Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–14

DOI: 10.1177/0962280216668554
smm.sagepub.com
SAGE

Comparison of statistical approaches dealing with time-dependent confounding in drug effectiveness studies

Mohammad Ehsanul Karim,[1,2] John Petkau,[3] Paul Gustafson,[3] Robert W Platt,[1,2,4,7] and Helen Tremlett[5]; The BeAMS Study Group[6]

Abstract

In longitudinal studies, if the time-dependent covariates are affected by the past treatment, time-dependent confounding may be present. For a time-to-event response, marginal structural Cox models are frequently used to deal with such confounding. To avoid some of the problems of fitting marginal structural Cox model, the sequential Cox approach has been suggested as an alternative. Although the estimation mechanisms are different, both approaches claim to estimate the causal effect of treatment by appropriately adjusting for time-dependent confounding. We carry out simulation studies to assess the suitability of the sequential Cox approach for analyzing time-to-event data in the presence of a time-dependent covariate that may or may not be a time-dependent confounder. Results from these simulations revealed that the sequential Cox approach is not as effective as marginal structural Cox model in addressing the time-dependent confounding. The sequential Cox approach was also found to be inadequate in the presence of a time-dependent covariate. We propose a modified version of the sequential Cox approach that correctly estimates the treatment effect in both of the above scenarios. All approaches are applied to investigate the impact of beta-interferon treatment in delaying disability progression in the British Columbia Multiple Sclerosis cohort (1995–2008).

Keywords

Bias (epidemiology), causality, confounding factors (epidemiology), epidemiologic methods, inverse probability weighting, longitudinal studies, models, survival analysis

1 Introduction

Longitudinal studies can include regular measurements of clinical symptoms and disease activity as covariates, and it is natural that the values may change over time. Since the predictive ability of baseline covariates may decrease over the follow-up time, consideration of the full history of these time-dependent covariates, rather than just the baseline covariates would be preferable.[1] However, if these covariates are affected by previous treatment and predicts the future treatment decision and future outcome conditional on the past treatment exposure, then such covariates are popularly known as "time-dependent confounders."[2,3] If the causal effect of treatment is of interest, the estimated hazard ratio may be biased whether or not the time-dependent confounders are included as covariates in a time-dependent Cox model analysis.[2,4] In the presence of time-dependent confounding, marginal structural Cox models (MSCM) are frequently used to estimate the causal effect of a time-dependent treatment

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada
[2]Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, Montreal, Canada
[3]Department of Statistics, University of British Columbia, Vancouver, Canada
[4]Department of Pediatrics, McGill University, Montreal, Canada
[5]Department of Medicine, Division of Neurology and Centre for Brain Health, University of British Columbia, Vancouver, Canada
[6]"The BeAMS Study, Long-term Benefits and Adverse Effects of Beta-Interferon for Multiple Sclerosis": A Shirani, Y Zhao, C Evans, E Kingwell, ML van der Kop and J Oger
[7]Research Institute of the McGill University Health Centre, Montreal, Canada

Corresponding author:
Mohammad E Karim, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Purvis Hall, 1020 Pine Avenue West, Montreal, Quebec H3A1A2, Canada.
Email: ehsan.karim@mail.mcgill.ca

exposure.[5,6] Sometimes MSCM estimates are unstable due to use of variable inverse probability of treatment weights (IPTW).[7–9] The sequential Cox approach has been proposed as an alternative to the MSCM approach.[9]

In a previous simulation study, we showed that a simplified implementation of the sequential Cox approach (considering only baseline covariate adjustment) performed well compared to a Cox proportional hazards model fit with time-dependent exposure while adjusting for baseline confounders.[10] To the best of our knowledge, no attempt has been made to explore the appropriateness of the sequential Cox approach in adjusting for the effect of time-dependent confounding in a simulation setting.

To overcome the limitations of this approach revealed by our simulations, we also propose a modified version of the sequential Cox approach in this paper. The primary focus of this paper is to assess the performance of these sequential Cox approaches for dealing with a time-dependent confounder. A secondary aim of this paper is to examine how these methods perform in the presence of a time-dependent covariate which does not interact with the past treatment condition (i.e. is not a "time-dependent confounder"). To do this, we simulate survival data with time-dependent treatment exposure. Two different conditions are considered for simulation: (1) a time-dependent confounder is present, and (2) a time-dependent covariate is present along with a baseline covariate. To assess their suitability in an application, we apply these methods to investigate the impact of time-varying beta-interferon treatment in delaying disability progression in subjects from the British Columbia (BC) Multiple Sclerosis (MS) database (1995–2008).[11,12]

The remainder of the paper is organized as follows. In the next section, we describe the notation and design of the simulation study, the methods used to address time-dependent confounding, and the metrics used to evaluate their performances. Then we summarize the simulation and the MS data analysis results. The paper concludes with a discussion of the results, and the implications and limitations of the current study.

2 Methods

2.1 Notation

Consider a hypothetical longitudinal study consisting of n subjects ($i = 1, 2, \ldots, n$). Let $t_0 = 0$ be the start of follow-up or the time of the baseline clinic visit. Baseline covariates L_0 (binary or continuous) are recorded at baseline. Follow-up continues until the time of failure T or the time of censoring T^C. Let that $[t_m, t_{m+1})$ constitutes the mth interval (say, mth month in the follow-up). At intervals $m = 0, 1, 2, \ldots, K$, regular measurements of the binary treatment status A_m (= 1 for treated and 0 otherwise) are recorded. Let C_m be the binary indicator of censoring (= 1 if censored due to dropout or artificial censoring and 0 otherwise). Let $\bar{a}_m = (a_0, a_1, \ldots, a_m)$ be the observed realizations of the treatment history $\bar{A}_m$ up to interval m, and similarly, let $\bar{l}_m$ and $\bar{c}_m$ be the observed realizations of the covariate history $\bar{L}_m$ and the censoring history $\bar{C}_m$ up to interval m, respectively. The binary indicator of failure by time t_{m+1} is defined as $Y_{m+1} = I(T \leq t_{m+1})$. As the sequential Cox approach does not allow for treatment discontinuation, we assume that the subjects may initiate treatment at most once and that they continue on the treatment thereafter until the end of their follow-up. Let treatment initiation occur at time T^A.

2.2 Analysis approaches

Brief characteristics of the analysis approaches are shown in Table 1. We describe these methods in detail in the following sections using the notation defined above.

2.2.1 Cox model with time-dependent treatment and covariates

In the presence of baseline confounders L_0 and time-dependent covariates L_m, one way to express the hazard function through the time-dependent Cox model is as follows

$$\lambda(m|L_0, L_m) = \lambda_0(m) \exp(\tilde{\psi}_1 A_m + \psi_2 L_0 + \psi_3 L_m) \tag{1}$$

where m is the visit index, $\lambda_0(m)$ is the unspecified baseline hazard function, $\tilde{\psi}_1$ is the log-hazard ratio (log-HR) of the time-dependent treatment status (A_m), and ψ_2 and ψ_3 are the vectors of log-HRs for the baseline covariates L_0 and the time-dependent covariates L_m, respectively.

2.2.2 MSCMs

If the time-dependent covariate L_m is influenced by past exposure, i.e. if L_m is a time-dependent confounder, playing a dual role as a confounder and an intermediate variable in the causal pathway between treatment and

Table 1. Description of the Cox models used in the approaches under consideration.

Approach	Stratified	Time-dependent covariate history	Weight adjusted
TD-Cox	No	Full	No
MSCM	No	Full	Yes, IPTC[a]
Sequential Cox[b]	Yes	Up to the new baseline[c]	Yes, IPC[d]
Modified sequential Cox[b]	Yes	New baseline and afterwards[e]	Yes, IPC[d]

IPC: inverse probability of censoring; IPT: inverse probability of treatment; IPTC: inverse probability of treatment and censoring; MSCM: marginal structural Cox model; TD-Cox: Cox model with time-dependent exposure.

[a]Pooled logistic regression is used to estimate the IPTC weights.

[b]Robust (sandwich) estimate is used to obtain SEs.

[c]For the sequential Cox approach, covariate values are collected at three time points for each mini-trial: at baseline, at the interval of treatment start, and at the previous interval (the lagged value): $\tilde{L}_m = (L_0, L_{m-1}, L_m)$. Here, time-fixed covariates collected at the original baseline (i.e. L_0) are included in the analysis.

[d]Aalen's additive regression model is used to estimate the IPCW.

[e]For the modified sequential Cox approach, the time-dependent covariate values are collected at the new baseline and then at subsequent intervals (i.e. $\vec{L}_m = (L_m, L_{m+1}, \ldots, L_K)$). Time-fixed covariates collected at the original baseline (i.e. L_0) are also included in the analysis.

outcome, then ψ_1 as estimated from equation (1) may be biased[2] (i.e. may deviate from the target parameter ψ_1; discussed in detail in Web Appendix §A). Instead of using L_m as a covariate, the MSCM approach uses it to calculate IPTW that are person-time-specific measures of the degree to which L_m confounds the treatment selection process.

Stabilized inverse probability of treatment and censoring (IPTC) weight, sw_m, can be obtained by multiplying stabilized IPTW, sw_m^T, by stabilized inverse probability of censoring weights (IPCW), sw_m^C,[5] where

$$sw_m^T = \prod_{j=0}^{m} \frac{pr(A_j = a_j \,|\bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0)}{pr(A_j = a_j \,|\bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0, \bar{L}_j = \bar{l}_j)} \tag{2}$$

and

$$sw_m^C = \prod_{j=0}^{m} \frac{pr(C_j = 0|\bar{C}_{j-1} = 0, \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0)}{pr(C_j = 0|\bar{C}_{j-1} = 0, \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0, \bar{L}_{j-1} = \bar{l}_{j-1})} \tag{3}$$

The weights sw_m are used in the time-dependent Cox model with hazard function modeled as follows to weight the contribution of each person-time observation so that the confounding due to L_m is removed

$$\lambda(m|L_0) = \lambda_0(m) \exp(\psi_1 A_m + \psi_2 L_0) \tag{4}$$

where ψ_1 is log-HR of the time-dependent treatment status (A_m). Note that IPCW is used only if nonrandom censoring is present. When the numerators in equations (2) and (3) are replaced by 1, these become the unstabilized IPTC weights, w_m. We used pooled logistic regression[2,5] to estimate the IPTC weights. Estimation procedure details are included in Web Appendix §B.

2.2.3 Sequential Cox approach

Suppose that at least one subject initiates treatment in the mth interval $[t_m, t_{m+1})$. We want to mimic a randomized clinical trial for each such interval. The mini-trial corresponding to the mth interval (hereafter referred to as the mth mini-trial) involves only subjects who have not previously received any treatment. Among the subjects at-risk at t_m the subjects initiating treatment during the interval ($t_m < T^A \leq t_{m+1}$) are considered as the treated group, while the remaining subjects are considered as the control group. These control subjects are artificially censored at their times of later treatment initiation ($T^A > t_{m+1}$) to avoid confounding due to treatment. As these subjects are artificially censored, the analysis must be adjusted using IPCW.

