_i is a vector of baseline covariates, $\bar{Q}_i(U_i)$ is a vector of time-dependent covariates collected up to time U_i

We now define the observed data counting process as $N_i(t) = I(U_i \leq t, \Delta_i = 1)$ and the at-risk process as $Y_i(t) = I(U_i \geq t)$. We make the consistency assumption that links the observed data processes with the potential outcome processes.

Assumption 1 (Consistency)—If $\Gamma_i = 1$, $N_i(t) = N_i^{(V_i)}(t)$ and $Y_i(t) = Y_i^{(V_i)}(t)$ for all t; if $\Gamma_i = 0$, $N_i(t) = N_i^{(v)}(t)$ and $Y_i(t) = Y_i^{(v)}(t)$, for $v \geq V_i$ and all t.

If the treatment regime that the individual actually followed is observed ($\Gamma_i = 1$), the observed data processes are the potential outcome processes under the treatment regime actually followed; if the treatment regime is censored ($\Gamma_i = 0$), then this individual is

consistent to any regime "υ" for $\upsilon \geq V_i$, and the observed outcome processes equal the potential outcome processes if the individual followed regime "υ".

In order to use the distribution of the observed data to estimate the parameters in the dr-MSM (5), we require the assumption of no unmeasured confounding (Robins, 2004).

Assumption 2 (No Unmeasured Confounding)—The hazard of optional treatment discontinuation is

$$\begin{aligned}\mu(\upsilon, F) &= \lim_{h \to 0} h^{-1} P(\upsilon \leq V < \upsilon + h, \Gamma = 1 | V \geq \upsilon, F) \\ &= \lim_{h \to 0} h^{-1} P(\upsilon \leq V < \upsilon + h, \Gamma = 1 | V \geq \upsilon, \bar{H}_\upsilon) = \mu\left\{\upsilon, \bar{H}(\upsilon)\right\},\end{aligned} \tag{6}$$

where $\bar{H}(\upsilon) = \{X, \bar{Q}(\upsilon)\}$.

Assumption 2 implies that the decision of discontinuing treatment optionally at time υ is independent of the future prognosis, given the past history of treatments and covariates up to time υ. This assumption holds if the set of historical covariates contains all prognostic factors for the outcome that affect the decision of going off treatment optionally at υ. This is a key assumption to identifying the parameters in the dr-MSM based on the observed data; however, this assumption can not be tested empirically. Therefore, it requires careful consultation with the subject matter experts on collecting a rich set of covariates to ensure this assumption holds.

From the hazard of optional discontinuation $\mu\{\upsilon, \bar{H}(\upsilon)\}$ defined in (6), we now define

$$K\{\upsilon, \bar{H}(\upsilon)\} = \exp\left[-\int_0^\upsilon \mu\{\omega, \bar{H}(\omega)\} \mathrm{d}\omega\right] \tag{7}$$

and

$$f\{\upsilon, \bar{H}(\upsilon)\} = \mu\{\upsilon, \bar{H}(\upsilon)\} K\{\upsilon, \bar{H}(\upsilon)\}. \tag{8}$$

For individual i, denote $K_i(\upsilon) = K\{\upsilon, \bar{H}_i(\upsilon)\}$ and $f_i(\upsilon) = f\{\upsilon, \bar{H}_i(\upsilon)\}$ for shorthand. These can be viewed as the probability of individual i not having optionally discontinued before time υ and the probability of optionally discontinuing at time $[\upsilon, \upsilon + \mathrm{d}\upsilon)$, respectively. We also impose a sufficient condition for $\mu\{\upsilon, \bar{H}(\upsilon)\}$.

Assumption 3 (Positivity)—There exist constants δ_1 and δ_2 such that with probability 1, $0 < \delta_1 \leq \mu\{\upsilon, \bar{H}(\upsilon)\} \leq \delta_2$ for υ in the support of U.

Assumption 3 implies that for some constants ε_1 and ε_2, with probability 1, $K\{\upsilon, \bar{H}(\upsilon)\} \geq \varepsilon_1 > 0$ and $f\{\upsilon, \bar{H}(\upsilon)\} \geq \varepsilon_2 > 0$ for υ in the support of U; that is, each individual has a positive

probability of discontinuing treatment for optional reasons at any time υ before having a clinical event or being censored.

3. Estimation: unbiased estimating equations

Define $\Lambda_0(t) = \int_0^t \lambda_0(\upsilon)\mathrm{d}\upsilon$ as the cumulative hazard function at time t and in what follows, we derive our estimators for β and $\Lambda_0(t)$, $t \geq 0$.

