



## Practice of Epidemiology

### Marginal Structural Cox Models for Estimating the Association Between $\beta$ -Interferon Exposure and Disease Progression in a Multiple Sclerosis Cohort

Mohammad Ehsanul Karim\*, Paul Gustafson, John Petkau, Yinshan Zhao, Afsaneh Shirani, Elaine Kingwell, Charity Evans, Mia van der Kop, Joel Oger, and Helen Tremlett

\* Correspondence to: Mohammad Ehsanul Karim, Department of Statistics, Faculty of Science, University of British Columbia, 3182 Earth Sciences Building, Vancouver, British Columbia V6T 1Z4, Canada (e-mail: ehsan@stat.ubc.ca).

Initially submitted September 24, 2013; accepted for publication April 17, 2014.

Longitudinal observational data are required to assess the association between exposure to  $\beta$ -interferon medications and disease progression among relapsing-remitting multiple sclerosis (MS) patients in the “real-world” clinical practice setting. Marginal structural Cox models (MSCMs) can provide distinct advantages over traditional approaches by allowing adjustment for time-varying confounders such as MS relapses, as well as baseline characteristics, through the use of inverse probability weighting. We assessed the suitability of MSCMs to analyze data from a large cohort of 1,697 relapsing-remitting MS patients in British Columbia, Canada (1995–2008). In the context of this observational study, which spanned more than a decade and involved patients with a chronic yet fluctuating disease, the recently proposed “normalized stabilized” weights were found to be the most appropriate choice of weights. Using this model, no association between  $\beta$ -interferon exposure and the hazard of disability progression was found (hazard ratio = 1.36, 95% confidence interval: 0.95, 1.94). For sensitivity analyses, truncated normalized unstabilized weights were used in additional MSCMs and to construct inverse probability weight-adjusted survival curves; the findings did not change. Additionally, qualitatively similar conclusions from approximation approaches to the weighted Cox model (i.e., MSCM) extend confidence in the findings.

bias (epidemiology); causality; confounding factors (epidemiology); epidemiologic methods; inverse probability weighting; marginal structural Cox model; multiple sclerosis; survival analysis

Abbreviations: EDSS, Expanded Disability Status Scale; HR, hazard ratio;  $\beta$ -IFN,  $\beta$ -interferon; IPC, inverse probability of censoring; IPT, inverse probability of treatment; IPTC, inverse probability of treatment and censoring; MS, multiple sclerosis; MSCM, marginal structural Cox model.

Multiple sclerosis (MS) is a disease associated with damage to the myelin and nerve fibers in the brain and spinal cord. It is a lifelong disease, typically manifesting in early adulthood, affecting an estimated 2.0–2.5 million people worldwide (1). A relapsing-remitting course is the most common presenting MS phenotype; these patients can experience periods of acute worsening, known as an attack or relapse, followed by relapse-free periods with partial or full recovery. Disability may gradually worsen over time, ultimately becoming irreversible. As evident from various clinical trials, immunomodulatory drugs such as  $\beta$ -interferon ( $\beta$ -IFN) may reduce the risk of an MS relapse and increase the duration of relapse-

free periods over the short term (2–6). However, their impact on longer-term outcomes such as irreversible disability is unclear.

There is a real need to determine whether  $\beta$ -IFNs positively influence the course of MS disease over the long term, particularly in the “real-world” clinical practice setting. Observational studies are the most pragmatic means of addressing this need. However, findings from recent observational studies have been contradictory with respect to the impact of  $\beta$ -IFN (7–9). Possible explanations for these inconsistencies include selection bias, immortal time bias, and inappropriate use of analytical tools (10, 11). Hence, the

association between  $\beta$ -IFN and the progression of disability in clinical practice remains undetermined.

Recently, researchers assessed the association of  $\beta$ -IFN with time to irreversible disability outcomes among relapsing-remitting MS patients treated in the real-world clinical practice setting of British Columbia, Canada, using a Cox model with time-dependent treatment exposure (9). They compared  $\beta$ -IFN-treated patients with 2 separate control cohorts—a “historical” cohort (patients who first became  $\beta$ -IFN-eligible prior to the approval of  $\beta$ -IFN in Canada in 1995) and a “contemporary” cohort (patients who first became  $\beta$ -IFN-eligible after the approval of  $\beta$ -IFN but remained unexposed to  $\beta$ -IFN). While this approach represented a considerable improvement over previous studies (12), concern remained about the potential for indication bias when the contemporary control cohort was considered (9). Despite adjustment for a number of baseline characteristics, there were also concerns raised about the inability to adjust for subsequent (postbaseline) treatment decisions (9, 13–16). Furthermore, since disease activity, such as relapses, can drive decision-making with respect to starting or stopping  $\beta$ -IFN treatment (17) and might also be associated with the outcome (18), relapses could be considered a potential time-dependent confounder. Simply incorporating such confounders into a time-dependent Cox model as covariates may be inadequate to adjust for selection bias and confounding (19).

Marginal structural Cox models (MSCMs) allow estimation of the causal associations between treatment exposure and survival responses (e.g., time to disability) in the presence of time-dependent confounding and selection bias (19, 20). These models depend on model-based estimates of the inverse probability of the observed treatment and censoring status of each patient to achieve causal interpretation of the findings. Simulation studies with short-term follow-up have repeatedly shown that MSCMs are advantageous in terms of obtaining consistent estimates of associations with time-varying treatment exposures (21–24). When studying MS, a chronic disease, extended observation periods are needed, which may contribute to the construction of highly variable weights (25) and subsequently may lead to an inefficient estimate of the causal association. Furthermore, how robust these models are when follow-up lengths differ for individual patients, as is the case in clinical practice, is largely unknown. To assess and address these practical challenges, we explored the use of different weighting approaches in MSCMs to estimate the causal association between  $\beta$ -IFN and time to irreversible MS disability in a cohort of relapsing-remitting MS patients from British Columbia, Canada.

## METHODS

### Study population and measurements

This cohort study included data that were collected prospectively from MS patients who were registered at a British Columbia MS clinic and were eligible to receive  $\beta$ -IFN (all preparations of  $\beta$ -IFN were considered as 1 therapeutic class). In Canada, the first  $\beta$ -IFN was licensed for clinical use in July 1995. Therefore, patients who became eligible

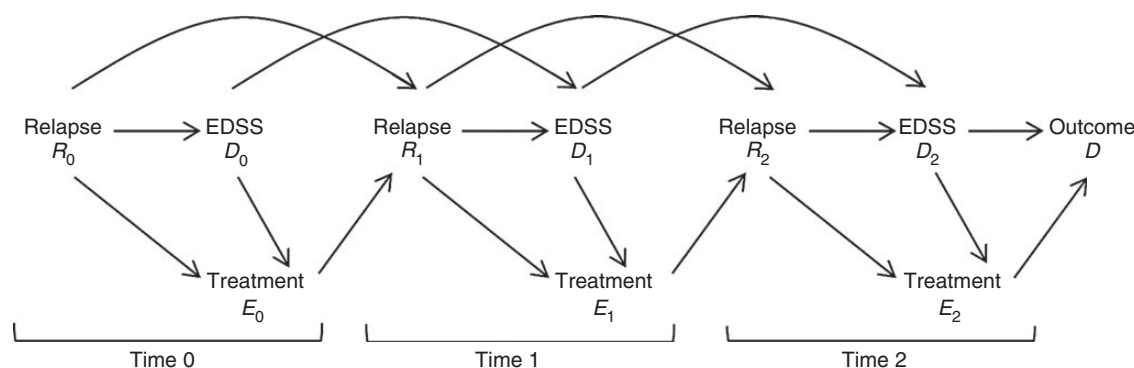
for  $\beta$ -IFN treatment for the first time between July 1, 1995, and December 31, 2004, were included (only the contemporary control cohort was considered). Broad eligibility criteria for receiving  $\beta$ -IFN treatment were adapted from the British Columbia government's reimbursement scheme—that is, adults (aged  $\geq 18$  years) who had a diagnosis of definite MS with a relapsing-onset course and were able to walk (Expanded Disability Status Scale (EDSS) score  $\leq 6.5$ ). The first MS clinic visit at which a patient met the  $\beta$ -IFN eligibility criteria was considered the patient's baseline date (time = 0). The end of follow-up was December 31, 2008. The study was approved by the University of British Columbia's Clinical Research Ethics Board.