In the analysis, we adjust for the baseline confounders L_0 measured at inclusion or baseline, the time-dependent covariates L_m measured at the start of the interval when patient started the treatment and the lagged covariates L_{m-1} consisting of the lagged value measured at the previous interval of the treatment start. Adjustment of these covariate values should help to reduce bias in the estimation of the treatment effect from the mth mini-trial data.[9]

Let us denote $\tilde{L}_m = (L_0, L_{m-1}, L_m)$. Here, L_0 includes the covariate values measured at baseline (i.e. time-static baseline covariates as well as the time-varying covariate or confounder value at time point $m = 0$).

We assume that the different mini-trials may have different baseline hazard functions but all subjects in the same mini-trial have the same baseline hazard function. Under this assumption, use of a stratified Cox model is appropriate. Therefore, one way to model the hazard function for the mth mini-trial is[9]

$$\lambda^m(j|L_0, \bar{L}_j) = \lambda_{0m}(j)\exp(\psi_1'' A_j + \psi_2'' \tilde{L}_m); \quad j \geq m \tag{5}$$

where $\lambda_{0m}(j)$ is the unspecified baseline hazard function for stratum m, ψ_1'' is log-HR of the time-dependent treatment status, and ψ_2'' is the vector of log-HRs for the time-dependent covariates $\tilde{L}_m$. This hazard function should be weighted by IPCW (equation (3)). It was suggested that the resulting estimate should bear a causal interpretation under the assumptions of no unmeasured confounders and correct model specification for the hazard ratio and the censoring weights.[9] We used Aalen's additive regression model[9,13] to estimate the IPCW.

We can fit a stratified Cox model to the combined data of all mini-trials (pseudo-data), stratified by the treatment initiation time. Inclusion of the same subject more than once invalidates the SE obtained from the stratified weighted Cox analysis. Computationally demanding resampling methods are suggested[9,13] to obtain a correct SE. We used a robust (sandwich) estimate instead to save computational time similar to other simulation studies.[14,15] An illustrative data construction example is provided in Web Appendix §C and the corresponding software implementation details are provided in Web Appendix §D.

2.2.4 Modified sequential Cox approach

As each of the mini-trials mimics a clinical trial, we propose to analyze the mini-trial data accordingly. Each mini-trial is created based on a particular month of treatment initiation. These treatment initiation months are considered as the new baselines (new time-0) for the corresponding mini-trials. Note that the time-fixed covariates measured at the original baseline (i.e. L_0) are included in the analysis. For covariates that vary over time, we consider all the information from the new baseline to the study endpoint to analyze the data, as we would do in a clinical trial setting. For example, time-varying covariate or confounder values collected at the interval of treatment start and onward (i.e. $\vec{L}_m = (L_m, L_{m+1}, \ldots, L_K)$ for the mth mini-trial) are included in the model for adjustment.[16] Unlike the original proposal,[9] we do not use any time-dependent confounder (or time-dependent covariate) values prior to the new baseline for adjustment. Therefore, the hazard function for the mth mini-trial can be expressed as

$$\lambda^m(j|L_0, \bar{L}_j) = \lambda_{0m}(j)\exp(\psi_1' A_j + \psi_2' \vec{L}_m); \quad j \geq m \tag{6}$$

where ψ_1' is log-HR of the time-dependent treatment status and ψ_2' is the vector of log-HRs for the time-dependent covariates $\vec{L}_m$. However, the process of creating the pseudo-population remains the same. The proposed changes occur only in the analysis stage. Theoretically proving the equivalence of the target parameters ψ_1' from equation (6) and ψ_1 from equation (4) is not easy, but we have provided a heuristic justification of such equivalence later based on a Monte Carlo experiment. In some context, additional techniques, such as matching[17,18] or use of propensity scores[19] or adherence adjustments,[16] may be useful to make the subjects within a mini-trial more comparable, but those approaches are not considered here.

2.3 Design of simulation

We adopt the data generation process of Young et al.[20] to simulate survival times where time-dependent confounding is present. To simulate survival times with time-dependent covariates (none of which are time-dependent confounders), we adapt the permutation algorithm.[21] Descriptions of these algorithms are presented in Web Appendices §E and §F, respectively.

2.4 Simulation specifications

In our Monte Carlo study, we generated $N = 1000$ datasets with $n = 2500$ subjects, each followed for up to $m = 10$ subsequent monthly visits for each setting under consideration. We set $\lambda_0 = 0.01$ (on a monthly scale) to represent a rare disease condition and $\lambda_0 = 0.10$ (on a monthly scale) for a more frequent disease condition. We discuss a

Table 2. Two simulation settings under consideration.

	Simulation I	Simulation II
Algorithm	Young et al.[20]	Abrahamowicz et al.[21]
Time-varying treatment	Yes	Yes
Baseline covariate	No	Yes
Time-varying covariate	No	Yes
Time-varying confounder	Yes	No

brief description of the two simulations under consideration in Table 2. Below we provide the detailed specifications of the simulation scenarios.

2.4.1 Simulation I

In our implementation of the algorithm,[20] counterfactual failure time $T_{i\bar{0}}$ s are sampled from an exponential distribution, with constant λ_0 rate of monthly events throughout the follow-up. The binary time-dependent confounder, L_m, is modeled by the following covariates: a binary covariate $I(T_{\bar{0}} \leq c)$, previous treatment status A_{m-1}, and the lagged variable L_{m-1}

$$\begin{aligned} \text{logit}(p_L) &= \text{logit } Pr(L_m = 1 | A_{m-1}, L_{m-1}, Y_m = 0; \boldsymbol{\beta}) \\ &= \beta_0 + \beta_1 I(T_{\bar{0}} < c) + \beta_2 A_{m-1} + \beta_3 L_{m-1} \end{aligned} \tag{7}$$

with associated parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3) = (\log(3/7),\ 2,\ \log(1/2),\ \log(3/2))$, $c = 30$, and $Y_m = I(T \leq t_m)$. Here, the time-dependent covariate L_m is moderately affected by prior treatment A_{m-1} ($\beta_2 = \log(1/2) = -0.3$).

We model binary treatment status at each stage A_m with the factors current symptom L_m, past symptom L_{m-1}, and previous treatment status A_{m-1} as

$$\begin{aligned} \text{logit}(p_A) &= \text{logit } Pr(A_m = 1 | L_m, A_{m-1}, L_{m-1}, Y_m = 0; \alpha) \\ &= \alpha_0 + \alpha_1 L_m + \alpha_2 L_{m-1} + \alpha_3 A_{m-1} \end{aligned} \tag{8}$$

with associated parameters $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (\log(2/7),\ 1/2,\ 1/2,\ 10)$. Current treatment status A_m is made heavily dependent on the previous treatment status A_{m-1} by setting a high parameter value ($\alpha_3 = 10$). That way, we emulate the situation where subjects switch to treatment at most once and keep on using the treatment without much interruption or discontinuation. The true causal effect parameter (i.e. treatment effect) is set to be $\psi_1 = 0.5$ (in equation (4)).

Equations (7) and (8) define L_m as a time-dependent confounder affected by prior treatment.[2] In particular, past treatment exposure status A_{m-1} affects the time-dependent confounder L_m, which then predicts future treatment exposure A_m. L_m is also associated with the future failure status Y_{m+1} via $I(T_0 \leq c)$. Here, T_0 is the untreated counterfactual survival time and c is an arbitrary cut point used to generate the binary variable $I(T_0 \leq c)$. The value of c affects the degree of variability in the indicator variable $I(T_0 \leq c)$.[22] Without $I(T_0 \leq c)$, there would not be any confounding in the exposure–outcome relationship. The confounding here arises via the path: $Y_{m+1} \leftarrow I(T_0 \leq c) \rightarrow L_m \rightarrow A_m$. This indicator variable $I(T_0 \leq c)$ therefore dictates the degree to which T_0 affects L_m for a chosen value of c.

2.4.2 Simulation II

We assume an exponential distribution for generating failure times T with constant $\lambda_0 = 0.01$ rate of monthly events throughout the follow-up. A uniform distribution $U(1, 60)$ months is assumed to generate censoring times T^C, i.e. administrative censoring is set at five years of follow-up. Treatment initiation time T^A is generated from a uniform distribution $U(0, 10)$ (in months). Additionally, we consider sex as a baseline confounder in these data. A subject's sex is generated based on a Bernoulli distribution where the probability of being male is 0.3. We also add one time-dependent confounder L_m, which could represent cumulative disease activity, for example, such that higher cumulative disease activity has a higher risk (a log-HR of $\psi_3 = \log(1.5)$). This time-dependent confounder L_m is generated based on a Bernoulli distribution where the probability of disease activity increment is 0.75, accumulating the disease activity over at most $m = 10$ periods of time.

Table 3. Characteristics of the simulation settings under consideration.

Rates	Simulation I	Simulation II
Failure	0.143	0.084
Always treated	0.261	0.051
Never treated	0.046	0.150
Partially treated	0.692	0.799
Discontinuation	0.001[a]	–
Mean visits	9.367	8.943

[a]Simulation I allows a few exceptions (19 out of 25,000) where there are discontinuations. However, the proportion of discontinuation in the simulation I dataset is negligible (0.00076) and we do not expect any noticeable impact in the results due to this small number of exceptions.

The permutation algorithm[21] is used to generate survival data where binary treatment A_m is time dependent but the confounder L_0 is fixed at its baseline value. Arbitrarily, the effect parameters for treatment and sex on the survival outcome are set such that the treatment has a harmful effect (a log-HR of $\psi_1 = 0.5$) and males are at a lower risk than females (a log-HR of $\psi_2 = -0.7$).

2.5 Performance metrics

We assessed the performance of the various approaches by the following measures:

- Bias $= \sum_{i=1}^{N} (\hat{\psi}_{1i} - \psi_1)/N$: The average difference between the true and $N = 1000$ estimated parameters (log-HR);
- SD $= \sqrt{\sum_{i=1}^{N} (\hat{\psi}_{1i} - \bar{\psi}_1)^2/(N-1)}$ where $\bar{\psi}_1 = \sum_{i=1}^{N} \hat{\psi}_{1i}/N$;
- Model-based SE: The average of $N = 1000$ estimated standard errors of the estimated causal effect;
- Coverage probabilities of model-based nominal 95% CIs: Proportion of $N = 1000$ datasets in which the true parameter is contained in the nominal 95% CI.

The above quantities were defined in terms of MSCM parameter ψ_1 from equation (4). In order to define the same measures from other analysis approaches under consideration, we assumed that the true target parameters (defined in equations (1), (4), and (6), respectively) are the same for all approaches, e.g. $\tilde{\psi}_1 = \psi_1 = \psi_1' = 0.5$. A heuristic justification of such assumption of equivalence is provided in Section 3.3.

3 Simulation results

3.1 Description of the simulated data

To describe the data obtained from the simulation settings under the rare event condition, we generated datasets with a larger number of subjects (25,000) with up to 10 subsequent visits from each simulation algorithm. The characteristics of the treated, untreated, and partially treated groups; their failure rates; and average number of visits are listed in Table 3.

3.2 Rare event condition

We present the results from the rare event condition ($\lambda_0 = 0.01$ in a monthly timescale) in the two simulation settings.