To motivate our estimators, we first assume that we are able to observe potential outcomes. If we fix the value of "υ", then the process $M^{(\upsilon)}(t) = N^{(\upsilon)}(t) - \int_0^t \exp\{\beta^T z_\upsilon(\omega)\} Y^{(\upsilon)}(\omega)\mathrm{d}\Lambda_0(\omega)$ is a mean zero martingale. By the martingale property of $M^{(\upsilon)}(t)$, we have a system of estimating equations for β and $\Lambda_0(t)$, $t \geq 0$:

$$\mathrm{d}M^{(\upsilon)}(t) = 0, t \geq 0, \qquad \int_0^\infty z_\upsilon(t)^T \mathrm{d}M^{(\upsilon)}(t) = 0, \quad (9)$$

where $\mathrm{d}M^{(\upsilon)}(t) = \mathrm{d}N^{(\upsilon)}(t) - \mathrm{d}\Lambda_0(t) \exp\{\beta^T z_\upsilon(t)\} Y^{(\upsilon)}(t)$. (9) was chosen because it would lead to the usual Breslow estimator (Breslow and Clayton, 1993) for $\Lambda_0(t)$ and the maximum partial likelihood estimator (Cox, 1972, 1975) for β.

We now integrate (9) over "υ" with a weight function $\theta(\upsilon)$ and sum over all individuals, to obtain the estimating equation

$$\int_0^\infty \mathrm{d}M_i^{(\upsilon)}(t)\theta(\upsilon)\mathrm{d}\upsilon = 0, t \geq 0, \qquad \int_0^\infty z_\upsilon(t)^T \mathrm{d}M^{(\upsilon)}(t)\theta(\upsilon)\mathrm{d}\upsilon = 0\,. \quad (10)$$

Given the constraints (1), (2), and the fact that $z_\upsilon(t) = I_0$ for $\upsilon > t$, we have $\mathrm{d}M^{(\upsilon)}(t) = \mathrm{d}M^{(t)}(t) = \mathrm{d}N^{(t)}(t) - \mathrm{d}\Lambda_0(t) Y^{(t)}(t)$ for $\upsilon > t$, and (10) can be further expressed as

$$\sum_{i=1}^n \left\{ \int_0^t \mathrm{d}M_i^{(\upsilon)}(t)\theta(\upsilon)\mathrm{d}\upsilon + \mathrm{d}M_i^{(t)}(t) \int_t^\infty \theta(\upsilon)\mathrm{d}\upsilon \right\} = 0, \quad t \geq 0, \quad (11)$$

$$\sum_{i=1}^n \int_t^\infty \int_0^t z_\upsilon(t)\mathrm{d}M_i^{(\upsilon)}(t)\theta(\upsilon)\mathrm{d}\upsilon = 0\,. \quad (12)$$

The fundamental problem for the estimating equations (11) and (12) is that not all potential outcomes can be observed for a particular individual, and therefore (11) and (12) are infeasible. In order to use the observed data, we follow the idea of Inverse-Probability-Risk-Set-Weighting (Robins, 2002; Hernán et al., 2006; Robins et al., 2008).

We first consider individuals who were are at risk at time t: $\{i: Y_i(t) = 1\}$. We consider three subsets of individuals: group (a) $\{i: V_i \le t, \Gamma_i = 1\}$; group (b) $\{i: V_i > t\}$; and group (c) $\{i: V_i \le t, \Gamma_i = 0\}$, and weighting the contributions for those three sets of individuals so that they mimic the contributions in (11) and (12) had all potential outcomes been observed at time t. The following theorem presents the unbiased observed data estimating functions for β and $\Lambda_0(t)$, $t \ge 0$, with its proof presented in Web Appendix 1 of the Supplementary Material available online.

Theorem 1 (Unbiased estimating functions)

The unbiased observed data estimating functions for β and $\Lambda_0(t)$, $t \ge 0$ are

$$\sum_{i=1}^{n} I(V_i \le t, \Gamma_i = 1)[\mathrm{d}N_i(t) - \mathrm{d}\Lambda_0(t)\exp\{\beta^T z_{V_i}(t)\}Y_i(t)]\frac{\theta(V_i)}{f_i(V_i)}$$
$$+ \sum_{i=1}^{n} I(V_i \ge t)\{\mathrm{d}N_i(t) - \mathrm{d}\Lambda_0(t)Y_i(t)\}\frac{\bar{\theta}(t)}{K_i(t)} + \sum_{i=1}^{n}\frac{I(V_i \le t, \Gamma_i = 0)}{K_i(V_i)}\int_{V_i}^{t}[\mathrm{d}N_i(t) - \mathrm{d}\Lambda_0(t)\exp$$
$$\{\beta^T z_v(t)\}Y_i(t)]\theta(v)\mathrm{d}v + \sum_{i=1}^{n}\frac{I(V_i \le t, \Gamma_i = 0)}{K_i(V_i)}\{\mathrm{d}N_i(t) - \mathrm{d}\Lambda_0(t)Y_i(t)\}\bar{\theta}(t), \quad (13)$$

for $t \ge 0$, and

$$\sum_{i=1}^{n}\int_0^{\infty} I(V_i \le t, \Gamma_i = 1)z_{V_i}(t)[\mathrm{d}N_i(t) - \mathrm{d}\Lambda_0(t)\exp\{\beta^T z_{V_i}(t)\}Y_i(t)]\frac{\theta(V_i)}{f_i(V_i)} \quad (14)$$
$$+ \sum_{i=1}^{n}\int_0^{\infty}\frac{I(V_i \le t, \Gamma_i = 0)}{K_i(V_i)}\int_{V_i}^{t} z_v(t)[\mathrm{d}N_i(t) - \mathrm{d}\Lambda_0(t)\exp\{\beta^T z_v(t)\}Y_i(t)]\theta(v)\mathrm{d}v,$$

respectively, where $\bar{\theta}(t) = \int_t^{\infty}\theta(v)dv$.