The study outcome (irreversible progression of disability) was based on the EDSS (26), a widely used standardized rating system with scores ranging from 0 (no disability) to 10 (death from MS). Our outcome was time to reaching a sustained EDSS score of 6 (“sustained EDSS 6”). An EDSS score of 6 indicates that the patient requires intermittent or unilateral constant assistance (cane, crutch, or brace) to walk about 100 meters with or without resting. Since it is possible to move back and forth along the EDSS, sustained EDSS 6 (i.e., confirmed after at least 150 days, with all subsequent EDSS scores being 6 or greater) was adopted in this study as an indicator of irreversible disability progression (9, 27, 28).

Since a patient's  $\beta$ -IFN exposure status might change during follow-up, this was considered as a time-dependent variable.  $\beta$ -IFN exposure was defined as “any versus none” on a monthly basis. This could be considered an improvement on the previous study design (9), in which only 1 treatment initiation and 1 termination date were considered for each treated patient. Potential confounders included age at baseline, sex, disease duration at baseline, EDSS score at baseline, and relapses.

The relapse variable was selected to be included in the model as a time-varying factor for the following reasons. Firstly, relapses may be associated with the outcome (disability progression). Studies have shown that early relapses may have a significant impact on later disability progression, even though the strength of this association may diminish with time (28). Secondly,  $\beta$ -IFNs have been shown to reduce relapse rates (2–6); therefore, a patient's relapse status may be affected by prior  $\beta$ -IFN treatment. Thirdly, the presence (or absence) of relapses might influence treatment decisions, that is, the determination of whether to start or stop using a  $\beta$ -IFN. Finally, the risk of a relapse is not constant over time; it typically decreases as the patient's disease duration and age increase (29). Therefore, considering only those relapses that occurred prior to a patient's baseline date may be insufficient. Instead, we considered the cumulative number of relapses in the last 2 years (hereafter called “cumulative relapses”) as a time-dependent confounder.

Cumulative relapses could be an intermediate variable between treatment exposure and disability progression; a simplified version of this hypothesized causal relationship is outlined in Figure 1 (also see Web Appendix 1, available at <http://aje.oxfordjournals.org/>). We also examined whether cumulative relapses were an important predictor of subsequent treatment choices.



**Figure 1.** The hypothesized causal relationships in the treatment of multiple sclerosis, with 3 time points:  $j = 0, 1, 2$ . Here  $E_j$  denotes the binary  $\beta$ -interferon exposure variable that is measured immediately after the time-dependent confounders  $R_j$  (cumulative relapses) and  $D_j$  (disability progression index—that is, Expanded Disability Status Scale (EDSS) score) for the  $j$ th time period. The time-dependent confounder  $R_j$  at time  $j$  is affected by prior treatment  $E_{j-1}$ . According to the causal diagram,  $R_0$  imposes confounding on the  $E_0$ - $D$  relationship (since relapse frequency may dictate the subsequent treatment choice and residual disability left by frequent relapses may accumulate over time, leading to irreversible disability), but  $R_1$  is an intermediate variable for the same relationship (51, 52) (since the prior  $\beta$ -interferon treatment may reduce relapse frequency, which may allow the patient more time to recover from residual disability left by past relapses and may contribute to slower progression of disability over time).

## Statistical methods

**Conventional Cox model.** We defined the model notation as follows. If patient  $i$  was followed from the month of  $\beta$ -IFN eligibility ( $t = 0$ ) to  $T_i$  months with treatment exposure in month  $t$  represented by  $A_{it}$  ( $1 =$  under treatment,  $0 =$  not under treatment), then  $a_{it}$  was the realization of  $A_{it}$ . The patient's baseline covariates were recorded in the vector  $L_{i0}$ , consisting of baseline EDSS score, disease duration, age, and sex. If  $\lambda_i(t | L_{i0})$  was the hazard of reaching sustained EDSS 6 at month  $t$  for patient  $i$  with baseline covariates  $L_{i0}$ , one way to model such data was with the time-dependent Cox proportional hazards model:

$$\lambda_i(t | L_{i0}) = \lambda_0(t) \exp(\beta_1 A_{it} + \beta_2 L_{i0}), \quad (1)$$

where  $\lambda_0(t)$  is the unspecified baseline hazard,  $\beta_2$  is the vector of log hazard ratios for the baseline covariates, and  $\beta_1$  is the log hazard ratio of the patient's current  $\beta$ -IFN status ( $A_{it}$ ). The addition of cumulative relapses ( $L_{it}$ ) as a covariate in this model may have failed to adjust for this time-dependent confounder (see Web Appendix 2). Hence, the MSCM approach (19, 30) was applied instead.

**Marginal structural Cox model.** Within a counterfactual framework, in the pseudopopulation, MSCMs enabled the conceptual comparison of the hazard functions for patients who never received  $\beta$ -IFN (complete nonexposure during follow-up) with those who received  $\beta$ -IFN continuously (complete exposure). To accomplish this, the partial likelihood function of the Cox model (or its approximations; see Web Appendix 3) was modified such that the contribution of patient  $i$  to the risk set at time  $t$  was weighted by the inverse probability of treatment and censoring (IPTC) weight,  $w_{it}$ , to remove the possible confounding effects of both time-varying and baseline confounders (19).

**Weighting schemes.** The stabilized inverse probability of treatment (IPT) weight ( $sw$ ) for patient  $i$  at month  $t$  was given by

$$sw_{it}^T = \prod_{j=0}^t \frac{\text{pr}(A_{ij} = a_{ij} | \bar{A}_{i(j-1)} = \bar{a}_{i(j-1)}, L_{i0} = l_{i0})}{\text{pr}(A_{ij} = a_{ij} | \bar{A}_{i(j-1)} = \bar{a}_{i(j-1)}, L_{i0} = l_{i0}, \bar{L}_{ij} = \bar{l}_{ij})}, \quad (2)$$

where  $\bar{A}_{ij} = \bar{a}_{ij}$  and  $\bar{L}_{ij} = \bar{l}_{ij}$  are the observed treatment history and time-varying confounder history, respectively, from baseline to time  $j$ . The stabilized IPT weights were inversely related to a function of the time-varying confounder cumulative relapses, since this variable appeared only in the denominator of the weights, whereas the baseline covariates were included in both the numerator and the denominator (see equation 2). The weights  $sw_{it}^T$  down-weighted the person-time contributions when cumulative relapses were a strong predictor of treatment status in the subsequent time periods, after controlling for the baseline covariates. Assuming that the denominators of the weight models were correctly specified, these weights created a pseudopopulation in which cumulative relapses no longer predicted subsequent  $\beta$ -IFN treatment status (31). The estimates of the  $\beta$ -IFN treatment association in this pseudopopulation would be the same as those in the original target population (32).

Generally, when the numerator in equation 2 is replaced by 1, these weights become the unstabilized IPT weights,  $w_{it}^T$  (19), which simultaneously control for time-varying covariates and baseline covariates. Unlike MSCMs using stabilized weights, MSCM analyses using unstabilized weights do not need further adjustment for the baseline covariates (32). Such weights also yield consistent causal estimates that are associated with substantial variability (31).