3.2.1 Results from simulation I

Results from simulation I are reported in Table 4. MSCM with treatment status (A_m) is fitted to validate the data-generating algorithm. The corresponding stabilized weights are generated based on the relationship between treatment status (A_m) and the time-dependent confounder L_m. The level of bias is negligible compared to other approaches under consideration and the average coverage probability of the model-based nominal 95% CIs is 0.942. These results are now considered as the ideal for comparison purposes for this simulation setting. The time-dependent Cox models provide biased estimates (assuming $\tilde{\psi}_1 = \psi_1 = 0.5$) in the presence of this time-dependent confounder. The average coverage probability of the model-based nominal 95% CIs for the method is very low.

Table 4. Comparison of the analytical approaches to adjust for time-dependent confounding from simulation I (one time-dependent confounder and time-dependent treatment exposure) of 1000 datasets, each containing 2500 subjects followed for up to 10 time intervals.

Approach	Bias	*SD*	*se*	Coverage probability
TD-Cox[a]	0.438	0.168	0.169	0.251
Sequential Cox[b,c]	0.451	0.280	0.267	0.636
Modified Sequential Cox[d,e]	0.015	0.221	0.222	0.956
MSCM[f]	0.029	0.201	0.205	0.942

MSCM: marginal structural Cox model; TD-Cox: Cox model with time-dependent exposure.
[a]Includes the time-dependent confounder L_m as a covariate. In the presence of a time-dependent confounder, the time-dependent Cox model is not appropriate but the results are retained for comparison purposes.
[b]Adjusts for $\tilde{L}_m$.
[c]For the stabilized IPCWs, the numerator model adjusts for A_m, while the denominator model adjusts for A_m and $\tilde{L}_m$ via Aalen's additive regression.
[d]Adjusts for lagged values of A_m, the time-dependent confounder $\vec{L}_m$, and lagged values of $\vec{L}_m$. Note that, baseline covariates are not present in this setting.
[e]For the stabilized IPCWs, the numerator model adjusts for A_m, while the denominator model adjusts for A_m, $\vec{L}_m$, and lagged values of $\vec{L}_m$ via Aalen's additive regression.
[f]The stabilized IPTW numerator model adjusts for time index and lagged values of A_m, while the denominator model additionally adjusts for current and lagged values of L_m to predict future treatment status via pooled logistic models.

The bias of sequential Cox analysis is comparable to that with the time-dependent Cox analysis, while the corresponding effect estimates are considerably more variable. The properties of the modified sequential Cox approach, on the other hand, are comparable to that of MSCM. The corresponding average coverage probability of the model-based nominal 95% CIs for the method is comparable to that with MSCM.

3.2.2 Results from simulation II

Results from simulation II are reported in Table 5. The time-dependent Cox model with treatment status (A_m), baseline covariate (L_0), and time-dependent covariate (L_m) is fitted to validate the data-generating permutation algorithm. The level of bias is negligible and the average coverage probability of the model-based nominal 95% CIs is 0.952. These results are considered as the ideal for comparison purposes for this simulation setting. When the sequential Cox approach is used in this simulation setting, we observe some bias. We apply MSCM with L_m treated as a time-dependent confounder, even though L_m is only a time-dependent covariate that is not affected by the past treatment. The corresponding bias is negligible and the average coverage probability of the model-based nominal 95% CIs is 0.952. As L_m is not a time-dependent confounder in this simulation, the similarity between the estimates obtained from MSCM and the time-dependent Cox model is not surprising. The properties of the modified sequential Cox approach are again comparable to MSCM with reasonable average coverage probability of the model-based nominal 95% CIs.

3.3 Marginal versus conditional interpretations

We need to take into account the different interpretations of the target quantities ($\tilde{\psi}_1$, ψ_1, and ψ_1' from equations (1), (4), and (6), respectively) being estimated by the three approaches under consideration. The clinical context should dictate whether the target parameter should bear a marginal or conditional interpretation. Conditional interpretations are generally useful in deciding personalized drug choices, whereas marginal interpretations may be more useful in making generalized policy decisions for a heterogeneous group of patients.[15]

The modified sequential Cox approach and its proposed modification emulate a sequence of conditionally randomized treatment assignments. This is done by first reorganizing the observed data and then stratifying the combined data based on the month of treatment initiation, conditioning on the pretreatment covariate values. The estimated ψ_1' from a modified sequential Cox approach (as well as estimated ψ_1'' from a sequential Cox approach), therefore, bears a conditional interpretation,[9,13] as does the estimated log-hazard ratio $\tilde{\psi}_1$ from a time-dependent Cox model.[23,24]

In contrast, MSCM estimates the log-hazard ratio ψ_1 between two counterfactual scenarios: all subjects are treated at a given time versus none of the same subjects are treated at that time. The target quantity of interest ψ_1 estimated from MSCM is the causal effect of the treatment. Assuming the MSCM assumptions hold, this quantity

Table 5. Comparison of the analytical approaches to adjust for time-dependent covariate from simulation II (one baseline covariate, one time-dependent covariate, and time-dependent treatment exposure) of 1000 datasets, each containing 2500 subjects followed for up to 10 time intervals.

Approach	Bias	*SD*	*se*	Coverage probability
TD-Cox[a]	0.000	0.164	0.162	0.952
Sequential Cox[b,c]	0.271	0.189	0.184	0.688
Modified Sequential Cox[d,e]	−0.022	0.231	0.234	0.961
MSCM[f,g]	−0.001	0.163	0.162	0.952

MSCM: marginal structural Cox model; TD-Cox: Cox model with time-dependent exposure.
[a]The baseline covariate L_0 and time-dependent covariate L_m are included.
[b]Adjusts for L_0 and $\tilde{L}_m$.
[c]In the stabilized IPCW model, the numerator model adjusts for A_m and L_0, while the denominator model adjusts for A_m, L_0, and $\tilde{L}_m$ via Aalen's additive model.
[d]Adjusts for baseline covariates L_0, lagged values of A_m, the time-dependent confounder $\vec{L}_m$, and lagged values of $\vec{L}_m$.
[e]For the stabilized IPCWs, the numerator model adjusts for A_m and baseline variable L_0, while the denominator model adjusts for L_0, A_m, $\vec{L}_m$, and lagged values of $\vec{L}_m$ via Aalen's additive regression.
[f]Adjusts for only L_0.
[g]For the stabilized IPTWs, the numerator model adjusts for the time index, L_0, and lagged values of A_m, while the denominator model additionally adjusts for current and lagged values of L_m to predict future treatment status via pooled logistic models.

Table 6. Series of Monte Carlo studies for the simulation I setting, each simulation with larger cohorts.

Approach	Bias (SD) from N = 100 cohorts			
Cohort size (*n*)	n = 5000	n = 10,000	n = 50,000	n = 100,000
MSCM	0.013 (0.150)	0.007 (0.107)	0.011 (0.049)	0.003 (0.033)
Modified sequential Cox	0.011 (0.157)	0.003 (0.100)	0.004 (0.048)	-0.001 (0.037)

MSCM: marginal structural Cox model.

should match the estimated treatment effect from a randomized clinical trial.[25,26] In that sense, the MSCM approach mimics randomized clinical trial data setting by appropriately weighting observational data. As the corresponding outcome or hazard model does not condition on any time-dependent covariates that affect future treatment, the log-hazard ratio estimated from a MSCM is a marginal or population-averaged quantity.[27] It is possible to extend the MSCM by incorporating baseline covariates in the hazard model.[28] When we use stabilized weights in a MSCM, the treatment effect is marginal with respect to the time-dependent confounders, but conditional with respect to the baseline covariates.[11,29,30]

Comparing the MSCM target parameter estimate with any conditional treatment effect estimate (from time-dependent Cox model approach: which is considered as the standard for comparison in simulation II) is not straightforward when noncollapsible measures, such as hazard ratio (HR) or odds ratio (OR), are employed.[30–34] Establishing the equivalence of the target parameters (log-hazard ratios $\tilde{\psi}_1$, ψ_1, and ψ_1' from equations (1), (4), and (6), respectively) from the approaches under consideration may not be easy. However, the difference between the conditional and marginal parameters is expected to be negligible when the event rate in the time intervals under consideration is small.[14,33] In our simulations, the event rate was 1% in each month interval. In this scenario, noncollapsibility of the HRs should not have any noticeable impact on the findings. We performed the following numerical experiment to support this proposition.[14,15] We generated a very large cohort ($n = 100{,}000$) under the simulation II settings. As expected, the time-dependent Cox model and MSCM yield almost identical values of the treatment effect estimates (log-hazard ratio of $\hat{\tilde{\psi}}_1 = 0.5084$ versus $\hat{\psi}_1 = 0.5091$). This further provides some evidence that noncollapsibility of the HR does not affect our simulation II findings.

Similarly, we performed another Monte Carlo experiment for the simulation I setting with $N = 100$ larger cohorts ($n = 100{,}000$) to investigate whether ψ_1' and ψ_1 quantities estimated via the modified sequential Cox and MSCM, respectively (which is considered as the standard for comparison in simulation I) differed systematically in the settings we investigated. Both approaches produce very similar values of the treatment effect estimates on average (see Table 6). For this setting, this shows that the target parameters for these two approaches are not materially different. To check the adequacy of the sample size $n = 2500$ chosen in our original

Table 7. Summary of the estimated parameters from the multiple sclerosis (MS) patients' data from British Columbia, Canada (1995–2008).

Approach	$\widehat{HR}$	$se(\widehat{HR})$	95% CI	Weights	
				Average (SD)	Range
TD-Cox[a]	1.29	0.23	0.91–1.82		
Sequential Cox[b,c]	1.23	0.32	0.74–2.07	1.00 (0.01)	0.74–1.63
Modified sequential Cox[d,e]	1.36	0.26	0.93–1.99	1.00 (0.01)	0.92–1.24
MSCM[f,g]	1.31	0.23	0.92–1.84	1.00 (0.06)	0.37–1.60

MSCM: marginal structural Cox model; TD-Cox: Cox model with time-dependent exposure.
[a]Adjusts for baseline covariates L_0 (sex, EDSS score, age, and disease duration), and for the time-dependent confounder L_m "cumulative relapses."
[b]Adjusts for L_0, A_m, and $\bar{L}_m$.
[c]The stabilized IPCW numerator model adjusts for A_m and L_0, while the denominator model additionally adjusts for L_m and lagged values of L_m via Aalen's additive model.
[d]Adjusts for baseline covariates L_0, lagged values of A_m, the time-dependent confounder $\vec{L}_m$, and lagged values of $\vec{L}_m$.
[e]For the stabilized IPCWs, the numerator model adjusts for A_m and baseline variable L_0, while the denominator model adjusts for L_0, A_m, $\vec{L}_m$, and lagged values of $\vec{L}_m$ via Aalen's additive regression.
[f]Adjusts for the potential baseline confounders L_0.
[g]The stabilized IPTW numerator model adjusts for a restricted cubic spline of the follow-up time index, baseline confounders L_0, and lagged values of A_m to predict future treatment status. The denominator model additionally adjusts for the current and lagged values of cumulative relapses (L_m) via the pooled logistic models.

simulation I, this study is also repeated for other cohort sizes $n = 5000$, 10,000, and 50,000 (see Table 6). The results look very similar, and expectedly the SDs are decreasing with increment of sample sizes.