Remark 1

For groups (a) and (b), there is only one contribution for each individual for a given "v"; namely, regime "V_i" for individuals in group (a), and regime "t" for individuals in group (b). However, group (c) is the most complicated as there are contributions for regimes "v" for any $v \ge V_i$.

The intuition for the weighting scheme in (13) producing unbiased estimation functions is the following. For individuals in group (a), weighting by $\theta(V_i)/f_i(V_i)$ creates a pseudo-sample where all individuals were randomly assigned to each of regimes "V_i". For individuals in group (b), weighting by $\bar{\theta}(t)/K_i(t)$ creates a pseudo-sample where all individuals followed regime "t". For individuals in group (c), weighting by $\theta(v)/K_i(V_i)$

creates a pseudo-sample where all individuals followed regime "υ" for any $\upsilon \geq V_i$. Thus, integrating over all "υ"'s implicitly removes confounding by creating a pseudo-population in which individuals were randomly assigned to each of regimes "υ".

3.1 Algorithm for estimating β

We have noted that $z_\upsilon(t)$ can take on one of $(p+1)$ vector of dummy variables $I_0, \ldots, I_p$. Let us define for an individual i in group (c) the intervals κ_{ij} between time points $(V_i, t]$, corresponding to $[t - \min(l_j, t - V_i), t - \min(l_{j-1}, t - V_i)]$, for $j = 1, \ldots, p$, where $l_0 = 0$. Note that if $t - V_i > l_j$, then $\kappa_{ij} = [V_i, V_i]$, which we take as the null set. See Figure 1 for an example for visualization of κ_{ij}.

In this case, (13) can be written as

$$\sum_{i=1}^{n} I(V_i \leq t, \Gamma_i = 0) \frac{\sum_{j=0}^{p} \{dN_i(t) - d\Lambda_0(t) \exp(\beta^T I_j) Y_i(t)\} \int_{\kappa_{ij}} \theta(\upsilon) d\upsilon}{K_i(V_i)} + \sum_{i=1}^{n} I(V_i \leq t, \Gamma_i = 0) \frac{\{dN_i(t) - d\Lambda_0(t) Y_i(t)\} \bar{\theta}(t)}{K_i(V_i)}.$$

If $(t - V_i)$ does not contain the jth interval, then κ_{ij} is a null set and $\int_{\kappa_{ij}} \theta(\upsilon) d\upsilon = 0$.

This suggests the following strategy. For each individual i, create $p+1$ records with time independent covariates $I_0, \ldots, I_p$, so that for the ith individual, there are $p+1$ records $\{U_i, \Delta_i, V_i, \bar{H}_i(V_i), I_i\}_{j=0}^{p}$. At any risk set time t, we create a weight function $\omega_{ij}(t)$, $i = 1, \ldots, n$, $j = 1, \ldots, p$, as follows.

Group a. With $V_i \leq t$ and $\Gamma_i = 1$,

$$\omega_{ij}(t) = \begin{cases} \dfrac{\theta(V_i)}{f_i(V_i)}, & \text{if } l_{j-1} \leq t - V_i < l_j, \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

We note that only one of the $(p+1)$ records for the ith individual would have a positive weight.

Group b. With $V_i > t$,

$$\omega_{ij}(t) = \begin{cases} \dfrac{\bar{\theta}(t)}{K_i(t)}, & \text{for } j = 0, \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

We note that only the 0th record for the ith individual would have a positive weight.

Group c. With $V_i \leq t$ and $\Gamma_i = 0$,

$$\omega_{ij}(t) = \begin{cases} \dfrac{\int_{\kappa_{ij}} \theta(\omega)\mathrm{d}\omega}{K_i(V_i)}, & \text{for } j = 1, \ldots, p, \\ \dfrac{\bar{\theta}(t)}{K_i(V_i)}, & \text{for } j = 0\,. \end{cases} \quad (17)$$

We note that the interval κ_{ij} may be null for some j, in which case the weight is zero.

We can now write (13) and (14) as

$$\sum_{i=0}^{n}\sum_{j=0}^{p} \omega_{ij}(t)\{\mathrm{d}N_i(t) - \mathrm{d}\Lambda_0(t)\exp(\beta^T I_j)Y_i(t)\}, \quad (18)$$

and

$$\sum_{i=0}^{n}\sum_{j=0}^{p} \int_0^{\infty} \omega_{ij}(t) I_j\{\mathrm{d}N_i(t) - \mathrm{d}\Lambda_0(t)\exp(\beta^T I_j)Y_i(t)\}\,. \quad (19)$$