Consistent estimation of  $\beta_1$  from censored data can be achieved by incorporating inverse probability of censoring (IPC) weights in the analysis (33). Using logic similar to that leading to the IPT weights for uncensored patients, the stabilized IPC weight for patient  $i$  at month  $t$  is

$$sw_{it}^C = \prod_{j=0}^t \frac{pr(C_{ij} = 0 | \bar{C}_{i(j-1)} = 0, \bar{A}_{i(j-1)} = \bar{a}_{i(j-1)}, L_{i0} = l_{i0})}{pr(C_{ij} = 0 | \bar{C}_{i(j-1)} = 0, \bar{A}_{i(j-1)} = \bar{a}_{i(j-1)}, L_{i0} = l_{i0}, \bar{L}_{i(j-1)} = \bar{l}_{i(j-1)})}, \quad (3)$$

where  $C_{ij}$  denotes the binary censoring status (taking the value 1 if the  $i$ th patient was censored in the  $j$ th month and 0 otherwise) and  $\bar{C}_{ij} = \bar{c}_{ij}$  is the observed censoring history up to time  $j$ . The overall stabilized IPTC weights,  $sw_{it}$ , are obtained by multiplying  $sw_{it}^T$  by  $sw_{it}^C$  (30, 32).

We applied logistic regression models to estimate the unknown conditional probabilities (appearing in equations 2 and 3) from the data (see Web Appendix 4 and Web Table 1).

The normalized IPTC weights were calculated, normalizing each weight by the mean weight of the corresponding risk set (22):

$$w_{it}^{(n)} = \frac{w_{it} N_t}{\sum_i Y_{it} w_{it}}, \quad sw_{it}^{(n)} = \frac{sw_{it} N_t}{\sum_i Y_{it} sw_{it}}, \quad (4)$$

where  $Y_{it}$  indicates whether patient  $i$  belonged to the risk set at time  $t$ ,  $sw_{it}^{(n)}$  represents the normalized stabilized weight, and  $N_t = \sum_i Y_{it}$  is the total number of patients in the risk set at time  $t$ .

In order to take within-subject correlation (34) into account, confidence intervals based on robust standard errors are usually evaluated, which may be asymptotically conservative (30, 35). Therefore, we calculated the 95% confidence intervals for the causal estimate on the basis of 500 nonparametric bootstrap samples (36–38).

**IPTC-weighted survival estimates.** IPTC weight-adjusted Kaplan-Meier survival curves did not require assumptions related to parametric survival or the Cox model. We used unstabilized IPTC weights ( $w$  or  $w^{(n)}$ ) to adjust the survival curves. This had the added advantage of yielding marginal estimates that provided direct causal interpretations without first requiring fitting of the MSCM model (39); hence, constructing such curves served as a sensitivity analysis. However, these weights can be highly variable compared with  $sw^{(n)}$ , and the adjusted survival curves are prone to distortion in the presence of extreme weights. Truncation of extreme weights was applied as one ad hoc solution to assuage the problem of extreme weights (40).

Sample code and practical guidance on implementing the weights in such direct and approximate MSCM approaches via various R packages (41) (R Foundation for Statistical Computing, Vienna, Austria) are provided in Web Appendix 5.

## RESULTS

Of 1,697 patients included in the study, 1,297 were female (76%). The mean age at baseline was 39.7 years (standard deviation, 9.7), the mean disease duration from symptom onset was 7.0 years (standard deviation, 7.7), and the median EDSS score was 2 (interquartile range, 1–3).

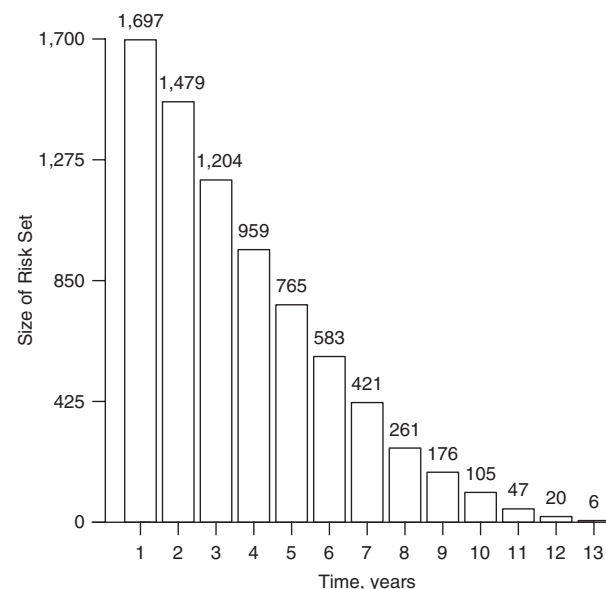
The mean follow-up time was 4.0 years (interquartile range, 1.7–6.0 (4.3 years)), the maximum being 12.7 years.

In total, there were 6,890 person-years of follow-up and 2,530 person-years of  $\beta$ -IFN exposure. In all, 829 patients remained untreated during follow-up. Patients at risk of reaching the outcome at the beginning of each year are shown in Figure 2. Overall, 138 patients reached the outcome of sustained EDSS 6. Further description of the data is provided in Web Appendix 6 (see Web Table 2).

## Time-dependent weights

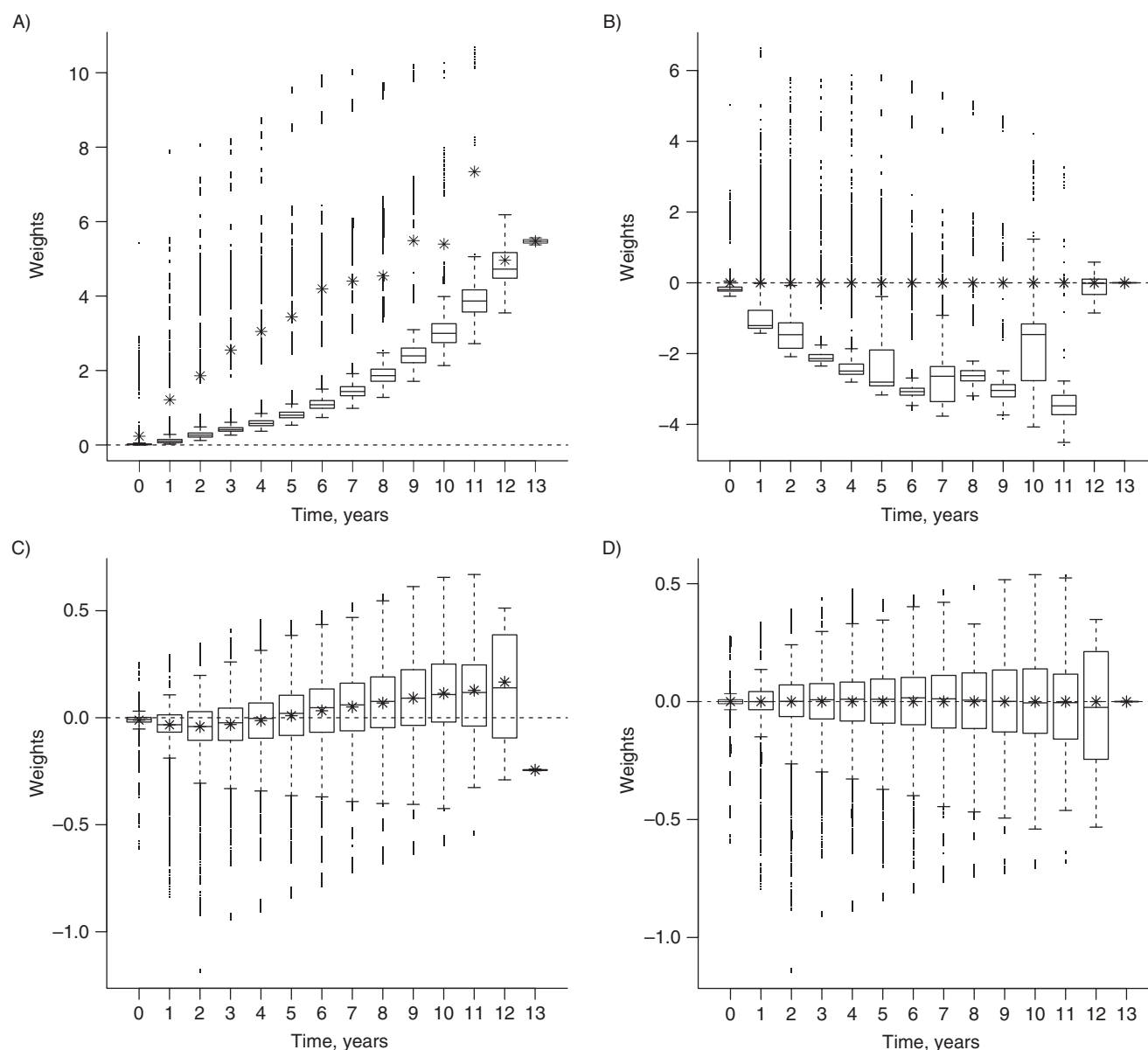
We found the cumulative relapse variable to be a good predictor of subsequent treatment choices, as evidenced by the significance in the model for the IPT weights (2-sided  $P < 0.001$ ; see Web Table 1) and also for the IPC weights (2-sided  $P = 0.03$ ).