3.4 When more events are available

The trends in the bias from the more frequent event condition ($\lambda_0 = 0.1$ in a monthly timescale) are similar compared to those in the rare event condition (see Web Tables H.1 and H.2). As expected, the standard errors are much less than in the corresponding analyses when failure rates are rare. Bias is slightly lower in some cases. One noticeable difference is observed in simulation setting I: in the presence of the time-dependent confounder, when the failure rate is more frequent, the bias of the time-dependent Cox and MSCM approaches is reduced to minimal levels, whereas considerable bias is still apparent with the sequential Cox approach. On the other hand, the average coverage probability of the model-based nominal 95% CIs from the time-dependent Cox approach is smaller than that of MSCM. The modified sequential Cox approach estimates are still associated with good statistical properties.

4 Application in MS

We apply these methodologies to the BC MS cohort data (1995–2008).[11] The dataset was used in previous studies[11,12,35–38] to estimate the effect of β-IFN on time to irreversible disability outcomes. As before, irreversible progression of disability is measured by sustained expanded disability status scale (EDSS) 6 which is confirmed after at least 150 days, with all subsequent EDSS scores being 6 or greater. Web Appendix §G describes the baseline characteristics, eligibility, and exclusion criteria of the MS cohort.

Potential baseline confounders L_0 include age, sex, disease duration, and EDSS score. Also, we consider the cumulative number of relapses in the previous two years (hereafter called "cumulative relapses") as a time-dependent confounder L_m.[11] Once the subjects initiate β-IFN, we assume they continue taking the drug without any discontinuation until they develop the outcome or are censored, as is assumed in our simulations and previous pharmacoepidemiologic studies.[9,13,39] As found in the previous study[11] using this cohort, we consider the MSCM estimates to be ideal in this time-dependent confounding context. Results are reported in Table 7.

The IPCWs in the modified sequential Cox approach are less variable than the IPWs in MSCM. IPCWs are estimated separately for each mini-trial.[9] When they are estimated from the aggregated dataset instead,[8] or when IPCWs were estimated via pooled logistic regression models,[2,5,39] the HR estimates are very similar (see Web Appendix §I). No matter how they are constructed, the IPCWs from the mini-trials are well behaved, i.e. the

averages are close to one and they have low variability (most are within the range of 0.9–1.1 and the distributions are unimodal and symmetric; see Web-Figures I.1 and I.2).

5 Discussion

In observational studies, the estimation of a treatment effect is challenging in the absence of randomization. In longitudinal studies, additional complexity arises in the presence of time-dependent confounding. MSCMs are popularly used to deal with this problem in the survival analysis setting. MSCMs handle time-dependent confounding by reweighting the data in such a way that the confounding effect of the time-dependent confounder is removed. Then adjusting for the baseline confounders (but not the time-dependent confounder) in the reweighted pseudo-population is adequate to obtain the counterfactual or causal effect of the treatment under the identifiability conditions.[8,28] Sometimes, the MSCM estimates may be unstable due to use of IPTWs and an alternate analysis or view of the data may be helpful.

The sequential Cox approach was proposed as an alternative method to the MSCMs for estimation of the treatment effect from complex observational data settings where the treatment is timedependent and censoring may be nonrandom.[9] This approach restructures the data in such a way that a sequence of subsets of data (mini-trials) are created based on intervals of treatment initiation. Aggregation of all the mini-trial data produces the pseudo-population. In this pseudo-population, subjects initiating treatment at each interval are compared to those who do not initiate treatment, conditional on covariates at respective intervals as well as baseline covariates. Although IPTWs are avoided in the sequential Cox approach, IPCWs are still required. These weights are less variable and more stable than IPTWs[7,9] and appropriately handle the artificial censoring at later treatment start dates. We proposed a modified version of this approach that deal with analyses differently than the original proposal.

The treatment effect that is estimated from a MSCM is a marginal estimate as it is obtained by averaging over subjects with different hazards. This estimate does not condition on the time-dependent confounder. Rather the time-dependent confounder plays a role in creating weights for the MSCM model fitting. These weights are used to create the pseudo-data which are free from time-dependent confounding and mimic a clinical trial situation. Similarly, the sequential Cox approach contrasts subjects within mini-trials, where each mini-trial includes subjects who did and did not initiate the treatment. Controlling for current values of the time-dependent covariates as well as baseline confounders should make the treated subjects conditionally exchangeable with control subjects at the time of treatment initiation. Within each mini-trial, treatment assignment could be considered as random among the comparable subjects.

While the MSCM approach provides a marginal estimate of the treatment effect, the sequential Cox approach and its modified version provide conditional estimates. Both approaches are equipped to adjust for baseline confounders. Although the mechanisms and interpretations behind the sequential Cox approach and MSCM are different, both claim to achieve the same goal of estimating the causal effect of treatment in the presence of time-dependent confounders. Generally marginal and conditional estimates may not be directly comparable due to noncollapsibility.[9,15] By examining whether these quantities differed systematically in the settings investigated, we showed that the use of noncollapsible measure has not been an issue in the specific simulation settings we considered.

To the best of our knowledge, ours is the first study to use simulation studies to investigate the characteristics of a sequential Cox approach that has been suggested as being suitable in the context of time-dependent confounding. The first simulation setting (simulationI) deals with the situation where the time-dependent covariate is affected by the prior treatment (i.e. is a time-dependent confounder). When a time-dependent confounder is present, MSCM is known to be an appropriate method and hence results from this method are used as the standard for comparison in this simulation setting. In this simulation process, we generate data such that the time-dependent confounder dictates the treatment assignment in the following periods. Among the subjects selected for a mini-trial based on those initiating treatment or at risk in a given period, only current (and lagged) values of the time-dependent covariates are used as adjustments in the sequential Cox approach. We proposed a modified version of sequential Cox approach, based on controlling for the time-dependent variable (confounder or covariate) values after the treatment initiation. We investigated via simulation whether such adjustments are sufficient. In the second setting (simulation II), we have a baseline covariate and a time-dependent covariate. As the time-dependent Cox model is appropriate for simulation setting II, we use these results as the standard for comparison.

Previously, the sequential Cox approach was shown to work very well in comparison to the time-dependent Cox approach in the absence of any time-varying covariate or confounder.[10] However, we do not find the sequential

Cox approach to be as effective as MSCM when a time-dependent confounder is present (simulation I) or even when a time-dependent covariate which is not a time-dependent confounder is present (simulation II). The sequential Cox approach does not seem to remove the effects of time-dependent confounding adequately, especially when the event rate is small. Based on our simulation findings, when we need to consider time-dependent confounders in order to adequately model a disease process, we recommend the use of the MSCM approach or the modified sequential Cox approach, as both are capable of producing estimates of the treatment effect that are close to the MSCM target parameter ψ_1.

We apply the methods under consideration to estimate the effect of β-IFN on disease progression. The modified sequential Cox approach produces effect estimate similar to MSCM. A sensitivity analysis of the sequential Cox approach without using IPC weights yielded very similar results, implying little impact of artificial censoring due to later treatment initiation.

The focus for the sequential Cox approach and its modified version is on recreating the covariate process at each treatment start using the mini-trial approach.[7] Such focused and detailed scrutiny could yield insights about the data which may be hard to extract using a MSCM approach. For example, the data associated with a given mini-trial can be extracted and separated quite easily from the combined mini-trial data (pseudo-population), it is straightforward to compare the effects of early versus late treatment initiation. It is also possible to estimate the treatment effect for patients with a specific level of a time-dependent covariate at treatment initiation. Variance estimation is a challenge in the sequential Cox and similar methods.[15] To account for possible multiple entry of the same control subjects in different mini-trials, we used a robust (sandwich) estimator. In our simulation studies, the average standard errors are slightly lower than the empirical standard deviations in most cases for the sequential Cox approach. However, for the modified version, these estimates are very close. For more accurate estimate of the standard errors, bootstrap or jackknife estimates could be used.[9,13] Unlike our simulation settings, if the time-dependent covariates are very strongly affected by the prior treatment, further adjustments[16–19] may be necessary. Using the same simulation scheme used in this study, future studies could assess the adequacy of the sequential Cox approaches and other similar methods[17,27,40,41] under such extreme setting.

Similar to other simulation studies, we investigated a few possible scenarios. However, the assumptions underlying our data simulation are consistent with patterns typical in observational survival studies where associated covariates are measured regularly. For more complex disease scenarios where an investigator may wish to assess different treatment strategies (i.e. switching between therapies) over the course of time, our assumption of no discontinuations or interruptions in the treatment is restrictive and may not be suitable.[8] Further simulation studies are required to assess the effect on the precision of the estimates when varying the sample size of the simulated data as well as the number of simulated datasets generated from the algorithms considered in this study.[42] Future research could focus on analytical derivation of the effect estimates from the sequential Cox approach and its modified version in an effort to theoretically justify the proposed analysis roadmap.

Acknowledgments

We gratefully acknowledge the British Columbia Multiple sclerosis Clinic neurologists who contributed to the study through patient examination and data collection (current members listed here by primary clinic): UBC MS Clinic: Dr A. Traboulsee, MD, FRCPC (UBC Hospital MS Clinic Director and Head of the UBC MS Programs); Dr A-L. Sayao, MD, FRCPC; Dr V. Devonshire, MD, FRCPC; Dr S. Hashimoto, MD, FRCPC (UBC and Victoria MS Clinics); Dr J. Hooge, MD, FRCPC (UBC and Prince George MS Clinics); Dr L. Kastrukoff, MD, FRCPC (UBC and Prince George MS Clinics); Dr J. Oger, MD, FRCPC. Kelowna MS Clinic: Dr D. Adams, MD, FRCPC; D. Craig, MD, FRCPC; Dr S. Meckling, MD, FRCPC. Prince George MS Clinic: Dr L. Daly, MD, FRCPC. Victoria MS Clinic: Dr O. Hrebicek, MD, FRCPC; Dr D. Parton, MD, FRCPC; Dr K Pope, MD, FRCPC. The views expressed in this paper do not necessarily reflect the views of each individual acknowledged. None have received compensation for their role in the study.