Setting (18) equal to zero, we would obtain the estimator for $\mathrm{d}\Lambda_0(t)$ (for fixed β) as

$$\mathrm{d}\widehat{\Lambda}_0(t) = \frac{\sum_{i=1}^{n}\sum_{j=0}^{p}\omega_{ij}(t)\mathrm{d}N_i(t)}{\sum_{i=1}^{n}\sum_{j=0}^{p}\omega_{ij}(t)\exp(\beta^T I_j)Y_i(t)}\,. \quad (20)$$

Substituting (20) into (19), we obtain the estimating equation to solve for β; namely, $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^T$ is the solution to the estimating equation

$$\sum_{i=1}^{n}\sum_{j=0}^{p} \int_0^{\infty} \omega_{ij}(t)\left\{I_j - \frac{\sum_{i=1}^{n}\sum_{j=0}^{p}\omega_{ij}(t)I_j\exp(\beta^T I_j)Y_i(t)}{\sum_{i=1}^{n}\sum_{j=0}^{p}\omega_{ij}(t)\exp(\beta^T I_j)Y_i(t)}\right\}\mathrm{d}N_i(t) = 0, \quad (21)$$

which is the partial score equation for a proportional hazards model with time-dependent weights. Therefore, the estimator for β can be calculated by standard software. Using the theory of estimating equations, $\hat{\beta}$ is consistent and asymptotically normal, and its variance can be estimated by the robust variance estimate from standard software. We give a heuristic proof of these asymptotic results in Web Appendix 2 of the Supplementary Material.

The proof of Theorem 1 and the asymptotic properties of the estimators assume that the hazard of optional treatment discontinuation $\mu\{v, \bar{H}(v)\}$ is known and correctly specified. In actuality, this must be estimated using the observed data. It is convenient in such cases to posit a proportional hazard model with time-dependent covariates for this purpose; that is,

$\mu\{t, \bar{H}(t)\} = \mu_0(t)g\{\bar{H}(t); \alpha\}$. We can obtain estimators for $\mu_0(t)$ and α using standard software for time-dependent covariates. Using the standard counting process argument (Andersen et al., 1993), $\hat{K}_i(t) = \exp[-\int_0^t \hat{\mu}_0(\omega)g\{\bar{H}_i(\omega); \hat{\alpha}\}\mathrm{d}\omega]$ is a uniformly root-n consistent estimator for $K_i(t)$ and these estimators can be substituted for $K_i(t)$ and $K_i(V_i)$ in the estimating functions (13). However, we still need an estimator for $\theta(V_i)/f_i(V_i)$. We note that the function $\theta(v)$ can be an arbitrary function of v leading to an unbiased estimating function, although the choice of $\theta(v)$ may affect the efficiency of the resulting estimation. By judicious choice, $\theta(v)$ can serve as a stabilized weight (Hernán et al., 2000) and also allow a consistent estimator for $\theta(v)/f_i(v)$. Specifically, we consider $\theta(v) = \mu_0(v)\exp\{-\int_0^v \mu_0(\omega)\mathrm{d}\omega\}$. In this case, $\theta(v)/f_i(v) = \exp\{-\int_0^v \mu_0(\omega)\mathrm{d}\omega\}/[g\{\bar{H}_i(v); \alpha\}K_i(v)]$. Because α, $K_i(v)$ and $\int_0^v \mu_0(\omega)\mathrm{d}\omega$ have uniformly root-n consistent estimators, we can replace $\theta(V_i)/f_i(V_i)$ in (13) by $\hat{\theta}(V_i)/\widehat{f_i}(V_i) = \exp\{-\int_0^{V_i} \hat{\mu}_0(\omega)\mathrm{d}\omega\}/[g\{\bar{H}_i(V_i); \hat{\alpha}\}\hat{K}_i(V_i)]$ to obtain a consistent estimator for β. To understand the stabilized weights, suppose that V_i is unconfounded, the stabilized weights are essentially 1, which would lead to efficient estimation for β with no confounding. With confounding, the stabilized weights will not be constant but will vary around 1; however, they will still tend to be much less variable than the unstabilized weights.

If the model for the hazard of optional discontinuation is correctly specified, the estimator for β obtained by solving (21) with the estimated weights is still consistent and asymptotically normal. However, the asymptotic variance will be affected by the estimation of the weights in a complicated fashion. Therefore, for variance estimation, we suggest using the nonparametric bootstrap (Efron and Tibshirani, 1994) method.

To summarize, the algorithm for developing an estimator for β is the following.

Step 1. Using the data $\{V_i, \Gamma_i, \bar{H}_i(V_i)\}$, $i = 1, \ldots, n$, derive the estimator for the proportional hazards model $\mu\{v, \bar{H}(v)\} = \mu_0(v)g\{\bar{H}(v); \alpha\}$, and obtain the estimators for $\theta(v)$, $K_i(v)$ and $\theta(v)/f_i(v)$.

Step 2. Create $p + 1$ records for each individual i; namely, for individual i, (U_i, Δ_i, I_j), $j = 0, \ldots, p$, where $I_0 = (0, \ldots, 0)$ and $I_j = (0, \ldots, 1, \ldots, 0)$, where the jth element is 1.