The IPTC weights varied not only from patient to patient but also by time. As the number of patients at risk decreased monotonically over time, the variation in the IPTC weights increased with follow-up time. As shown in Figure 3, parts A–D, in addition to such increasing variability, a clear upward trend over time was evident in the unstabilized weights  $w$ . The mean values at successive time points were much closer to 1 after stabilization ( $sw$ ). However, an upward trend in the mean weights was still apparent as follow-up



**Figure 2.** Number of multiple sclerosis patients at risk of reaching a sustained Expanded Disability Status Scale (EDSS) score of 6 during the first month of each follow-up year after baseline, British Columbia, Canada, 1995–2008. Failure to continue to the next risk set results from either censoring or reaching a sustained EDSS score of 6. Analyses were performed by month, but the plot is drawn by year for simplicity.





**Figure 3.** Distribution of various inverse probability of treatment and censoring weighting schemes for each year of follow-up among multiple sclerosis patients from British Columbia, Canada, 1995–2008. A) unstabilized ( $w$ ); B) normalized unstabilized ( $w^{(n)}$ ); C) stabilized ( $sw$ ); D) normalized stabilized ( $sw^{(n)}$ ). The mean value in each boxplot is indicated by an asterisk (\*). Weights are shown on a logarithmic scale. Years were used on the horizontal axes instead of months for better visual display. Note that the plots do not have identical scales on the vertical axes.

progressed. As expected, this trend was eliminated when the stabilized weights were normalized ( $sw^{(n)}$ ) (22). When the unstabilized weights were normalized ( $w^{(n)}$ ), even though the mean weight at each time point was 1, the distributions of the weights were highly variable and skewed.

The mean value and standard deviation of the unstabilized, unnormalized weights ( $w$ ) were much larger than those of the other weights (Table 1), and the resulting causal association estimate was further removed from the null, with a much wider confidence interval. Normalization resulted in

a mean weight of 1 and a markedly reduced standard deviation. Stabilization of the weights had an even greater impact on reducing the standard error of the causal estimate.

A necessary condition for correct model specification is that the mean weight is 1 (42), ideally in each time period rather than just overall. Additionally, a smaller range is an indication of well-behaved weights (40), which generally leads to a smaller confidence interval for the association estimate. In terms of these desirable properties,  $sw^{(n)}$  behaved better than the other schemes: These weights had a smaller range,

**Table 1.** Different Versions of IPTC Weights and Corresponding Causal Estimates of the Association Between  $\beta$ -Interferon Treatment and the Hazard of Reaching a Sustained EDSS Score of 6 Among Multiple Sclerosis Patients From British Columbia, Canada, 1995–2008

Scheme <sup>a</sup>	Stabilized	Normalized	Estimated Weight		Causal Estimate	
			Mean (Log SD)	Range	HR	95% Bootstrap CI <sup>b</sup>
$w$	No	No	28.17 (6.44)	1–43,985.38	1.54	0.09, 26.38
$w^{(n)}$	No	Yes	1 (2.45)	0.01–753.47	1.36	0.18, 10.40
$sw$	Yes	No	0.99 (–2.12)	0.30–1.95	1.36	0.95, 1.94
$sw^{(n)}$	Yes	Yes	1 (–2.18)	0.32–1.71	1.36	0.95, 1.94 <sup>c</sup>

Abbreviations: CI, confidence interval; EDSS, Expanded Disability Status Scale; HR, hazard ratio; IPTC, inverse probability of treatment and censoring; SD, standard deviation.

<sup>a</sup> The inverse probability of treatment numerator model included the baseline covariates EDSS score, age, disease duration, sex, treatment status in the previous time interval, and restricted cubic spline (48) of the follow-up month number. The denominator model included the covariates considered in the numerator model and the time-dependent covariate “cumulative number of relapses for last 2 years,” as well as its interaction with treatment status in the previous time interval. The same model specifications were used to generate the inverse probability of censoring weights. With the stabilized versions of the weights, the hazard ratio model of the marginal structural Cox model must include adjustment for the baseline covariates, but this is not necessary with the unstabilized versions of the weights.

<sup>b</sup> Based on 500 nonparametric bootstrap samples with patients as sampling units.

<sup>c</sup> The 95% CI of the causal association estimate obtained using  $sw^{(n)}$  was the smallest, although it was equal to that obtained using  $sw$  when results were calculated to 2 decimal places.

and there was no tendency for the mean to deviate from 1 even after a long period of follow-up (see Figure 3D). This supported the use of  $sw^{(n)}$  in this application.

### Causal association of $\beta$ -IFN with sustained EDSS score of 6

Since the normalized stabilized weights ( $sw^{(n)}$ ) had better properties, we relied on the corresponding MSCM estimates (see Table 2). The evidence of an association between current  $\beta$ -IFN exposure status and the hazard of reaching a sustained EDSS score of 6 was inconclusive.

To verify the results, we also obtained the estimates from several approaches that approximate the MSCM (see Table 3). All of the estimates from the models based on  $sw^{(n)}$  were consistent. The conclusion concerning the causal association between  $\beta$ -IFN status and time to sustained EDSS score of 6 did not change with the modeling choices.

In a complementary analysis, we considered longitudinal EDSS values as an additional time-varying confounder, instead of treating EDSS score as a baseline covariate (see Table 4). Additionally, we evaluated the impact of weight trimming (43) to assess the sensitivity of the findings to the positivity assumption (see Web Appendix 7). We also repeated the analysis after selecting patients via more restricted eligibility criteria (see Web Appendix 8 and Web Table 3). Further analyses were conducted to check the impact of cumulative exposure to  $\beta$ -IFN over the last 2 years on the same outcome (see Web Appendix 9 and Web Table 4). We also assessed the impact of including cumulative relapses in the last year (see Web Appendix 10 and Web Table 5). None of these sensitivity analyses resulted in statistical evidence for an association with treatment.

### IPTC weighting for estimation of survival curves

We plotted IPTC weight ( $w^{(n)}$ )-adjusted Kaplan-Meier survival curves (see Figure 4, parts A–F). However, the large drops in the survival plot in Figure 4B were driven by only a few large weights. Therefore, we investigated the sensitivity of these adjusted Kaplan-Meier curves after progressively truncating  $w^{(n)}$ .

**Table 2.** Fit of the Marginal Structural Cox Model With the Normalized Stabilized IPTC Weights ( $sw^{(n)}$ ) for Time to a Sustained EDSS Score of 6 (Sustained EDSS 6) in Estimating the Causal Association Between  $\beta$ -Interferon Treatment and the Hazard of Reaching Sustained EDSS 6 Among Multiple Sclerosis Patients From British Columbia, Canada, 1995–2008

Covariate	Estimated Log HR <sup>a</sup>	HR <sup>b</sup>	95% Bootstrap CI <sup>c</sup>
$\beta$ -Interferon use	0.31	1.36	0.95, 1.94
EDSS score	0.54	1.72	1.54, 1.92
Disease duration, decades	–0.19	0.83	0.66, 1.05
Age, decades	0.28	1.32	1.08, 1.62
Sex <sup>d</sup>	–0.22	0.80	0.55, 1.17

Abbreviations: CI, confidence interval; EDSS, Expanded Disability Status Scale; HR, hazard ratio; IPTC, inverse probability of treatment and censoring.