Declaration of conflicting interest

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article:

MEK has received accommodation costs from the endMS Research and Training Network (2011, 2012), Statistical Society of Canada (2016) to present at conferences, and from Pacific Institute for the Mathematical Sciences (2013), The Canadian Statistical Sciences Institute (2016) to attend workshops. Over the past three years, JP has received consulting fees and/or fees

for service on Data Safety Monitoring Boards from EMD Serono, Myelin Research Foundation and Novartis. PG has received consulting fees from Biogen to serve on an advisory board. RWP has received fees for service for consulting from Abbvie, Amgen, and Searchlight Pharma, for teaching from Novartis, and for scientific steering committee membership from Pfizer. HT has received speaker honoraria and/or travel expenses to attend conferences from the Consortium of MS Centres (2013), the National MS Society (2012, 2014, 2016), Teva Pharmaceuticals (2011), ECTRIMS (2011, 2012, 2013, 2014, 2015, 2016), UK MS Trust (2011), the Chesapeake Health Education Program, US Veterans Affairs (2012), Novartis Canada (2012), Biogen Idec (2014), American Academy of Neurology (2013, 2014, 2015, 2016). All speaker honoraria are either declined or donated to an MS charity or to an unrestricted grant for use by her research group. YZ declares no conflicts of interest. AS has received travel grants to present at and attend conferences from the endMS Research and Training Network (2010, 2011), ECTRIMS (2010, 2011), and the Consortium of MS Centres (2012). EK has had travel and accommodation costs covered to present at and attend conferences from the endMS Research and Training Network (2008, 2011), the International Society for Pharmacoepidemiology (2010, 2013) and Bayer Schering Pharma (2010). CE has received travel grants to present and attend conferences from the endMS Research and Training Network (2011), and the European Committee for Treatment and Research in MS (2011). MLvdK declares no conflicts of interest. Over the past five years, JO has received speaker honoraria, consulting fees, travel grants and/or research grants and/or educational grants from Aventis, Bayer, Biogen-Idec, BioMS, Corixa, Genentech, Novartis, Serono, Shering, Talecris and Teva-neurosciences. JO receives fees for services from Bayer, from Novartis and from Biogen Idec to serve on advisory committees.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a studentship from the Multiple Sclerosis (MS) Society of Canada (MEK), by a postdoctoral fellowship from the Canadian Network for Observational Drug Effect Studies (CNODES) (MEK) and by a grant from the National Multiple Sclerosis Society (RG 4202-A-2, PI: HT) and the Canadian Institutes of Health Research (MOP-93646, PI: HT). JP holds research grants from the National MS Society, and the Natural Sciences and Engineering Research Council of Canada. PG is supported by the Natural Sciences and Engineering Research Council of Canada. RWP holds the Albert Boehringer I Chair in Pharmacoepidemiology and is supported in part by a National Scholar (Chercheur-national) of the Fonds de Récherche du Québec-Santé (FQR-S) and is a member of the Research Institute of the McGill University Health Centre, which is supported by core funds from FQR-S. HT is supported by the Canada Research Chair Program and currently receives research support from the National Multiple Sclerosis Society, the Canadian Institutes of Health Research, the Multiple Sclerosis Society of Canada, and the Multiple Sclerosis Scientific Research Foundation. In addition, in the last five years HT has received research support from the Multiple Sclerosis Society of Canada (Don Paty Career Development Award), the Michael Smith Foundation for Health Research (Scholar Award), and the UK MS Trust. YZ receives research funding from the Canadian Institutes of Health Research, the MS Society of Canada, and the National MS Society. AS was funded through a Postdoctoral Fellowship from the MS Society of Canada, and grants from the Canadian Institutes of Health Research (MOP-93646; PI=H.T.) and the National MS Society (RG 4202-A-2; PI=H.T.). EK was supported by Postdoctoral Fellowships from the Michael Smith Foundation for Health Research and the MS Society of Canada. CE was funded through grants from the Canadian Institutes of Health Research (MOP-93646; PI=H.T.) and the National MS Society (RG 4202-A-2; PI=HT), and the Michael Smith Foundation for Health Research. M.L.v.d.K. is supported by a Canadian Institutes of Health Research Doctoral Award—Doctoral Foreign Study Award (October 2012), offered in partnership with the Canadian Institutes of Health Research (CIHR) Strategy for Patient-Oriented Research, and the CIHR HIV/AIDS Research Initiative. JO receives support from the Christopher Foundation and the University of British Columbia (UBC). Over the past five years Dr Oger has received speaker honoraria, consulting fees, travel grants and/or research grants and/or educational grants of less than $5000 each from Bayer, Biogen-Idec, BioMS, Glaxo-Smith-Klein, Novartis, Merck-Serono, Talecris, and Teva-neurosciences. Dr Oger has received fees for services from Bayer and from Biogen Idec to serve on advisory committees.

References

1. Sylvestre MP and Abrahamowicz M. Comparison of algorithms to generate event times conditional on time-dependent covariates. *Stat Med* 2008; **27**: 2618–2634.
2. Hernán MA, Brumback B and Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; **11**: 561–570.
3. Fewell Z, Hernán MA, Wolfe F, et al. Controlling for time-dependent confounding using marginal structural models. *Stata J* 2004; **4**: 402–420.
4. Robins JM. Causal inference from complex longitudinal data. In: Berkane M (ed.) *Latent variable modeling and applications to causality*. Vol. 120: New York: Springer-Verlag, 1997, pp.69–117.

5. Robins JM, Hernán MA and Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**: 550–560.
6. Austin PC and Platt RW. Survivor treatment bias, treatment selection bias, and propensity scores in observational research. *J Clin Epidemiol* 2010; **63**: 136–138.
7. Aalen OO. Armitage lecture 2010: understanding treatment effects: the value of integrating longitudinal data and survival analysis. *Stat Med* 2012; **31**: 1903–1917.
8. Lange T and Rod NH. Causal models. In: Klein JP, Van Houwelingen HC, Ibramhim JH and Scheike TH (eds) *Handbook of Survival Analysis*. London: Chapman-Hall, 2013, pp.135–151.
9. Gran JM, Røysland K, Wolbers M, et al. 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.
10. Karim ME, Gustafson P, Petkau J, et al. Comparison of statistical approaches dealing with immortal time bias in drug effectiveness studies. *Am J Epidemiol*, Epub ahead of print 25 July 2016. doi: 10.1093/aje/kwv445.
11. Karim ME, Gustafson P, Petkau J, et al. Marginal structural Cox models for estimating the association between β-Interferon exposure and disease progression in a multiple sclerosis cohort. *Am J Epidemiol* 2014; **180**: 160–171.
12. Shirani A, Zhao Y, Karim ME, et al. Association between use of interferon beta and progression of disability in patients with relapsing-remitting multiple sclerosis. *J Am Med Assoc* 2012; **308**: 247–256.
13. Røysland K, Gran JM, Ledergerber B, et al. Analyzing direct and indirect effects of treatment using dynamic path analysis applied to data from the Swiss HIV cohort study. *Stat Med* 2011; **30**: 2947–2958.
14. Xiao Y, Abrahamowicz M and Moodie EEM. Accuracy of conventional and marginal structural Cox model estimators: a simulation study. *Int J Biostat* 2010; **6**: 1–28.
15. Taylor JMG, Shen J, Kennedy EH, et al. Comparison of methods for estimating the effect of salvage therapy in prostate cancer when treatment is given by indication. *Stat Med* 2014; **33**: 257–274.
16. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008; **19**: 766–779.
17. Schaubel DE, Wolfe RA and Port FK. A sequential stratification method for estimating the effect of a time-dependent experimental treatment in observational studies. *Biometrics* 2006; **62**: 910–917.
18. Schaubel DE, Wolfe RA, Sima CS, et al. Estimating the effect of a time-dependent treatment by levels of an internal time-dependent covariate: application to the contrast between liver wait-list and posttransplant mortality. *J Am Stat Assoc* 2009; **104**: 49–59.
19. Lu B. Propensity score matching with time-dependent covariates. *Biometrics* 2005; **61**: 721–728.
20. Young JG, Hernán MA, Picciotto S, et al. Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime Data Anal* 2010; **16**: 71–84.
21. Abrahamowicz M, Mackenzie T and Esdaile JM. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. *J Am Stat Assoc* 1996; **91**: 1432–1439.
22. Moodie EEM, Stephens DA and Klein MB. A marginal structural model for multiple-outcome survival data: assessing the impact of injection drug use on several causes of death in the Canadian co-infection cohort. *Stat Med* 2014; **33**: 1409–1425.
23. Cook NR, Cole SR and Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' health study. *Am J Epidemiol* 2002; **155**: 1045–1053.
24. Gerhard T, Delaney JAC, Cooper-DeHoff RM, et al. Comparing marginal structural models to standard methods for estimating treatment effects of antihypertensive combination therapy. *BMC Med Res Methodol* 2012; **12**: 1.
25. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. *Stat Models Epidemiol Environ Clin Trials* 1999; **116**: 95–134.
26. Hernán MA, Brumback B and Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Stat Assoc* 2001; **96**: 440–448.
27. Kennedy EH, Taylor JMG, Schaubel DE, et al. The effect of salvage therapy on survival in a longitudinal study with treatment by indication. *Stat Med* 2010; **29**: 2569–2580.
28. Cole SR and Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008; **168**: 656–664.
29. Westreich D, Cole SR, Tien PC, et al. Time scale and adjusted survival curves for marginal structural Cox models. *Am J Epidemiol* 2010; **171**: 691–700.
30. Kaufman JS. Marginalia: comparing adjusted effect measures. *Epidemiology* 2010; **21**: 490–493.
31. Godin O, Elbejjani M and Kaufman JS. Body mass index, blood pressure, and risk of depression in the elderly: a marginal structural model. *Am J Epidemiol* 2012; **176**(3): 204–213.
32. Pang M, Kaufman JS and Platt RW. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Stat Methods Med Res*, Epub ahead of print 9 October 2013. DOI: 0962280213505804.
33. Xiao Y, Moodie EEM and Abrahamowicz M. Comparison of approaches to weight truncation for marginal structural Cox models. *Epidemiol Methods* 2012; **2**: 1–20.
34. Shrier I and Pang M. Confounding, effect modification, and the odds ratio: common misinterpretations. *J Clin Epidemiol* 2015; **68**: 470.

35. Shirani A, Zhao Y, Karim ME, et al. Investigation of heterogeneity in the association between interferon beta and disability progression in multiple sclerosis: an observational study. *Eur J Neurol* 2014; **21**: 835–844.
36. Karim ME. *Causal inference approaches for dealing with time-dependent confounding in longitudinal studies, with applications to multiple sclerosis research*. PhD Thesis, University of British Columbia, Vancouver, British Columbia, Canada, 2015.
37. Zhang T, Shirani A, Zhao Y, et al. Beta-interferon exposure and onset of secondary progressive multiple sclerosis. *Eur J Neurol* 2015; **22**: 990–1000.
38. Shirani A, Zhao Y, Petkau J, et al. Multiple sclerosis in older adults: the clinical profile and impact of interferon beta treatment. *BioMed Res Int* 2015; **2015**: 1–11.
39. Sterne JAC, Hernán MA, Ledergerber B, et al. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *Lancet* 2005; **366**: 378–384.
40. Van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scand J Stat* 2007; **34**: 70–85.
41. Toh S and Hernán MA. Causal inference from longitudinal studies with baseline randomization. *Int J Biostat* 2008; **4**: 1–30.
42. Burton A, Altman DG, Royston P, et al. The design of simulation studies in medical statistics. *Stat Med* 2006; **25**: 4279–4292.

Title: Comparison of Statistical Approaches Dealing with Time-dependent Confounding in Drug Effectiveness Studies
Authors: Karim, M. E.; Petkau, J.; Gustafson, P.; Platt, R.W.; Tremlett, H. and BeAMS study group.
Journal: Statistical Methods in Medical Research. First published online: Sep 21, 2016.
DOI: 10.1177/0962280216668554

WEB-APPENDIX

A MSCM Model Specification

In the presence of baseline covariates L_0, the hazard function can be expressed as the following time-dependent Cox model:

$$\lambda(m|L_0) \quad = \quad \lambda_0(m)\exp\big(\psi_1 A_m + \psi_2 L_0\big), \tag{A.1}$$

where m is the visit index, $\lambda_0(m)$ is the unspecified baseline hazard function, ψ_1 is the log-HR of the current treatment status (A_m) and ψ_2 is the vector of log-HRs for the baseline covariates. Here the impact of treatment is modelled based on only current exposure[1].

