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:

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

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:

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

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

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

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 $\bar{L}_m$.

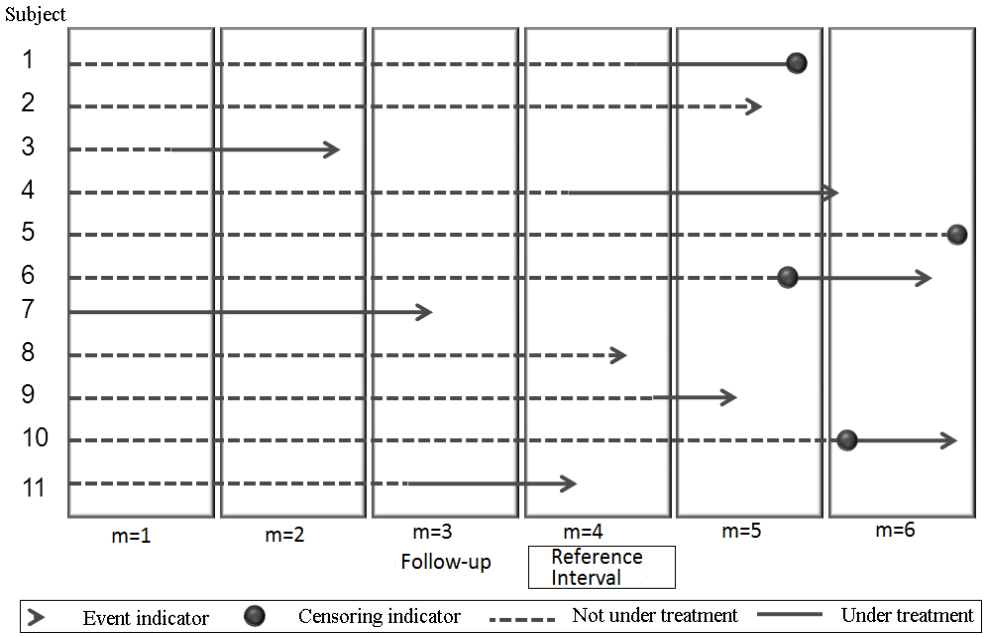
Using similar logic to that leading to the IPTW for uncensored patients, the stabilized IPCW can be obtained as[3]:

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

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.

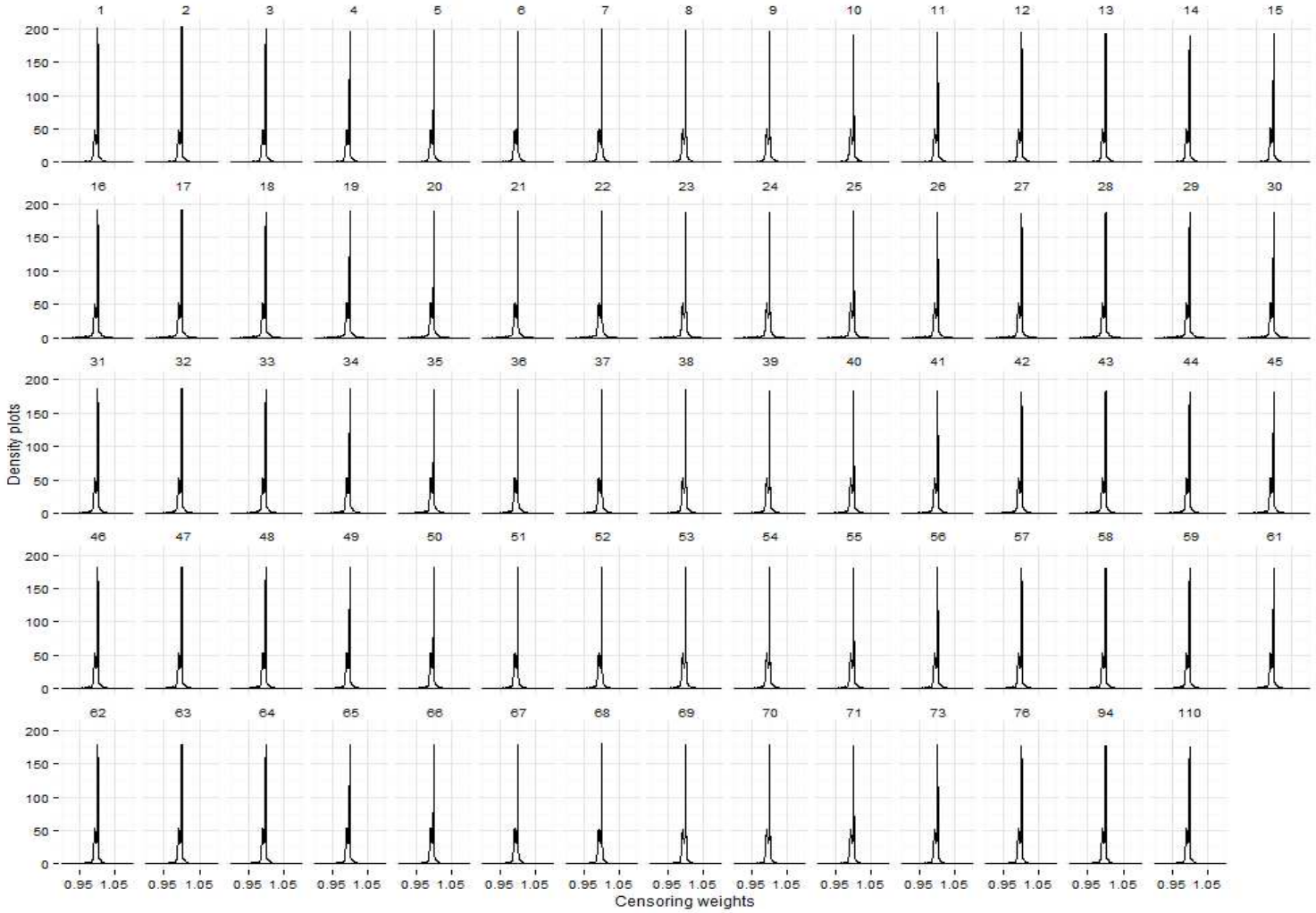
I Additional MS Data Analysis: Modified Sequential Cox Approach