<sup>a</sup> Estimated log HR from a marginal structural Cox model. The model also adjusted for the baseline covariates EDSS score, age, disease duration, and sex.

<sup>b</sup> Instantaneous risk of reaching a confirmed sustained EDSS score of 6.

<sup>c</sup> Based on 500 nonparametric bootstrap samples.

<sup>d</sup> Referent: male.

**Table 3.** Estimates of Association Between  $\beta$ -Interferon Treatment and Time to a Sustained EDSS Score of 6, Obtained Using Different Analytical Approaches, Among Multiple Sclerosis Patients From British Columbia, Canada, 1995–2008

Type of Adjustment, by Model Type	Measure of Association	95% CI
Cox		
Unweighted <sup>a</sup>	1.29 <sup>b</sup>	0.91, 1.82 <sup>c</sup>
Weighted by $sw^{(n)}$	1.36 <sup>b</sup>	0.95, 1.94 <sup>d</sup>
Pooled logistic		
Unweighted <sup>a</sup>	1.29 <sup>e</sup>	0.91, 1.82 <sup>c</sup>
Weighted by $sw^{(n)}$	1.36 <sup>e</sup>	0.96, 1.95 <sup>d</sup>
Poisson		
Weighted by $sw^{(n)}$	1.36 <sup>e</sup>	0.96, 1.95 <sup>d</sup>
Complementary log-log		
Weighted by $sw^{(n)}$	1.37 <sup>e</sup>	0.96, 1.95 <sup>d</sup>

Abbreviations: CI, confidence interval; EDSS, Expanded Disability Status Scale.

<sup>a</sup> Based on time-dependent  $\beta$ -interferon treatment exposure status and covariates measured at baseline: EDSS score, age, disease duration, and sex. This estimate does not have a causal interpretation and is shown for comparison purposes.

<sup>b</sup> The hazard ratio is the measure of association obtained from a Cox model.

<sup>c</sup> 95% CIs were calculated based on robust standard errors.

<sup>d</sup> 95% CIs were obtained from 500 nonparametric bootstrap samples.

<sup>e</sup> The hazard ratio from the Cox model was approximated by the odds ratio from the pooled logistic model (49, 50) (see Web Appendix 3) or, under the infrequent event assumption, by the standardized mortality ratio from Poisson regression or by the odds ratio from complementary log-log regression, respectively. The weighted Cox model (19, 22) was approximated using weighted versions of these models. Software specifications for these analyses are provided in Web Appendix 5.

As can be seen from Figure 4C, truncation of the 5% smallest and largest of the  $w^{(n)}$  freed the curves from the excess influence of a few extreme weights. In this application, the adjusted survival curves did not change dramatically with greater truncation (see Figure 4, parts D–F).

The magnitude of variability in the weights  $w^{(n)}$  affected not only the adjusted survival curve but also the 95% confidence interval for the causal association estimate obtained from the  $w^{(n)}$ -weighted MSCM. The confidence interval (95% bootstrap confidence interval: 0.18, 10.4; see Table 1) was wider than that obtained with  $sw^{(n)}$ , even though the two causal association estimates were the same (hazard ratio (HR) = 1.36). As before, truncation of the extreme weights was examined as another sensitivity analysis to increase the precision of the causal estimate (40). Truncating the 5% smallest and largest of the  $w^{(n)}$  had a substantial impact in this application: The 95% bootstrap confidence interval shrank to (0.64, 1.95) (see Table 5). Table 5 shows that despite improving the precision of the estimate of the  $\beta$ -IFN treatment association, this ad-hoc truncation approach did not alter the conclusion concerning the causal

**Table 4.** Fit of the Marginal Structural Cox Model With the Normalized Stabilized IPTC Weights ( $sw^{(n)}$ ) for Time to a Sustained EDSS Score of 6 (Sustained EDSS 6) in Estimating the Causal Association Between  $\beta$ -Interferon Treatment and Sustained EDSS 6 Among Relapsing-Onset Multiple Sclerosis Patients (Sensitivity Analysis Assessing the Impact of EDSS Score as an Additional Time-Varying Confounder), British Columbia, Canada, 1995–2008

Covariate	Estimated Log HR <sup>a</sup>	HR	95% Bootstrap CI <sup>b</sup>
$\beta$ -Interferon use	0.12	1.13	0.76, 1.68
Disease duration, decades	−0.02	0.98	0.82, 1.22
Age, decades	0.32	1.37	1.10, 1.63
Sex <sup>c</sup>	−0.36	0.70	0.47, 1.02

Abbreviations: CI, confidence interval; EDSS, Expanded Disability Status Scale; HR, hazard ratio; IPTC, inverse probability of treatment and censoring.

<sup>a</sup> The model adjusted for cumulative relapses and EDSS as time-varying confounders and the baseline covariates age, disease duration, and sex. Considering EDSS score as a time-varying confounder rather than a baseline covariate in the analysis does not contradict the causal diagram (Figure 1). All missing EDSS values were imputed via the last-value-carried-forward approach.

<sup>b</sup> Based on 500 nonparametric bootstrap sample estimates.

<sup>c</sup> Referent: male.

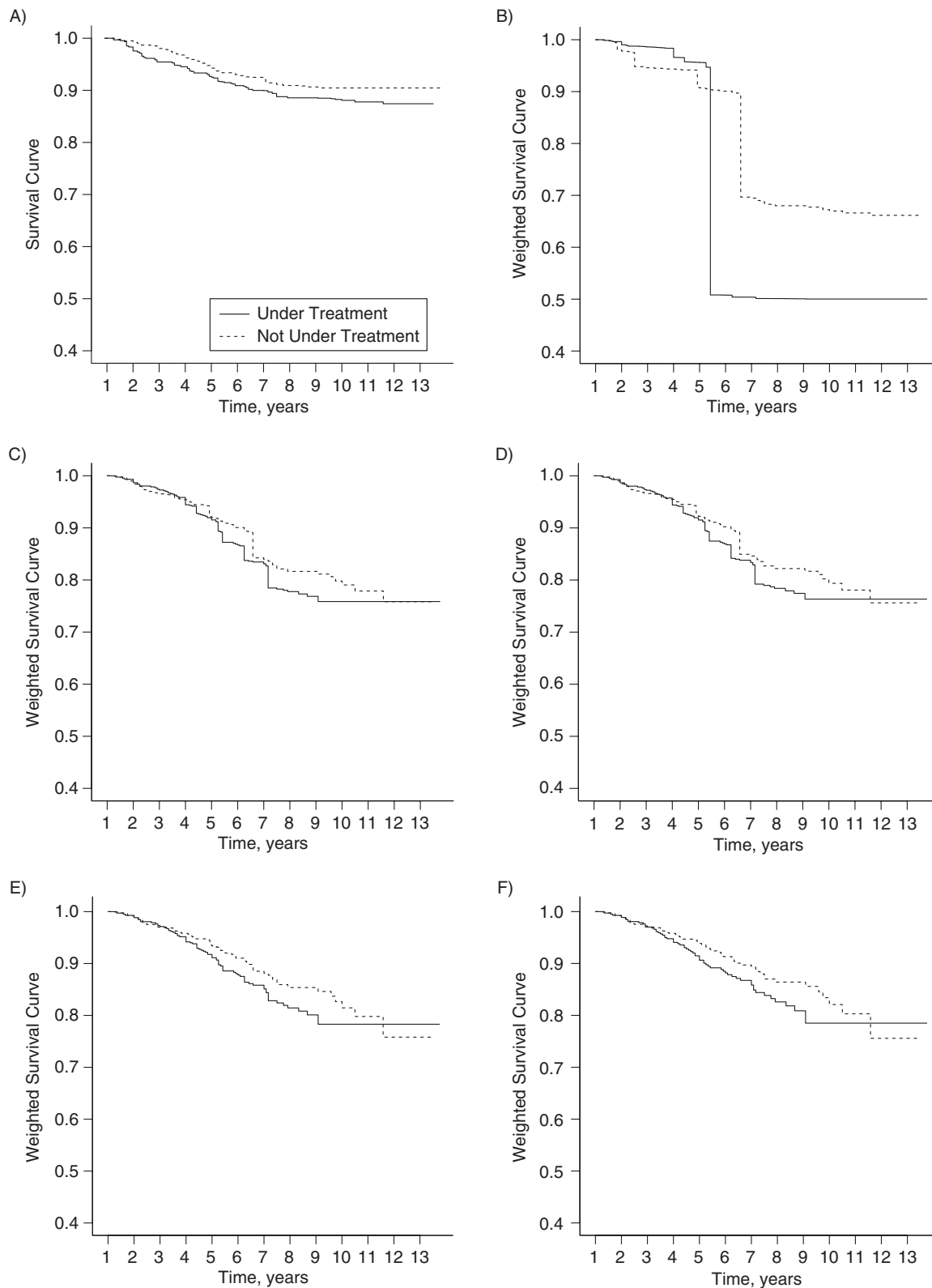
association between  $\beta$ -IFN and time to sustained EDSS score of 6.