In presence of a time-dependent confounder L_m, we may want to expand the above Cox model to:

$$\lambda(m|L_0, L_m) = \lambda_0(m)\exp\big(\psi_1 A_m + \psi_2 L_0 + \psi_3 L_m\big),$$

which may produce a biased estimate of ψ_1 if L_m is influenced by past exposure[1]. Nonetheless, as L_m is a confounder, we still need to adjust for confounding due to L_m somehow. IPWs are person-time specific measures of the degree to which L_m confounds the treatment selection process. Therefore, in MSCM, IPWs are used in the time-dependent Cox model formulation (equation (A.1)) to weight the contribution of each person-time observation so that the confounding due to L_m is removed.

B Model Specifications for Estimating the Weights

The unstabilized IPTW is expressed as:

$$w_m^T \quad = \quad \prod_{j=0}^{m} \frac{1}{pr(A_j = a_j|\bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0, \bar{L}_j = \bar{l}_j)}, \tag{B.1}$$

A pooled logistic regression model is used to estimate the probabilities in equation (B.1) as follows:

$$logit\ Pr(A_j = 1|\bar{A}_{j-1}, L_0, \bar{L}_j) \quad = \quad \alpha_0(j) + \alpha_1 A_{j-1} + \alpha_2 L_0 + \alpha_3 L_j. \tag{B.2}$$

Here, $\alpha_0(j)$ is a smooth function[1,2] of the month index j, A_j is the current treatment status, A_{j-1} is the treatment status at the previous time interval, L_0 is the collection of baseline covariates, and L_j is the time-varying confounder. The predicted probabilities from equation (B.2) yield the estimated probability of the subject's treatment status at time j. Multiplying the corresponding probabilities as indicated in equation (B.1) yields the probability of the observed exposure sequence over m time periods of a given subject.

To obtain the stabilized IPTW, we use the following formula:

$$sw_m^T = \prod_{j=0}^{m} \frac{pr(A_j = a_j|\bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0)}{pr(A_j = a_j|\bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0, \bar{L}_j = \bar{l}_j)}. \tag{B.3}$$

The numerator terms are estimated from:

$$logit\ Pr(A_j = 1|\bar{A}_{j-1}, L_0) = \alpha_0'(j) + \alpha_1' A_{j-1} + \alpha_2' L_0. \tag{B.4}$$

Dividing the estimated numerator probabilities of the subject's observed treatment status a_j by the corresponding estimated denominator probabilities yields the estimated IPTWs sw_m^T that account for the confounding due to $\bar{L}_m$.

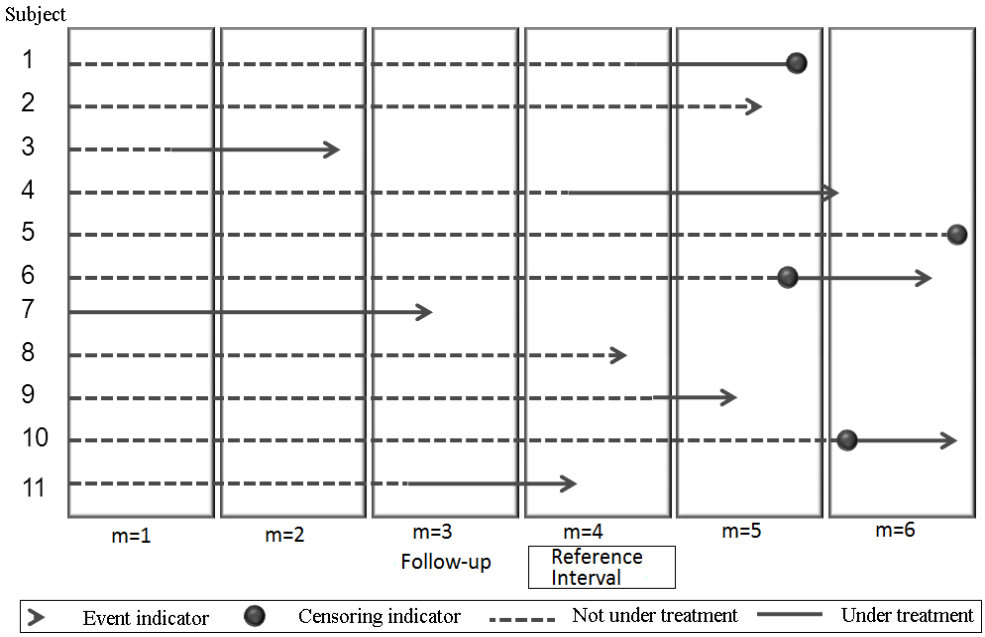
Using similar logic to that leading to the IPTW for uncensored patients, the stabilized IPCW can be obtained as[3]:

$$sw_m^C = \prod_{j=0}^{m} \frac{pr(C_j = 0|\bar{C}_{j-1} = 0, \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0)}{pr(C_j = 0|\bar{C}_{j-1} = 0, \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0, \bar{L}_{j-1} = \bar{l}_{j-1})}, \tag{B.5}$$

where C_j denotes the binary censoring status taking the value of 1 if the patient was censored in the j-th month and 0 otherwise. The overall stabilized IPTC weights sw_m are obtained by multiplying sw_m^T by sw_m^C[4].

C Constructing a Mini-trial in the Sequential Cox Approach

To illustrate the method, consider Web-Figure C.1, where the follow–up times for 11 subjects are outlined. Patient 1 was not under treatment when entering the study. This individual started taking the treatment in the $m = 4$th month and was censored during the 5th month. Similarly subject 5, who was never under treatment was censored during the 6th month.



Web-Figure C.1: An illustration of the sequential Cox approach

Now, suppose we want to create the mimicked trial considering the 4th month as the reference interval. We eliminate the subjects who received treatment before the 4th month, i.e., the 3rd, 7th and 11th subjects are discarded. Then for the subjects who started treatment after the 4th month, we censor them at the time of treatment start i.e., the 6th and 10th subjects are censored at the 5th and 6th months respectively. Then, under the assumption that treatment status remains the same for the entire month, subjects 1, 4 and 9 are considered the treated group and subjects 2, 5, 6, 8 and 10 are considered the control group, for the mimicked trial starting at the beginning of 4th month.

In this mimicked trial, a subject is considered either on treatment or off treatment during the entire duration of the follow-up. Therefore, this manipulated subset of the data mimics a clinical trial. A Cox proportional hazards model can be used to compare the survival experiences of these two groups. Similarly, we can identify the subjects for the treatment and control groups in the mimicked trials starting at the beginning of other months. This yields multiple mimicked trials, one for each of the time intervals (say, months) of treatment start. The intervals in which no subject initiates treatment do not have a corresponding mimicked trial.

One way to get a treatment effect estimate is to fit a stratified Cox model on the combined data of all mini-trials (pseudo-data), stratified by the treatment initiation time. In this paper, we used this approach. Alternatively, a simple Cox model weighted by IPCW can be run for each of the successive mini-trials to obtain separate estimates of the treatment effect for each mini-trial, leading to the name of this approach, the sequential Cox approach. An overall estimate of the treatment effect is obtained by simply averaging the treatment effect estimates from the separate mini-trials. Convergence may be an issue if some mini-trials have only a few subjects, which could be the case in mini-trials starting near the end of the follow-up. This may have an impact on the estimation of IPCW if we are estimating them separately for each mini-trial.

The overall estimate (from the above two approaches) requires two additional assumptions for causal interpretation: (1) the treatment effect is the same in all the mini-trials and (2) the treatment effect is unchanged for all covariate histories before the m-th interval, given the covariates at the m-th interval. However, if one is willing to interpret the overall effect estimate as an aggregated (averaged) effect over all the mini-trials, then the first assumption can be relaxed[5,6]. Whether the two estimators (using combined pseudo-data or averaging the results from the separate mini-trials) are estimating the same target parameter may depend on satisfying the stated assumptions.

D Implementation of the Sequential Cox Approach in R

The `coxph` function in the `survival` package[7] is used to fit both time-independent and time-dependent Cox PH models. The combined mini-trial (pseudo) dataset can become large due to repeated use of the same control subjects.

In the `coxph` function, the option `strata` is set to fit a stratified Cox model for the sequential Cox approach. Also, the options such as `cluster` and `robust = TRUE` are set to obtain the robust (sandwich) variance estimate. This is an approximate grouped jackknife variance estimate[8] when multiple observations per subject are present. Aalen's additive regression is fitted using the `aalen` function in the `timereg` package to estimate the IPCWs[6]. To obtain bootstrap estimates[9], the `lapply` function can be used on each bootstrap sample to estimate

the corresponding IPCWs and subsequently the HR from a Cox PH.

E MSCM Data Simulation Algorithm Pseudocode

A number of different simulation schemes are available in the literature to simulate survival times in the presence of a time-dependent confounder[10–16]. The algorithm we used[11,17] generates data satisfying the conditions of the following three models simultaneously: MSM, structural nested accelerated failure time model and a structural nested cumulative failure time model. The steps of this algorithm are also described elsewhere[11,12,16,18–20].

GET

$n \leftarrow 2500$;

$K \leftarrow 10$ (maximum follow-up);

$\lambda_0 \leftarrow 0.01$ (rare events) or 0.10 (frequent events);

$\beta \leftarrow [\log(3/7), 2, \log(1/2), \log(3/2)]$ (parameter vector for generating L);

$\alpha \leftarrow [\log(2/7), (1/2), (1/2), 10]$ (parameter vector for generating A);

$\psi_1 \leftarrow 0.5$ (true log-HR value of the treatment effect)

COMPUTE

FOR $ID = 1$ to n

INIT: $L_{-1} \leftarrow 0$; $A_{-1} \leftarrow 0$; $Y_0 \leftarrow 0$; $H_m \leftarrow 0$; $c \leftarrow 30$

$T_{\bar{0}} \sim$ Exponential(λ_0)

FOR $m = 1$ to K

$logit\ p_L \leftarrow$ logit $Pr(L_m = 1|L_{m-1}, A_{m-1}, Y_m = 0; \beta)$

$\leftarrow \beta_0 + \beta_1 I(T_0 < c) + \beta_2 A_{m-1} + \beta_3 L_{m-1}$

$L_m \sim$ Bernoulli(p_L)

$logit\ p_A \leftarrow$ logit $Pr(A_m = 1|L_m, L_{m-1}, A_{m-1}, Y_m = 0; \alpha)$

$\leftarrow \alpha_0 + \alpha_1 L_m + \alpha_2 A_{m-1} + \alpha_3 L_{m-1}$

$A_m \sim$ Bernoulli(p_A)

$H_m \leftarrow \int_0^{m+1} \lambda_{\bar{a}_j}(j)dj$

$\leftarrow H_m + \exp(\psi_1 \times A_m)$

IF $T_0 \geq H_m$

$Y_{m+1} \leftarrow 0$

ELSE

$Y_{m+1} \leftarrow 1$

$T \leftarrow m + (T_0^{-} - H_m) \times \exp(-\psi_1 \times A_m)$

END IF

ENDFOR m

ENDFOR ID

PRINT

ID, m, Y_{m+1}, A_m, L_m, A_{m-1}, L_{m-1}

F Survival Data Simulation via Permutation Algorithm

This algorithm has been validated for generating survival times conditional on time-dependent treatment[21] and also when time-dependent covariates are present[22]. This algorithm has been used in several other studies dealing with generating survival data with time-dependent covariates (see for example[23–27]). The algorithm has the following steps:

1. For each subject $i = 1, 2, \ldots, n$, generate the survival time T_i using a specified distribution.

2. For each subject i, generate the censoring time T_i^C using a specified distribution.

3. Find the observed survival time $T_i^* = min(T_i, T_i^C)$ and the binary censoring indicator $C_i = I(T_i \geq T_i^C) = 1$ if censored and 0 otherwise.

