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

Mohammad Ehsanul Karim [*1], **Paul Gustafson**[1], **John Petkau**[1], **Yinshan Zhao**[2,3], **Afsaneh Shirani**[2], **Elaine Kingwell**[2], **Charity Evans**[4], **Mia van der Kop**[5], **Joel Oger**[2] and **Helen Tremlett**[2]

[1] Department of Statistics, University of British Columbia (UBC), Vancouver, British Columbia, Canada.
[2] Department of Medicine, Division of Neurology and Brain Research Centre, UBC, Vancouver, British Columbia, Canada.
[3] MS/MRI Research Group, UBC, Vancouver, British Columbia, Canada.
[4] College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
[5] Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden.

[*] *Correspondence to Mohammad Ehsanul Karim, Department of Statistics, University of British Columbia, 3182 Earth Sciences Building, 2207 Main Mall, Vancouver, BC V6T 1Z4, Canada (e-mail: ehsan@stat.ubc.ca).*

WEB MATERIAL

Web Appendix 1: Rationale behind hypothesizing that cumulative relapses are lying on the causal path of β-IFN and disability progression

The exact mechanism of action of the β-interferon (β-IFN) drugs in multiple sclerosis (MS) has never been fully established and is one reason why estimating the effect of these drugs in MS is not straightforward. In the absence of randomization, establishing a causal link between drug exposure and outcome requires subject-specific knowledge and careful implementation of that knowledge in the analysis. Suggesting a plausible causal path is the first step.

Relapsing-remitting patients experience relapses followed by periods of remission in which partial or complete recovery occurs. Based on the results from randomized, double-blind, placebo-controlled studies, β-IFN treatments reduced the severity and frequency of relapses (1–5) and hence increased the period between relapses (4). Consequently, a patient has more time to recover from the residual disability left by the past relapse. This extended period of relapse-free time due to β-IFN exposure may eventually contribute to a slower progression of disability (2–4). However, it should be noted that while most natural history studies indicate that long-term there is minimal or no association between relapse rates and disability progression, a specific window of opportunity for relapses to contribute to disease progression may exist (6, 7).

Therefore, we hypothesized that within a short time interval the cumulative relapses are acting as an intermediate variable for the treatment and disability progression relationship, i.e., the relapse frequency is influenced by prior β-IFN treatment and a greater (lesser) relapse frequency will result in faster (slower) disability progression. Also, we assume that the cumulative relapse count in the previous time period is a confounder that may dictate the treatment choice in subsequent time periods. Furthermore, experiencing an increased number of cumulative relapses after initiating treatment will increase the probability of discontinuing treatment (8). Hence, in this relationship, cumulative relapse is treated both as an intermediate variable and a confounder.

The causal path described above could be considered as rather simplistic. It is possible that cumulative relapse and disability progression have an unmeasured common cause (for example, low serum vitamin D levels). Should this data be available, then we would add that variable to the causal path between cumulative relapse and Expanded Disability Status Scale (EDSS) score. Cumulative relapse would still be a time-dependent confounder and would need to be adjusted for accordingly.

Web Appendix 2: Rationale behind using a marginal structural Cox model (MSCM) instead of a Cox model

For a longitudinal study with N patients, let $i = 1, 2,\ldots,N$ be the patient index, $t = 0, 1,\ldots,T_i$ months be the follow-up time index, A_{it} be the binary treatment status at month t (1 = treated, 0 = untreated), and L_{i0} be the baseline covariates of patient i. One possible model would express the hazard function of the time-dependent Cox model as follows:

$$\lambda_i(t \mid L_{i0}) = \lambda_{0t} \exp\left(\beta_1 A_{it} + \beta_2 L_{i0}\right), \qquad (1)$$

where λ_{0t} is the unspecified baseline hazard function, $\boldsymbol{\beta}_2$ is the vector of log hazard ratios (HRs) for the baseline covariates and β_1 is the log HR of the current β-IFN status (A_{it}).