## DISCUSSION

When adapting an IPTC weight-based MSCM approach to explore the impact of  $\beta$ -IFN on progression of MS disability in the “real-world” clinical practice setting, we did not find a significant association between  $\beta$ -IFN exposure and MS disability progression.

The possibility that cumulative number of (prior) relapses may represent a time-dependent confounder lying on the causal pathway between  $\beta$ -IFN and disability progression led us to propose this MSCM approach (44). From the analysis, it was evident that the cumulative number of relapses in the previous 2 years was an important factor in the weight models. This highlights the importance of controlling for this type of time-dependent confounder and justifies the additional complexity of the MSCM approach. Further advantages of using such models include the ability to adjust for potential informative censoring.

Even though an extended follow-up period is essential to adequately capture the potential association between treatment and disease progression for chronic diseases such as MS, the duration of follow-up may vary considerably from patient to patient in observational settings. This feature of the data poses considerable challenges while applying the MSCM approach, especially when trying to obtain suitable weights. Over time, treatment exposure and other patient characteristics (e.g., age, disease duration, occurrence of relapses) change, further contributing to the complexity of the study design. To account for these changes, the weights at a given time point need to be



**Figure 4.** Inverse probability of treatment and censoring weight-adjusted Kaplan-Meier-type survival curves for the association between  $\beta$ -interferon treatment and time to a sustained Expanded Disability Status Scale score of 6 among multiple sclerosis patients from British Columbia, Canada, 1995–2008. A) Unweighted; B) adjusted by  $w^{(n)}$  (untruncated); C) adjusted by 5% truncated  $w^{(n)}$ ; D) adjusted by 10% truncated  $w^{(n)}$ ; E) adjusted by 25% truncated  $w^{(n)}$ ; F) adjusted by 50% truncated  $w^{(n)}$ . The truncated weights are derived from the normalized unstabilized weights ( $w^{(n)}$ ) so that the survival probabilities and hazard ratios are marginal estimates with a causal interpretation.



**Table 5.** Impact of Truncation of the Normalized Unstabilized IPTC Weights ( $w^{(n)}$ ) on the Estimated Causal Association Between  $\beta$ -Interferon Treatment and Time to a Sustained EDSS Score of 6 Among Multiple Sclerosis Patients From British Columbia, Canada, 1995–2008

Truncation Percentiles <sup>a</sup>	Estimated Weight		Treatment Association Estimate		
	Mean (Log SD)	Range	HR	SE <sup>b</sup>	95% Bootstrap CI <sup>b</sup>
None	1 (2.45)	0.01–753.47	1.36	1.41	0.18, 10.40
5 and 95	0.31 (–1.24)	0.04–0.93	1.11	0.32	0.64, 1.95
10 and 90	0.3 (–1.29)	0.05–0.83	1.13	0.31	0.66, 1.95
25 and 75	0.21 (–2.2)	0.09–0.35	1.17	0.25	0.77, 1.76
Median <sup>c</sup>	0.19 (– $\infty$ )	0.19–0.19	1.29	0.23	0.91, 1.82

Abbreviations: CI, confidence interval; EDSS, Expanded Disability Status Scale; HR, hazard ratio; IPTC, inverse probability of treatment and censoring; SD, standard deviation; SE, standard error.

<sup>a</sup> Truncation means that the extreme weights (determined by the selected percentile range) are replaced by the nearest percentile weight value.

<sup>b</sup> Based on 500 nonparametric bootstrap samples.

<sup>c</sup> Weighting by the median of the weights produces the same treatment association estimate and 95% CI as those obtained from the simple baseline covariate-adjusted Cox model (see Table 3).

obtained by combining weights for each previous time period in a multiplicative manner. For patients with an extended duration of follow-up, this may cause estimated weights for later periods to increase dramatically and the overall mean weights for these periods to deviate far from 1. In addition, as follow-up progresses, the decreasing number of patients “at risk” may further contribute to high variability in the weights. Deviation from a mean of 1 at any time point is an indication of possible weight model misspecification, whereas highly variable weights may decrease the precision of the causal association estimate (40). Furthermore, in the presence of very large weights, near nonpositivity may result in a biased and imprecise estimate of the treatment association (33, 45). The large variability in the follow-up periods of the MS patients prompted us to investigate the choice of appropriate weighting schemes for MSCMs.

Stabilization of the weights is generally advocated to decrease weight variation and hence increase the precision of MSCM estimates (19). However, the performance of these weights in the chronic disease context has not been well-studied. Here we noted that as the observation period increased, so did the upward trend of the weights. Even though the normalized weights ( $sw^{(n)}$ ) generally possess desirable properties irrespective of the length of the follow-up period (22), we could find no application of these newly proposed weights to the chronic disease context in the published literature. Application of  $sw^{(n)}$  completely eradicated the upward trends, in turn producing an association estimate with slightly higher precision compared with the other weighting schemes, suggesting the potential utility of such weights in studies with longer follow-up.

Adjusting for the time-dependent confounder “cumulative relapses” via IPTC weighting ( $sw^{(n)}$ ) moved the estimated

association for  $\beta$ -IFN treatment (HR = 1.36) away from the null in comparison with the unweighted Cox model (HR = 1.29). The corresponding 95% bootstrap confidence intervals from the MSCM analyses were wider than the 95% robust confidence intervals from the unweighted Cox model, appropriately reflecting more uncertainty as a consequence of using estimated weights. The association estimates were consistent for the various approximations of MSCM models that we considered; none provided evidence of a significant benefit of  $\beta$ -IFN exposure in disease progression.

We also explored the application of other weighting schemes, such as normalized unstabilized weights ( $w^{(n)}$ ). Using these weights, we constructed IPTC-weighted adjusted survival curves. These curves serve as sensitivity analyses, since their results are independent of fitting any MSCM. However, unstable survival estimates were produced as a result of a few very large weights. Moreover, as expected, use of the unstabilized weights resulted in larger standard errors of the MSCM estimators than those obtained from the stabilized versions. The ad-hoc strategy of truncating extreme weights produced more stable survival curves and increased the precision of the MSCM estimate based on  $w^{(n)}$ . Truncation at the 5% level was enough to produce quite stable and smooth survival curves, as well as  $w^{(n)}$ -based MSCM-estimated standard errors comparable to those based on  $sw^{(n)}$ .