4. Repeat steps 1-3 n times and sort survival status tuples (T_i^*, C_i) with respect to T_i^* in increasing order.

5. Generate n covariate matrices $X_i = (A_{im}, L_{i0}, L_{im})$ with dimensions $(m \times p)$, where the $m = 0, 1, \ldots, K$ rows correspond to the different time intervals or visits when measurements are taken and the p columns correspond to the predictor variables, including treatment (A_m), time-fixed and/or time-varying covariates (L_0 and/or L_m). For subject i, X_{im}, the m-th row of X_i, is a vector of variable values at time m.

6. According to the ordered T_i^* listed in step 3, begin assigning the survival status tuple (T_i^*, C_i) to covariate values from X_{im} as follows. At time T_i^*, variable values (treatment

and covariate) are sampled with probabilities p_{im} defined below based on the Cox model's partial likelihood:

$$p_{im} = \begin{cases} \frac{\exp(\psi X_{im})}{\sum_{j \in r_i} \exp(\psi X_{jm})}, & \text{if } C_i = 0 \\ \frac{1}{\sum_{j \in r_i} I(j \in r_i)}, & \text{if } C_i = 1, \end{cases}$$

where ψ is the vector of log-HRs for the corresponding variables and $I(j \in r_i)$ indicates whether a subject is within a given riskset r_i for time T_i^*.

7. The subject i with the covariate values X_{im} is assigned the observed time T_i^*. The selected X_{im} is removed from further calculation.

The permutation algorithm is implemented in the `PermAlgo` package in R[28].

G Summary of Selected Cohorts and Exclusion Criteria

The eligibility criteria used for β-IFN treatment are: patients have to be at least 18 years old, have an Expanded Disability Status Scale (EDSS) score of 6.5 or below (i.e., able to walk 20 meters without resting with constant bilateral support) and have definite MS with a relapsing-onset course. 2,671 patients met the eligibility criteria to receive β-IFN treatment between July 1995 and December 2004[29,30].

Web-Table G.1: Characteristics of the selected cohort of patients with relapsing-onset multiple sclerosis (MS), British Columbia, Canada (1995-2008).

Baseline characteristics	Ever-β-IFN exposed	Never-β-IFN exposed
Number	868	829
Women, n (%)	660 (76.0)	637 (76.8)
Disease duration, average (SD)	5.8 (6.6)	8.3 (8.5)
Age, average (SD)	38.1 (9.2)	41.3 (10.0)
EDSS score, median (range)	2.0 (0-6.5)	2.0 (0-6.5)
Relapse rate / year†, median (IQR)	0.5 (0-1.2)	0.5 (0-1.0)

† Over the 2 years prior to baseline.

Of these, patients who were exposed to a non-β-IFN immunomodulatory drug, a cytotoxic immunosuppressant for MS ($n = 172$), or an MS clinical trial ($n = 21$) prior to baseline were excluded from the analysis. If the exposure occurred after baseline, data were censored at the start of the exposure to the non-β-IFN treatment. Further exclusion criteria included unknown MS onset date ($n = 10$), insufficient EDSS measurements ($n = 436$), reaching of the outcome ($n = 218$) or the secondary progressive stage before the eligibility date ($n = 217$). Some patients met multiple exclusion criteria. As a result, $1,697$ patients were selected. A summary of their characteristics are reported in Web-Table G.1.

H Additional Simulation Results

H.1 When More Events are Available

Results from the more frequent event condition are presented in the Tables H.1-H.2 ($\lambda_0 = 0.10$ on a monthly scale).

Web-Table H.1: Comparison of the analytical approaches to adjust for time-dependent confounding from simulation-I (one time-dependent confounder and time-dependent treatment exposure) of $1,000$ datasets, each containing $2,500$ subjects followed for up to 10 time-intervals (frequent event case).

Approach	Bias	$SD(\hat{\psi}_1)$	$se(\hat{\psi}_1)$	Coverage Probability
TD-Cox[§]	0.044	0.067	0.065	0.888
Sequential Cox[#, †]	0.174	0.098	0.097	0.560
Modified Sequential Cox[*, @]	-0.035	0.074	0.073	0.924
MSCM[‡]	0.000	0.069	0.068	0.942

TD-Cox, Cox model with time-dependent exposure; MSCM, Marginal structural Cox model.

[§] Includes the time-dependent confounder L_m as a covariate. In the presence of a time-dependent confounder, the time-dependent Cox model is not appropriate but the results are retained for comparison purposes.

[#] Adjusts for $\tilde{L}_m$.

[†] For the stabilized IPCWs, the numerator model adjusts for A_m, while the denominator model adjusts for A_m and $\tilde{L}_m$ via Aalen's additive regression.

[*] Adjusts for lagged values of A_m, the time-dependent confounder $\vec{L}_m$, and lagged values of $\vec{L}_m$. Note that, baseline covariates are not present in this setting.

[@] For the stabilized IPCWs, the numerator model adjusts for A_m, while the denominator model adjusts for A_m, $\vec{L}_m$ and lagged values of $\vec{L}_m$ via Aalen's additive regression.

[‡] The stabilized IPTW numerator model adjusts for time index and lagged values of A_m, while the denominator model additionally adjusts for current and lagged values of L_m to predict future treatment status via pooled logistic models.

Web-Table H.2: Comparison of the analytical approaches to adjust for time-dependent covariate from simulation-II (one baseline covariate, one time-dependent covariate and time-dependent treatment exposure) of $1,000$ datasets, each containing $2,500$ subjects followed for up to 10 time-intervals (frequent event case).

Approach	Bias	$SD(\hat{\psi}_1)$	$se(\hat{\psi}_1)$	Coverage Probability
TD-Cox[§]	-0.002	0.059	0.060	0.960
Sequential Cox[#, †]	0.218	0.063	0.064	0.074
Modified Sequential Cox[*, @]	-0.034	0.083	0.083	0.945
MSCM[±, ‡]	-0.014	0.058	0.060	0.952

TD-Cox, Cox model with time-dependent exposure; MSCM, Marginal structural Cox model.

[§] The baseline covariate L_0 and time-dependent covariate L_m are included.

[#] Adjusts for L_0 and $\tilde{L}_m$.

[†] In the stabilized IPCW model, the numerator model adjusts for A_m and L_0, while the denominator model adjusts for A_m, L_0 and $\tilde{L}_m$ via Aalen's additive model.

[*] Adjusts for baseline covariates L_0, lagged values of A_m, the time-dependent confounder $\vec{L}_m$, and lagged values of $\vec{L}_m$.

[@] For the stabilized IPCWs, the numerator model adjusts for A_m and baseline variable L_0, while the denominator model adjusts for L_0, A_m, $\vec{L}_m$ and lagged values of $\vec{L}_m$ via Aalen's additive regression.

[±] Adjusts for only L_0.

[‡] For the stabilized IPTWs, the numerator model adjusts for the time index, L_0 and lagged values of A_m, while the denominator model additionally adjusts for current and lagged values of L_m to predict future treatment status via pooled logistic models.

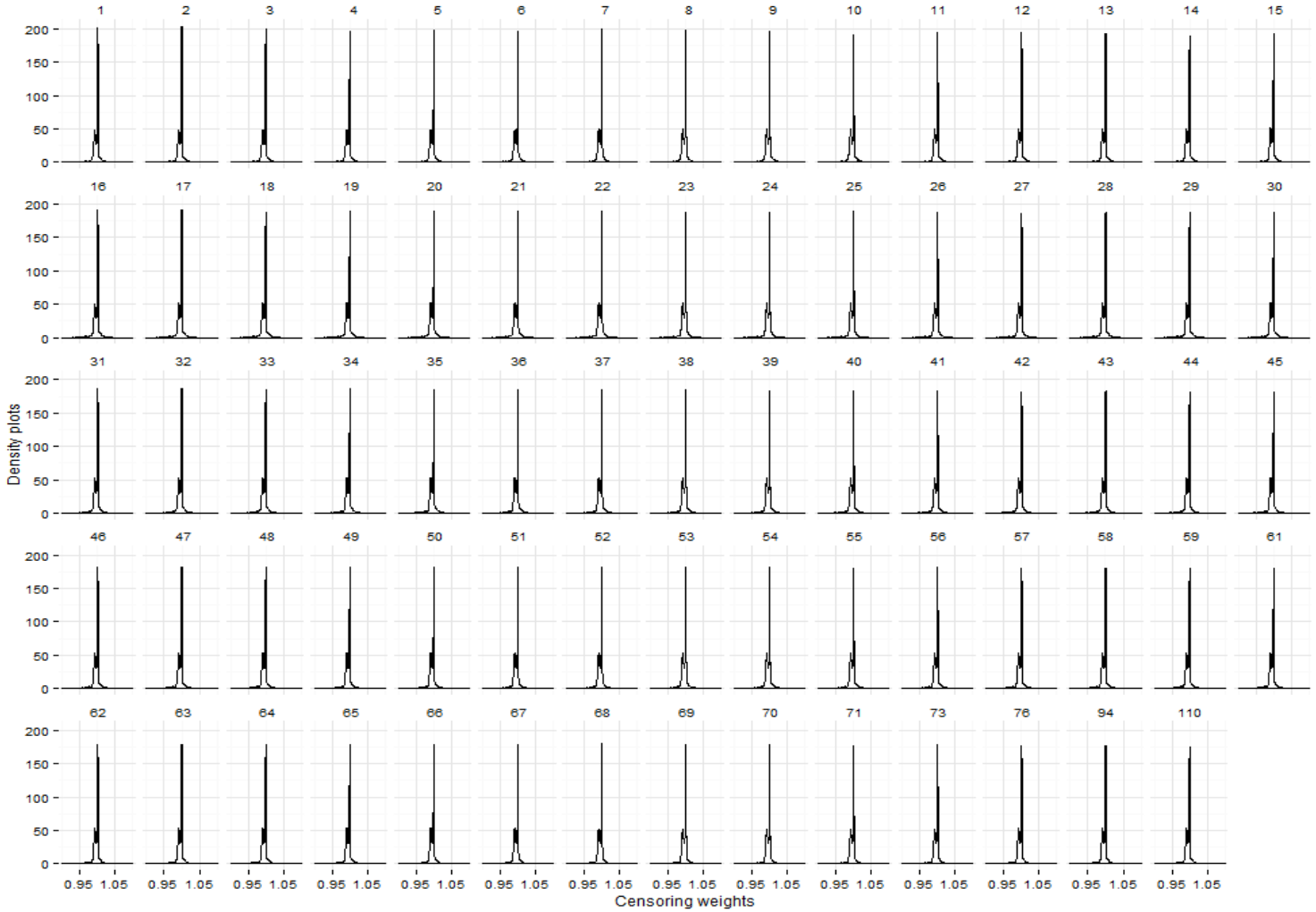
I Additional MS Data Analysis: Modified Sequential Cox Approach

The HRs for the treatment estimated using a the modified sequential Cox approach when IPCWs are calculated from different approaches are reported in Web-Table I.1. The analyses are adjusted for baseline covariates: sex, EDSS score, age, disease duration and time-dependent confounder 'cumulative relapse' measured at baseline, treatment initiation month and its lagged value.