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.

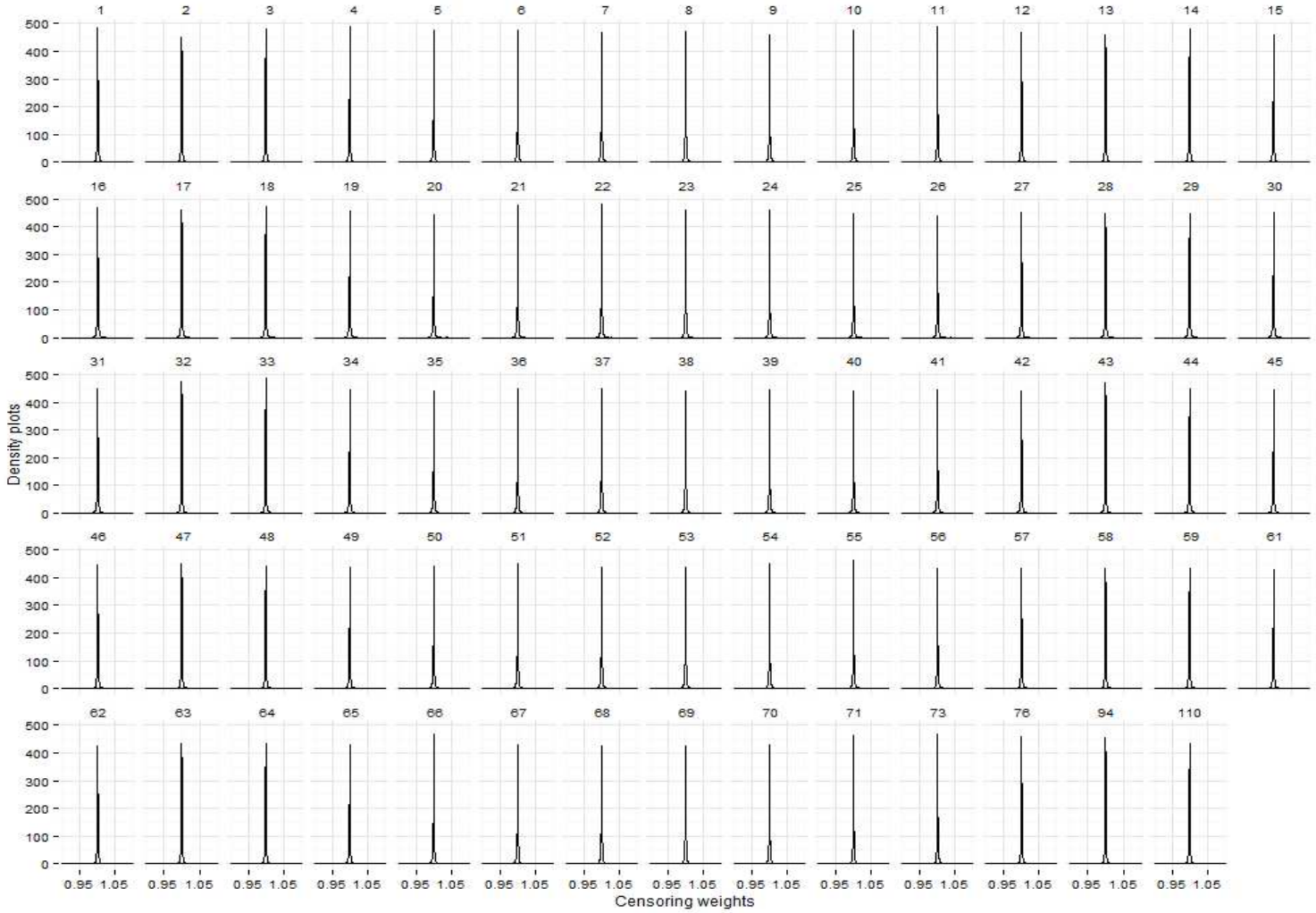
Web-Table I.1: Estimated hazard ratio using the modified sequential Cox approach to estimate the causal effect of β-IFN on time to sustained EDSS 6 for patients with relapsing-onset multiple sclerosis (MS), British Columbia, Canada (1995-2008), when IPCWs are calculated using different approaches.

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.