Step 3. For each of the $(p + 1)n$ records, create time-dependent weights $\omega_{ij}(t)$, $i = 1, \ldots, n$, $j = 0, \ldots, p$, and all values t for which there was an observed clinical event, according to (15), (16), and (17) with the estimated values of $\theta(t)$, $K_i(t)$ and $\theta(t)/f_i(t)$.

Step 4. Obtain estimator for $\beta = (\beta_1, \ldots, \beta_p)^T$ by solving (21) using the standard software; e.g., the function “coxph” in R with the weighting argument.

4. Simulation Study

We now evaluate the finite-sample performance of the proposed estimator on simulated data sets. We report the simulation results under the setting with a null effect of treatment discontinuation time below. Additional simulation studies to detect a non-zero legacy treatment effect are reported in Web Appendix 4 in the Supplementary Material online.

We simulate 2, 000 data sets. We first generate the covariate process if an individual stayed on treatment throughout time $\tau = 1.5$, $\bar{H}(\tau) = \{X, \bar{Q}(\tau)\}$ one time-independent (X) and one time-dependent ($Q(t)$). The time-independent covariate X is generated from a Bernoulli distribution with mean equal to 0.55. The time-dependent covariate is a 1×3 row vector generated from a multivariate normal distribution with mean equal to (0, 0, 0) and covariance equal to $0.7^{|i-j|}$ for $i, j = 1, 2, 3$. This vector represents the values of $Q(t)$ at times $t_1 = 0$, $t_2 = 0.5$, and $t_3 = 1$. We assume that the time-dependent variable remains constant between measurements.

We generate the time until mandatory treatment discontinuation, V_1, according to a proportional hazards model with baseline covariates, $\lambda_1\{t, H(0)\} = 1.1 \exp\{0.1X + 0.1Q(0)\}$, and the time until optional discontinuation, V_2, according to a proportional hazards model with time-dependent covariates, $\lambda_2\{t, \bar{H}(t)\} = 2 \exp\{0.5X + 0.5Q(t)\}$. See Web Appendix 3 for the steps to generate V_2. We generate the time until a clinical event, T, according to a proportional hazards model with baseline covariates,

$$\lambda_T\{t, H(0)\} = 2 \exp\{0.5X + 0.5Q(0)\}. \quad (22)$$

We consider an administrative type of censoring, where the censoring time C is generated from a uniform distribution from 1.2 to τ. Let $U = \min(T, C)$. If $T < C$, $\Delta = 1$; otherwise $\Delta = 0$. Finally, let $V = \min(V_1, V_2, T, C)$. If $V = V_1$, $\Gamma = 1$, which is the indicator of optional discontinuation, otherwise, $\Gamma = 0$. The observed time-dependent covariate process is $Q(t)$ if $t \leq V_2$ and $Q(t) + T$ if $t > V_2$ to reflect that the covariate process is affected after treatment discontinuation and is also related to survival time. The observed data is $\{U_i, \Delta_i, V_i, \Gamma_i, \bar{H}_i(V_i)\}$ for $i = 1, \ldots, n = 1, 000$. From our data generating mechanism, approximately 40% observed treatment discontinuation times are due to optional reasons, 22% are due to mandatory reasons, and 38% are censored by clinical events or censoring.

We assume that the dr-MSM is $\lambda(t) = \lambda_0(t) \exp\{\beta Z_v(t)\}$, where $Z_v(t) = 1$ if $v \leq t$, and $Z_v(t) = 0$ if $v > t$. In this model, β quantifies the relative hazard of patients that have ever stopped treatment as compared to those still on treatment. From the data generating model (22), the true value is $\beta = 0$.

We consider the following estimators for β: (i) a. the proposed estimator of β in model (5) and b. the proposed estimator for β in a model which adjusts for baseline covariates; that is, the inverse probability weighted estimator for β in the marginal structural model $\lambda_v(t; X) = \lambda_0(t) \exp\{\beta^T z_v(t) + \gamma^T X\}$; (ii) Naive I: the naive estimator, which is obtained by the time-dependent Cox proportional hazards model with (U_i, Δ_i) and the observed time-dependent covariate $I(V_i \leq t)$, for which we consider three cases for covariate adjustment: a. without covariates, b. with time-independent covariates, and c. with both time-independent and time-dependent covariates; and (iii) Naive II: consider Naive I, censoring those with mandatory discontinuation at the time they discontinue. For standard errors, we consider the robust estimate output from the standard software, which treats the weight function as known, and the bootstrap standard error with a bootstrap size of 100.

Table 1 shows the simulation results for estimating β. The naive methods are biased, even adjusting for covariates; whereas the proposal method has small biases. In our dr-MSM approach, whether or not we adjust for covariates further in model (5), the results remain close. Moreover, for the proposed estimators, both the standard software and the bootstrap work well for producing standard errors close to the true standard deviations and coverage rates close to the nominal coverage.