This study had limitations. In order to make a causal interpretation from the MSCM results, investigators require identifiability conditions such as positivity, consistency, conditional exchangeability, and correct MSCM model specification (40)—most of which are untestable assumptions. In addition, assuming that the IPTC weight models were correctly specified, truncation of the most extreme weights might have introduced bias into the  $\beta$ -IFN association estimates, reflecting the fundamental “bias-variance tradeoff” (40). Our assessment of disease progression was based on the EDSS, which has recognized limitations (46) and may not be able to distinguish differences due to natural aging from those due to MS disability. Additionally, one could consider EDSS score another time-dependent confounder. Our sensitivity analysis implementing this (based on imputed missing EDSS values) substantially moved the estimated hazard ratio towards the null (HR = 1.13, 95% confidence interval: 0.76, 1.68), considerably weakening the suggestion from the main analysis of a stronger association for treatment in the harmful direction. Therefore, the near-significant point estimate (HR = 1.36, 95% confidence interval: 0.95, 1.94) from the main results may have been due to residual confounding. Although we considered important confounders, residual confounding due to unmeasured covariates (both baseline and time-dependent) is still possible. Potential limitations of the observational study design in assessing the association between  $\beta$ -IFN and MS disease progression are similar to those described elsewhere (9).

In summary, use of the Cox model alone may be inadequate to handle the challenges of analyzing longitudinal observational data. The use of such tools may partly explain the seemingly inconsistent findings regarding the association between  $\beta$ -IFN and disability progression in the “real-world” MS clinical practice setting (8, 9). Here, we carefully implemented the MSCM analysis to adjust for potential indication bias and related changes in patient characteristics which might influence the

subsequent treatment decisions. Our analyses did not find any association between  $\beta$ -IFN exposure and time to development of a sustained EDSS score of 6 over the course of follow-up. Even though different approaches were used here, our conclusions are consistent with those of other studies (9, 47). Furthermore, none of the sensitivity analyses in the current study changed our conclusion regarding the causal association between  $\beta$ -IFN and MS disease progression. The consistency of the results from all of our MSCM analyses strengthens our confidence in the findings. The methods implemented here are adaptable to chronic disease settings beyond MS.

## ACKNOWLEDGMENTS

Author affiliations: Department of Statistics, Faculty of Science, University of British Columbia, Vancouver, British Columbia, Canada (Mohammad Ehsanul Karim, Paul Gustafson, John Petkau); Department of Medicine, Division of Neurology and Brain Research Centre, University of British Columbia, Vancouver, British Columbia, Canada (Yinshan Zhao, Afsaneh Shirani, Elaine Kingwell, Joel Oger, Helen Tremlett); University of Texas Southwestern Medical Center, Dallas, Texas (Afsaneh Shirani); MS/MRI Research Group, University of British Columbia, Vancouver, British Columbia, Canada (Yinshan Zhao); Division of Pharmacy, College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, Saskatchewan, Canada (Charity Evans); and Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden (Mia van der Kop).

This work was supported by a studentship from the Multiple Sclerosis (MS) Society of Canada (Toronto, Ontario, Canada) (awarded to M.E.K.) and by grants from the National MS Society (New York, New York) (grant RG 4202-A-2; Principal Investigator (PI): H.T.) and the Canadian Institutes of Health Research (CIHR) (Ottawa, Ontario, Canada) (grant MOP-93646; PI: H.T.). P.G. was supported by the Natural Sciences and Engineering Research Council of Canada (Ottawa, Ontario, Canada). J.P. held research grants from the CIHR, the National MS Society, and the Natural Sciences and Engineering Research Council of Canada. Y.Z. received research funding from the CIHR, the MS Society of Canada, and the National MS Society. A.S. was supported by a postdoctoral fellowship from the MS Society of Canada and by grants from the CIHR (grant MOP-93646; PI: H.T.) and the National MS Society (grant RG 4202-A-2; PI: H.T.). E.K. was supported by postdoctoral fellowships from the Michael Smith Foundation for Health Research (Vancouver, British Columbia, Canada) and the MS Society of Canada. C.E. was supported by grants from the CIHR (grant MOP-93646; PI: H.T.) and the National MS Society (grant RG 4202-A-2; PI: H.T.) and by the Michael Smith Foundation for Health Research. M.v.d.K. was supported by a CIHR Doctoral Award–Doctoral Foreign Study Award (October 2012), offered in partnership with the CIHR Strategy for Patient-Oriented Research and the CIHR HIV/AIDS Research Initiative. J.O. received support from the Christopher Foundation (Vancouver, British Columbia, Canada) and the University of British Columbia (Vancouver, British Columbia, Canada). H.T. was supported by the Canada

Research Chair Program and by an MS Society of Canada Don Paty Career Development Award and was a Michael Smith Foundation for Health Research Scholar. She also received research support from the National MS Society, the CIHR, and the MS Trust (Letchworth, United Kingdom).

We gratefully acknowledge the neurologists of the various British Columbia MS clinics (listed below) who contributed to the study through patient examination and data collection. We also thank Dr. P. Rieckmann (Sozialstiftung Bamberg Hospital, Bamberg, Germany) for helpful revisions of the original CIHR grant.

British Columbia MS clinic neurologists who contributed to the study: *University of British Columbia (UBC) MS Clinic*—Drs. A. Traboulsee (director of the UBC MS Clinic and head of the UBC MS Research Program), A.-L. Sayao, V. Devonshire, S. Hashimoto (UBC and Victoria MS clinics), J. Hooge (UBC and Prince George MS clinics), L. Kastrukoff (UBC and Prince George MS clinics), and J. Oger; *Kelowna MS Clinic*—Drs. D. Adams, D. Craig, and S. Meckling; *Prince George MS Clinic*—Dr. L. Daly; *Victoria MS Clinic*—Drs. O. Hrebicek, D. Parton, and K. Pope.

The views expressed in this paper do not necessarily reflect the views of any individual acknowledged. No one received compensation for their role in the study.

M.E.K. has had travel and accommodation costs for conference presentations covered by the endMS Research and Training Network (MS Society of Canada) (2011, 2012) and workshop attendance costs covered by the Pacific Institute for the Mathematical Sciences (Vancouver, British Columbia, Canada) (2013). Over the past 3 years, J.P. has received consulting fees and/or fees for service on Data Safety Monitoring Boards from Bayer AG (Leverkusen, Germany), BTG International Ltd. (London, United Kingdom), EMD Serono, Inc. (Mississauga, Ontario, Canada), Merck Serono (Darmstadt, Germany), the Myelin Repair Foundation (Saratoga, California), and Novartis International AG (Basel, Switzerland). A.S. has received travel grants for conference attendance and presentations from the endMS Research and Training Network (2010, 2011), the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS) (Basel, Switzerland) (2010, 2011), and the Consortium of MS Centers (Hackensack, New Jersey) (2012). E.K. has had travel and accommodation costs for conference attendance and presentations covered by the endMS Research and Training Network (2008, 2011), the International Society for Pharmacoepidemiology (Bethesda, Maryland) (2010, 2013), and Bayer (2010). C.E. has received travel grants for conference attendance and presentations from the endMS Research and Training Network (2011) and ECTRIMS (2011). Over the past 5 years, J.O. has received speaker honoraria, consulting fees, and grants for travel, research, and/or education from Sanofi-Aventis (Bridgewater, New Jersey), Bayer, Biogen Idec (Cambridge, Massachusetts), BioMS Medical Corporation (Edmonton, Alberta, Canada), Corixa Corporation (Marietta, Pennsylvania), Genentech (South San Francisco, California), Novartis, EMD Serono, Bayer, Talecris Biotherapeutics (Research Triangle Park, North Carolina), and Teva Neuroscience, Inc. (Kansas City, Missouri). J.O. also receives fees for service on advisory committees from Bayer,

Novartis, and Biogen Idec. H.T. has received speaker honoraria and/or travel expenses for conference attendance from the Consortium of MS Centers (2013), the MS Society of Canada (2013), the National MS Society (2012), the UBC MS Research Program, Bayer HealthCare Pharmaceuticals (San Francisco, California) (speaker, 2010; honoraria declined), Teva Pharmaceuticals (Petach Tikva, Israel) (speaker, 2011),ECTRIMS (2011, 2012, 2013), the MS Trust (2011), the Chesapeake Health Education Program, US Department of Veterans Affairs (Baltimore, Maryland) (2012; honorarium declined), Novartis Pharmaceuticals Canada Inc. (Dorval, Quebec, Canada) (2012), and Biogen Idec (2014; honorarium declined) (unless otherwise stated, all speaker honoraria are either donated to an MS charity or added to an unrestricted grant for use by H.T.'s research group). P.G., Y.Z., and M.v.d.K. declare no conflicts of interest.