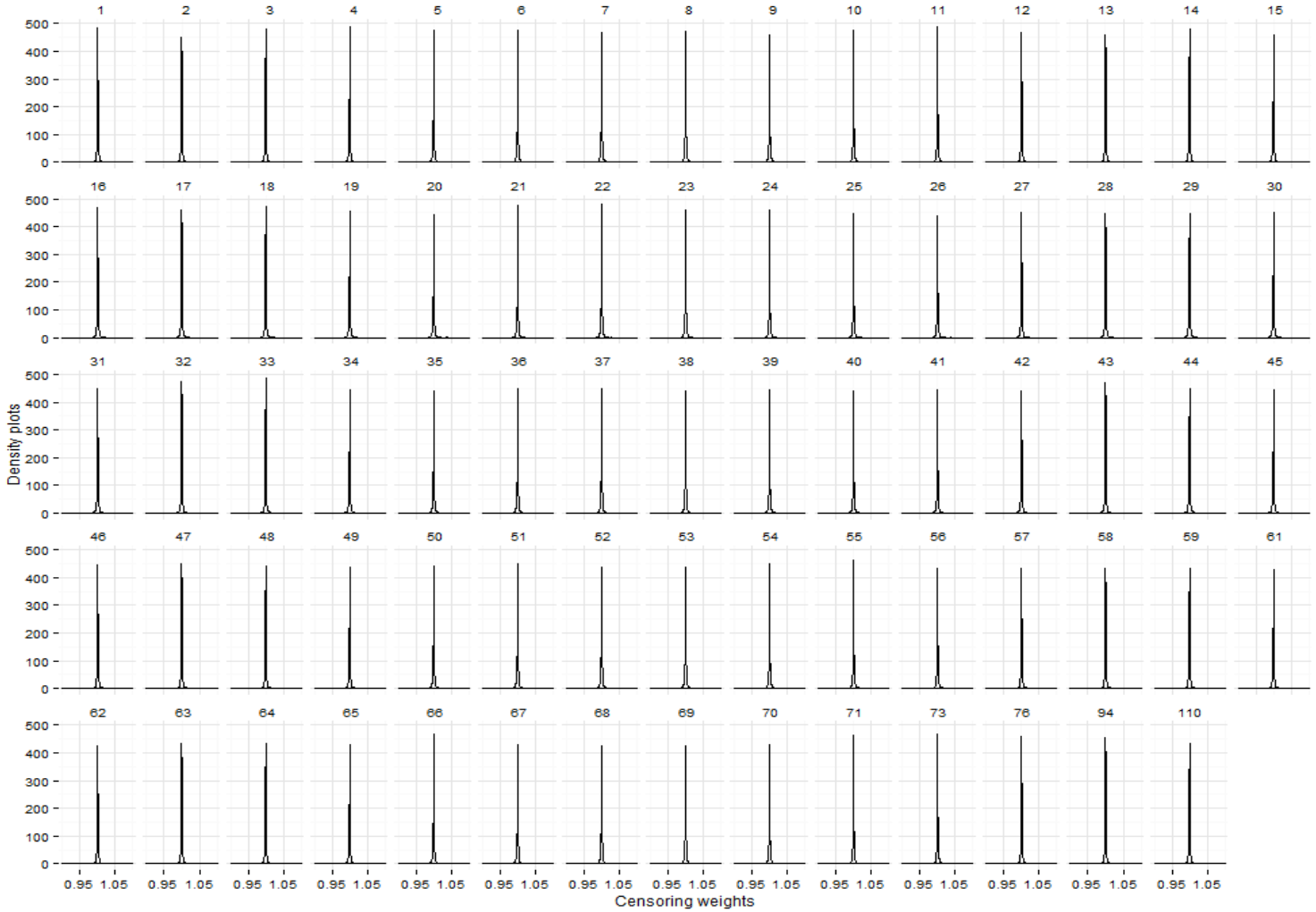
Web-Table I.1: Estimated hazard ratio using the modified sequential Cox approach to estimate the causal effect of β-IFN on time to sustained EDSS 6 for patients with relapsing-onset multiple sclerosis (MS), British Columbia, Canada (1995-2008), when IPCWs are calculated using different approaches.

IPCW estimation	HR	$se(\hat{HR})$	95% CI	Weights	
				Average (SD)	range
No weights	1.36	0.26	0.93 - 1.99		
Aalen's regression[‡]	1.36	0.26	0.93 - 1.99	1.00 (0.01)	0.92 - 1.24
Aalen's regression[†]	1.36	0.26	0.94 - 1.99	1.00 (0.02)	0.41 - 1.31
Pooled logistic[‡]	1.36	0.26	0.93 - 1.99	1.00 (0.01)	0.36 - 1.51
Pooled logistic[†]	1.36	0.26	0.93 - 1.99	1.00 (0.01)	0.95 - 1.15

[‡] IPCW estimated from each mini-trial separately.
[†] IPCW estimated from the aggregated data of all mini-trials.



Web-Figure I.1: Density plots of the estimated IPC weights via Aalen's additive regression from the MS data (estimated from the aggregated data of all mini-trials) in all the reference intervals using the modified sequential Cox approach



Web-Figure I.2: Density plots of the estimated IPC weights via pooled logistic from the MS data (estimated from the aggregated data of all mini-trials) in all the reference intervals using the modified sequential Cox approach

Web-References

[1] M.A. Hernán, B. Brumback, and J.M. Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5):561–570, 2000.

[2] Z. Fewell, M.A. Hernán, F. Wolfe, K. Tilling, H. Choi, and JA Sterne. Controlling for time-dependent confounding using marginal structural models. *The Stata Journal*, 4(4):402–420, 2004.

[3] J.M. Robins, S. Greenland, and F.C. Hu. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association*, 94(447):687–700, 1999.

[4] J.M. Robins, M.A. Hernán, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

[5] J.M. Gran, K. Røysland, M. Wolbers, V. Didelez, J.A.C. Sterne, B. Ledergerber, H. Furrer, V. von Wyl, and O.O. Aalen. 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. *Statistics in Medicine*, 29(26):2757–2768, 2010.

[6] T. Lange and N.H. Rod. Causal models. In *Handbook of Survival Analysis*, pages 135–151. CRC Press, 2013.

[7] T. Therneau. *A Package for Survival Analysis in S*, 2014. R package version 2.37-7, Last accessed: Sep-15,2014.

[8] T.M. Therneau. *Modeling Survival Data: Extending the Cox Model*. Springer, 2000.

[9] B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. CRC press, 1994.

[10] J. Bryan, Z. Yu, and M.J. van der Laan. Analysis of longitudinal marginal structural models. *Biostatistics*, 5(3):361–380, 2004.

[11] J.G. Young, M.A. Hernán, S. Picciotto, and J.M. Robins. Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime Data Analysis*, 16(1):71–84, 2010.

[12] Y. Xiao, M. Abrahamowicz, and E.E.M. Moodie. Accuracy of conventional and marginal structural Cox model estimators: A simulation study. *The International Journal of Biostatistics*, 6(2):1–28, 2010.

[13] D. Westreich, S.R. Cole, E.F. Schisterman, and R.W. Platt. A simulation study of finite-

sample properties of marginal structural Cox proportional hazards models. *Statistics in Medicine*, 31(19):2098–2109, 2012.

[14] W.G. Havercroft and V. Didelez. Simulating from marginal structural models with time-dependent confounding. *Statistics in Medicine*, 31(30):4190–4206, 2012.

[15] J.G. Young and E.J. Tchetgen Tchetgen. Simulation from a known Cox MSM using standard parametric models for the g-formula. *Statistics in Medicine*, 33(6):1001–1014, 2014.

[16] E.E.M. Moodie, D.A. Stephens, and M.B. Klein. A marginal structural model for multiple-outcome survival data: assessing the impact of injection drug use on several causes of death in the Canadian co-infection cohort. *Statistics in Medicine*, 33(8):1409–1425, 2014.

[17] J.G. Young, M.A. Hernán, S. Picciotto, and J.M. Robins. Simulation from structural survival models under complex time-varying data structures. In *JSM Proceedings, Section on Statistics in Epidemiology*, pages 1–6. American Statistical Association, 2008.

[18] R.A. Ali, M.A. Ali, and Z. Wei. On computing standard errors for marginal structural Cox models. *Lifetime Data Analysis*, 20(1):106–131, 2014.

[19] M.E. Karim. Causal inference approaches for dealing with time-dependent confounding in longitudinal studies, with applications to multiple sclerosis research, 2015. Ph.D. Thesis, University of British Columbia.

[20] N. Mojaverian, E.E.M. Moodie, A. Bliu, and M.B. Klein. The impact of sparse follow-up on marginal structural models for time-to-event data. *American journal of epidemiology*, 182(12):1047–1055, 2015.

[21] T. Mackenzie and M. Abrahamowicz. Marginal and hazard ratio specific random data generation: Applications to semi-parametric bootstrapping. *Statistics and Computing*, 12(3):245–252, 2002.

[22] M.P. Sylvestre and M. Abrahamowicz. Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine*, 27(14):2618–2634, 2008.

[23] M. Abrahamowicz and T.A. MacKenzie. Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. *Statistics in Medicine*, 26(2):392–408, 2007.

[24] M. Sylvestre and M. Abrahamowicz. Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine*, 28(27):3437–3453, 2009.

[25] A. Mahboubi, M. Abrahamowicz, R. Giorgi, C. Binquet, C. Bonithon-Kopp, and C. Quantin. Flexible modeling of the effects of continuous prognostic factors in relative survival. *Statistics in Medicine*, 30(12):1351–1365, 2011.

[26] M. Abrahamowicz, M. Beauchamp, and M. Sylvestre. Comparison of alternative models for linking drug exposure with adverse effects. *Statistics in Medicine*, 31(11-12):1014–1030, 2012.

[27] H. Gauvin, A. Lacourt, and K. Leffondré. On the proportional hazards model for occupational and environmental case-control analyses. *BMC Medical Research Methodology*, 13(1):18, 2013.

[28] M. Sylvestre, T. Edens, T. MacKenzie, and M. Abrahamowicz. *Package 'PermAlgo'*, 2010. Last accesses: Sep-10-2014.

[29] A. Shirani, Y. Zhao, M.E. Karim, C. Evans, E. Kingwell, M. van der Kop, J. Oger, P. Gustafson, J. Petkau, and H. Tremlett. Association between use of interferon beta and progression of disability in patients with relapsing-remitting multiple sclerosis. *Journal of American Medical Association*, 308(3):247–256, 2012.

[30] M. E. Karim, P. Gustafson, J. Petkau, Y. Zhao, A. Shirani, E. Kingwell, C. Evans, M. van der Kop, J. Oger, and H. Tremlett. Marginal structural Cox models for estimating the association between β-Interferon exposure and disease progression in a multiple sclerosis cohort. *American Journal of Epidemiology*, 180(2):160–171, 2014.