5. Application to the IMPROVE-IT Data

We present an analysis in IMPROVE-IT. The IMPROVE-IT data consists of 17; 706 on-treatment patients who received at least one dose of study medication. We included a subset of 17; 654 patients whose time-dependent covariates were measured in order to adjust for confounding. Among these patients, 8; 833 received statin monotherapy (the control treatment, $A = 0$), and 8; 821 received statin-ezetimibe therapy (the active treatment, $A = 1$). The primary end point was the first occurrence of cardiovascular (CV) death, nonfatal myocardial infarction (MI), rehospitalization for unstable angina, coronary revascularization (occurring at least 30 days after randomization), or stroke.

Patients were monitored for safety including safety laboratory tests (including liver function tests and creatine kinase levels), physical examinations, assessment of adverse events and clinic assessments. Exceeding these safety parameters or having adverse events requiring medication discontinuation were identified as events meriting mandatory discontinuation. Other cases were identified as optional discontinuation. Treatment discontinuation was defined as being off medication for 30 days (Blazing et al., 2016). Following the physicians' identification, among the overall patients, 7; 255 patients discontinued treatment for optional reasons prior to clinical event (mean 25:3 months, SD 23:6), 131 patients discontinued treatment for mandatory reasons (mean 20:5 months, SD 19:4), and all remaining subjects completed assigned treatment prior to clinical event. Among the statin monotherapy group, 3; 681 patients discontinued treatment for optional reasons (mean 25:0 months, SD 23:9) and 66 patients discontinued treatment for mandatory reasons (mean 19:8 months, SD 21:6). Among the intensive statin–ezetimibe therapy group, 3; 574 patients discontinued treatment for optional reasons (mean 25:1 months, SD 23:7) and 65 patients discontinued treatment for mandatory reasons (mean 20:1 months, SD 20:5). Because of variability in the time to discontinuation and in patient characteristics, IMPROVE-IT is uniquely placed to investigate the legacy effect of statin drugs on the clinical outcome after discontinuation.

The questions we address were to evaluate the effect that discontinuation from statin drugs had on a composite clinical outcome (CV death, MI and stroke), and to determine whether there was a benefit to the use of statins that continued even after treatment was discontinued. We considered the dr-MSM as

$$\lambda_v(t) = \lambda_0(t) \exp\{\beta_1 z_{v1}(t) + \beta_2 z_{v2}(t)\}, \quad (23)$$

where $\lambda_0(t)$ is the population hazard rate if patients never had treatment withdrawn, $z_{v1}(t) = 1$ if $0 \le t - v < 6$ and $z_{v2}(t) = 1$ if $6 \le t - v$; if $v > t$, then $z_{v1}(t) = z_{v2}(t) = 0$. The parameters

β_j would describe the relative hazard if all patients that have stopped using statin for a period of time (< 6 months or > 6 months) as compared to if all patients had never stopped using statin. For model (23) we considered two cases:

a. No adjustment for any covariates;

b. Adjusting for baseline covariates.

To build a model for $K\{\upsilon, \bar{H}(\upsilon)\}$ in (7), we considered the baseline covariates X, including age, smoking status, and other 15 baseline health outcome measures. We first fitted a Cox proportional hazards model to the data including the baseline covariates, the squared terms and all the first-order interactions, with l_1 penalty. In fitting the model, the tuning parameter was estimated using 10-fold cross-validation. This resulted in 23 significant baseline terms. The final proportional hazards model included the 23 selected baseline terms and all time-dependent covariates $Q(\upsilon)$, including low-density lipoprotein cholesterol level, apolipoprotein B level, indicator of malignant symptoms, indicator for ever having adverse events by time υ with three levels: "null" if there is no averse events by time υ, "possible" if there is only "possible" events but no "likely" events by time υ, and "likely" if there is ever a "likely" event by time υ. The adverse events were categorized as "possible" and "likely" to be associated with off treatment optionally by the clinical expert on the study.

Table 2 reports the results from the proposed dr-MSM approach and the naive methods as described in the simulation section, separately for the control and active treatment groups. For standard errors and p-values, we consider the usual robust estimate output from the standard software, which treats the weight function as known. Although the effect sizes may be a little different between the standard analyses and the dr-MSM analyses, qualitatively they all suggest that discontinuation of treatment is significantly harmful with hazard rates after treatment discontinuation increasing by 1:21-fold to 2:18-fold as compared to having stayed on treatment, and there is no legacy effect of statin drugs. It is important to note that only the dr-MSM analysis is designed to address the well-formulated question for comparing different time-to-discontinuation regimes. However, we must add the caveat that our methods assume that all confounders are measured whereas, we cannot be sure that some patients may have discontinued treatment optionally based on prognostic factors that have not been captured in our data which if is the case can distort the causal interpretation. That being said, clinicians in the IMPROVE-IT study have tried their best to record the reasons for optional discontinuation at patients' follow-up visits during the course of the trial. Therefore, we believe that we have adjusted for important confounders in this study and tried our best to answer the causal question based on the data we have.