## REFERENCES

- World Health Organization and Multiple Sclerosis International Federation. *Atlas: Multiple Sclerosis Resources in the World 2008*. Geneva, Switzerland: World Health Organization; 2008.
- IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology*. 1993;43(4):655–661.
- IFNB Multiple Sclerosis Study Group and University of British Columbia MS/MRI Analysis Group. Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Neurology*. 1995;45(7):1277–1285.
- Multiple Sclerosis Collaborative Research Group. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Ann Neurol*. 1996;39(3):285–294.
- PRISMS (Prevention of Relapses and Disability by Interferon  $\beta$ -1a Subcutaneously in Multiple Sclerosis) Study Group. Randomised double-blind placebo-controlled study of interferon  $\beta$ -1a in relapsing/remitting multiple sclerosis. *Lancet*. 1998;352(9139):1498–1504.
- Once Weekly Interferon for MS Study Group. Evidence of interferon  $\beta$ -1a dose response in relapsing-remitting MS: the OWIMS Study. *Neurology*. 1999;53(4):679–686.
- Brown MG, Kirby S, Skedgel C, et al. How effective are disease-modifying drugs in delaying progression in relapsing-onset MS? *Neurology*. 2007;69(15):1498–1507.
- Trojan M, Pellegrini F, Fuiani A, et al. New natural history of interferon- $\beta$ -treated relapsing multiple sclerosis. *Ann Neurol*. 2007;61(4):300–306.
- 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. *JAMA*. 2012;308(3):247–256.
- Renoux C, Suissa S. Immortal time bias in the study of effectiveness of interferon- $\beta$  in multiple sclerosis. *Ann Neurol*. 2008;64(1):109–110.
- Koch M, Mostert J, De Keyser J, et al. Interferon- $\beta$  treatment and the natural history of relapsing-remitting multiple sclerosis. *Ann Neurol*. 2008;63(1):125–126.
- Derfuss T, Kappos L. Evaluating the potential benefit of interferon treatment in multiple sclerosis. *JAMA*. 2012;308(3):290–291.
- Goodin DS, Reder AT, Cutter G. Treatment with interferon beta for multiple sclerosis [letter]. *JAMA*. 2012;308(16):1627–1628.
- Shirani A, Petkau J, Tremlett H. Treatment with interferon beta for multiple sclerosis—reply [letter]. *J Am Med Assoc*. 2012;308(16):1627–1628.
- Greenberg BM, Balcer L, Calabresi PA, et al. Interferon beta use and disability prevention in relapsing-remitting multiple sclerosis. *JAMA Neurol*. 2013;70(2):248–251.
- Shirani A, Zhao Y, Karim ME, et al. Interferon beta and long-term disability in multiple sclerosis. *JAMA Neurol*. 2013;70(5):651–653.
- Coles A. Multiple sclerosis: the bare essentials. *Pract Neurol*. 2009;9(2):118–126.
- 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(6):835–844.
- Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11(5):561–570.
- Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Stat Assoc*. 2001;96(454):440–448.
- 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(1):71–84.
- Xiao Y, Abrahamowicz M, Moodie EE. Accuracy of conventional and marginal structural Cox model estimators: a simulation study. *Int J Biostat*. 2010;6(2):Article 13.
- Westreich D, Cole SR, Schisterman EF, et al. A simulation study of finite-sample properties of marginal structural Cox proportional hazards models. *Stat Med*. 2012;31(19):2098–2109.
- Havercroft WG, Didelez V. Simulating from marginal structural models with time-dependent confounding. *Stat Med*. 2012;31(30):4190–4206.
- Xiao Y, Moodie EEM, Abrahamowicz M. Comparison of approaches to weight truncation for marginal structural Cox models. *Epidemiol Methods*. 2013;2(1):1–20.
- Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33(11):1444–1452.
- Tremlett H, Paty D, Devonshire V. Disability progression in multiple sclerosis is slower than previously reported. *Neurology*. 2006;66(2):172–177.
- Tremlett H, Yousefi M, Devonshire V, et al. Impact of multiple sclerosis relapses on progression diminishes with time. *Neurology*. 2009;73(20):1616–1623.
- Tremlett H, Zhao Y, Joseph J, et al. Relapses in multiple sclerosis are age- and time-dependent. *J Neurol Neurosurg Psychiatry*. 2008;79(12):1368–1374.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–560.
- Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry D, eds. *Statistical Models in Epidemiology, the Environment and Clinical Trials*. New York, NY: Springer-Verlag; 1999:95–134.
- Robins JM. Association, causation, and marginal structural models. *Synthese*. 1999;121(1):151–179.
- Robins JM, Greenland S, Hu FC. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *J Am Stat Assoc*. 1999;94(447):687–700.
- Cook NR, Cole SR, 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(11):1045–1053.



35. Cole SR, Hernán MA, Robins JM, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Am J Epidemiol.* 2003;158(7):687–694.
36. 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(26):2757–2768.
37. McCulloch M, Broffman M, van der Laan M, et al. Lung cancer survival with herbal medicine and vitamins in a whole-systems approach: ten-year follow-up data analyzed with marginal structural models and propensity score methods. *Integr Cancer Ther.* 2011;10(3):260–279.
38. Ali RA, Ali MA, Wei Z. On computing standard errors for marginal structural Cox models. *Lifetime Data Anal.* 2014; 20:106–131.
39. Westreich D, Cole SR, Tien PC, et al. Time scale and adjusted survival curves for marginal structural Cox models. *Am J Epidemiol.* 2010;171(6):691–700.
40. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168(6): 656–664.
41. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2012. (<http://www.R-project.org/>). (Accessed January 1, 2014).
42. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health.* 2006; 60(7):578–586.
43. Platt RW, Delaney JA, Suissa S. The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding. *Eur J Epidemiol.* 2012;27(2): 77–83.
44. van der Wal WM, Noordzij M, Dekker FW, et al. Comparing mortality in renal patients on hemodialysis versus peritoneal dialysis using a marginal structural model. *Int J Biostat.* 2010; 6(1):Article 2.
45. Robins J, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. *Stat Med.* 2008; 27(23):4678–4721.
46. Willoughby EW, Paty DW. Scales for rating impairment in multiple sclerosis: a critique. *Neurology.* 1988;38(11): 1793–1798.
47. Ebers GC, Traboulsee A, Li D, et al. Analysis of clinical outcomes according to original treatment groups 16 years after the pivotal IFNB-1b trial. *J Neurol Neurosurg Psychiatry.* 2010;81(8):907–912.
48. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York, NY: Springer-Verlag; 2001.
49. Thompson WA Jr. On the treatment of grouped observations in life studies. *Biometrics.* 1977;33(3):463–470.
50. D’Agostino RB, Lee ML, Belanger AJ, et al. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Stat Med.* 1990;9(12): 1501–1515.
51. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10(1):37–48.
52. Glymour MM. Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS, eds. *Methods in Social Epidemiology.* San Francisco, CA: Jossey-Bass/Wiley; 2006:393–428.