6. Discussion

In clinical trials, after proving the effectiveness of a treatment, researchers are often interested in the effect of the treatment discontinuation time on the outcome. The typical approach would use multi-state models for event history analysis (Frydman, 1992, 1995; Andersen and Keiding, 2002; Putter et al., 2007) and focus on the transition intensities from optional/mandatory treatment discontinuation to the clinical outcome. However, these transition intensities would only be associative. In contrast, we evaluate the causal effect of

optional treatment discontinuation time on the clinical outcome. Importantly, studying the effect of mandatory discontinuation on the outcome is arguably not interesting for the purpose of intervention. On the other hand, mandatory discontinuation is consistent with how a treatment regime must be administrated. The objective here is to develop an instructive demonstration of how careful conceptualization of our problem leads to unambiguous definition of a sensible treatment effect and to valid inferences, shaping a principled approach to deal with treatment discontinuation.

Extending our method to allow for informative censoring, using the idea of inverse probability of censoring weighting (Hernán et al., 2006), is an interesting subject for future investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Jennifer White for providing the IMPROVE-IT data used in this paper and her help and advice in interpreting the data. The IMPROVE-IT trial was sponsored by Merck & Co., Inc. The second author was sponsored by NIH grants: R01 HL118336 and P01 CA142538.

References

Aalen OO. A linear regression model for the analysis of life times. Stat. Med. 1989; 8:907–925. [PubMed: 2678347]

Andersen PK, Borgan O, Gill RD, Keiding N. Statistical Models based on Counting Processes. Springer-Verlag; New York: 1993.

Andersen PK, Keiding N. Multi-state models for event history analysis. Stat. Methods Med. Res. 2002; 11:91–115. [PubMed: 12040698]

Blazing MA, Giugliano RP, de Lemos JA, Cannon CP, Tonkin A, Ballantyne CM, Lewis BS, Musliner TA, Tershakovec AM, Lokhnygina Y, et al. On-treatment analysis of the improved reduction of outcomes: Vytorin efficacy international trial (improve-it). Am. Heart J. 2016; 182:89–96. [PubMed: 27914504]

Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. 1993; 88:9–25.

Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, Theroux P, Darius H, Lewis BS, Ophuis TO, Jukema JW, et al. Ezetimibe added to statin therapy after acute coronary syndromes. N. Engl. J. Med. 2015; 372:2387–2397. [PubMed: 26039521]

Cannon CP, Giugliano RP, Blazing MA, Harrington RA, Peterson JL, Sisk CM, Strony J, Musliner TA, McCabe CH, Veltri E, et al. Rationale and design of IMPROVE-IT (IMProved Reduction of Outcomes: Vytorin Efficacy International Trial): comparison of ezetimbe/simvastatin versus simvastatin monotherapy on cardiovascular outcomes in patients with acute coronary syndromes. Am. Heart J. 2008; 156:826–832. [PubMed: 19061694]

Cox DR. Regression models and life-tables. J. R. Stat. oc. Ser. B Stat. Methodol. 1972; 34:187–220.

Cox DR. Partial likelihood. Biometrika. 1975; 62:269–276.

Daniel R, Cousens S, De Stavola B, Kenward M, Sterne J. Methods for dealing with time-dependent confounding. Stat. Med. 2013; 32:1584–1618. [PubMed: 23208861]

Efron B, Tibshirani RJ. An introduction to the bootstrap. Boca Raton, FL: CRC press; 1994.

Frydman H. A nonparametric estimation procedure for a periodically observed three-state markov process, with application to AIDS. Journal of the Royal Statistical Society. Series B (Methodological). 1992; 54:853–866.

Frydman H. Semiparametric estimation in a three-state duration-dependent markov model from interval-censored observations with application to AIDS data. Biometrics. 1995; 51:502–511. [PubMed: 7662839]

Gran JM, Røysland K, Wolbers M, Didelez V, Sterne JA, Ledergerber B, Furrer H, von Wyl V, Aalen OO. A sequential cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. Stat. Med. 2010; 29:2757–2768. [PubMed: 20803557]

Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology. 2000; 11:561–570. [PubMed: 10955409]

Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. Basic Clin. Pharmacol. Toxicol. 2006; 98:237–242. [PubMed: 16611197]

Johnson BA, Tsiatis AA. Estimating mean response as a function of treatment duration in an observational study, where duration may be informatively censored. Biometrics. 2004; 60:315–323. [PubMed: 15180656]

Johnson BA, Tsiatis AA. Semiparametric inference in observational duration-response studies, with duration possibly right-censored. Biometrika. 2005; 92:605–618.

Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: main content. The International Journal of Biostatistics. 2010; 6 Article 8.

Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat. Med. 2007; 26:2389–2430. [PubMed: 17031868]

Robins J, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. Stat. Med. 2008; 27:4678–4721. [PubMed: 18646286]

Robins JM. Methodological Issues in AIDS Behavioral Research. Springer; New York: 2002. Analytic methods for estimating HIV-treatment and cofactor effects; 213–288.

Robins JM. Proceedings of the second seattle Symposium in Biostatistics. New York: Springer; 2004. Optimal structural nested models for optimal sequential decisions; 189–326.

Robins JM, Blevins D, Ritter G, Wulfsohn M. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology. 1992; 3:319–336. [PubMed: 1637895]

Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000; 11:550–560. [PubMed: 10955408]

Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 1974; 66:688–701.

Zhang M, Tsiatis AA, Davidian M, Pieper KS, Mahaffey KW. Inference on treatment effects from a randomized clinical trial in the presence of premature treatment discontinuation: the synergy trial. Biostatistics. 2010; 12:258–269. [PubMed: 20797983]

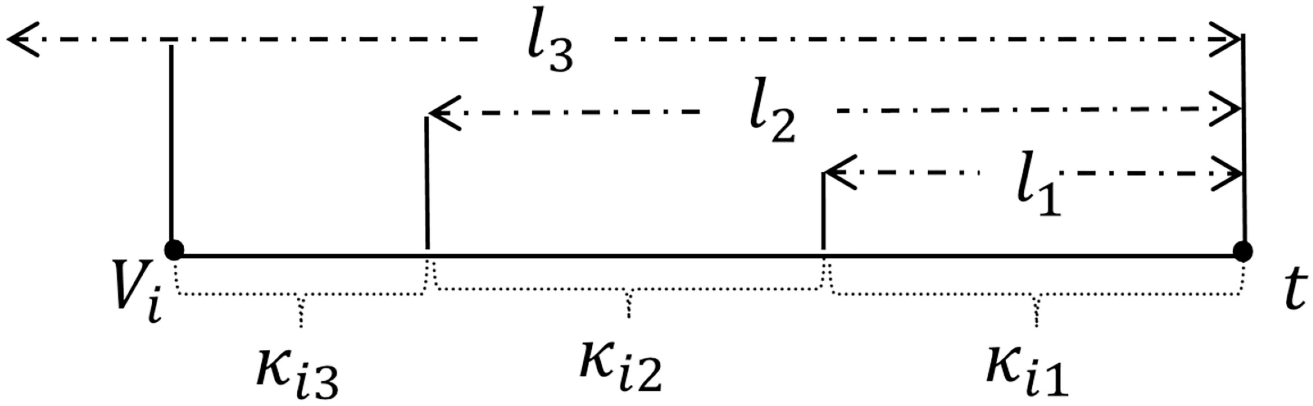

Figure 1.
An example for visualization of intervals κ_{ij}.

Table 1

Simulation results: mean and variance of estimates of β based on 2,000 simulated data sets: the true value of β is 0, SE1 is the standard error output from the standard software, that is, ignore the uncertainty associated with estimation of weights, SE2 is the bootstrap standard error with a bootstrap size of 100, CR1(2) is the coverage rate based on SE1(2).

	Mean Est.	SD	SE1	SE2	CR1	CR2
Naive I						
a. without covariates	0.263	0.081	0.080	0.081	0.091	0.102
b. with time-ind	0.219	0.082	0.082	0.081	0.229	0.251
c. plus time-dep	0.041	0.084	0.083	0.083	0.837	0.841
Naive II						
a. without covariates	0.332	0.088	0.088	0.088	0.034	0.041
b. with time-ind	0.276	0.089	0.089	0.089	0.123	0.133
c. plus time-dep	0.044	0.093	0.092	0.092	0.836	0.838
dr-MSM						
a. without covariates	0.004	0.081	0.081	0.078	0.949	0.950
b. with time-ind	0.004	0.080	0.080	0.078	0.947	0.948

Table 2

Results of Legacy Effect of Statin Treatment after Discontinuation: $\exp(\hat{\beta}_1)$ and $\exp(\hat{\beta}_2)$ are hazard ratios if patients that have stopped using statin for a period of time (<6 months and > 6 months, respectively) as compared to those still on statin drug.

	Statin monotherapy						**Statin-ezetimibe therapy**					
	$\exp(\hat{\beta}_1)$			**$\exp(\hat{\beta}_2)$**			**$\exp(\hat{\beta}_1)$**			**$\exp(\hat{\beta}_2)$**		
	Est	**S.E.**	***p***	**Est**	**S.E.**	***p***	**Est**	**S.E.**	***p***	**Est**	**S.E.**	***p***
Naive I												
a. without covariates	2.90	0.22	0.00	1.55	0.08	0.00	2.61	0.22	0.00	1.85	0.10	0.00
b. with time-ind	2.66	0.21	0.00	1.43	0.08	0.00	2.36	0.21	0.00	1.70	0.09	0.00
c. plus time-dep	2.41	0.18	0.00	1.28	0.07	0.00	2.11	0.19	0.00	1.53	0.09	0.00
Naive II												
a. without covariates	4.54	1.84	0.00	2.33	0.61	0.00	4.34	1.92	0.00	1.91	0.59	0.04
b. with time-ind	3.67	1.49	0.00	1.99	0.52	0.01	4.07	1.82	0.00	1.76	0.54	0.07
c. plus time-dep	2.80	1.15	0.01	1.50	0.41	0.14	2.93	1.30	0.02	1.20	0.38	0.55
dr-MSM												
a. without covariates	2.18	0.18	0.00	1.21	0.07	0.00	2.06	0.19	0.00	1.50	0.09	0.00
b. with covariates	2.19	0.18	0.00	1.25	0.07	0.00	2.04	0.19	0.00	1.53	0.09	0.00