Assuming no tied event times, we estimate $\beta = (\beta_1,\beta_2)$ by maximizing the partial likelihood (9):

$$PL(\beta) = \prod_{i=1}^{N}\prod_{t=0}^{T_i}\left(\frac{Y_{it}\exp\left(\beta_1 A_{it} + \beta_2 L_{i0}\right)}{\sum_{k=1}^{N} Y_{kt}\exp\left(\beta_1 A_{kt} + \beta_2 L_{k0}\right)}\right)^{dN_{it}},$$

where Y_{it} denotes whether patient i belongs to the risk set at time t, N_{it} is the number of events in the interval $[0,t]$ and dN_{it} denotes the number of new events for patient i at month t (increment from month t - 1, if any). This setting is more general than our case, where $N_{it} \leq 1$ and $dN_{it} = 1$ for at most 1 month.

However, ignoring the time-dependent confounder L_{it} (i.e., an intermediate variable lying in the causal pathway of the treatment and the outcome) may lead to a biased estimate of β. Simply including this variable in the Cox model as a covariate as,

$$\lambda_i(t \mid L_{i0}, L_{it}) = \lambda_{0t} \exp\left(\beta_1 A_{it} + \beta_2 L_{i0} + \beta_3 L_{it}\right), \qquad (2)$$

may still produce a biased estimate if L_{it} is influenced by past exposure (10).

Inverse probability of treatment and censoring (IPTC) weights (say, w, sw, $w^{(n)}$, $sw^{(n)}$) are person-time specific measures of the degree to which a time-dependent variable confounds the treatment selection and censoring processes. These are used in the time-dependent Cox model to weight the contribution of each person-time observation so that confounding due to L_{it} is

removed without changing the target parameter. In this way, MSCM facilitates correction for time-dependent confounding. In the MSCM, these IPTC weights are inserted in the partial likelihood function as follows (11–13):

$$PL_w(\beta) = \prod_{i=1}^{N} \prod_{t=0}^{T_i} \left(\frac{Y_{it} \exp\left(\beta_1 A_{it} + \beta_2 L_{i0}\right)}{\sum_{k=1}^{N} Y_{kt} w_{kt} \exp\left(\beta_1 A_{kt} + \beta_2 L_{k0}\right)} \right)^{dN_{it} \times w_{it}}.$$

The gradient with respect to the parameter vector **β** of the log of the weighted partial likelihood $PL_w(\beta)$ yields the score function $U_w(\beta)$. Equating $U_w(\beta)$ to zero yields a set of estimating equations that can be solved using an iterative method such as the Newton-Raphson algorithm or a penalized partial likelihood approach.

Web Appendix 3: Approximation of the marginal structural Cox model

Let D_t be an indicator of reaching EDSS 6 for the first time between the months t - 1 and t. The data for patients who did not reach sustained EDSS 6 and remained uncensored until follow-up month t can be modelled using the pooled logistic regression (logistic regression pooled over persons and times):

$$logit[Pr(D_{it} = 1 \mid D_{i(t-1)} = 0, A_{it}, L_{i0})] = \gamma_0(t) + \gamma_1 A_{it} + \gamma_2 L_{i0}. \qquad (3)$$

Here $\gamma_0(t)$ is a smooth function of the month index t, represented as a restricted cubic spline, which is often used to reduce weight variability. Just as for cubic polynomial regression, use of a restricted cubic spline forces the relationship to be smooth even on the edges (14, chapter 6); see the R code in the Web Appendix 5. The log OR of the current β-IFN status in this pooled logistic regression, γ_1, is generally a good approximation of the corresponding log hazard ratio obtained from the time-dependent Cox model (β_1), provided that censoring is ignorable (15) and relatively short intervals are chosen so that the probability of outcome occurrence in each time interval is small (16, 17). The corresponding likelihood function can be expressed as:

$$L(\gamma) = \prod_{i=1}^{N} \prod_{t=0}^{T_i} p_{it}^{D_{it}} (1 - p_{it})^{(1-D_{it})},$$

where $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ and $logit(p_{it}) = \gamma_0(t) + \gamma_1 A_{it} + \gamma_2 L_{i0}$.

Hernán et al. (10) suggested use of weighted pooled logistic regression to approximate MSCM (IPTC-weighted time-dependent Cox model) estimates of treatment association (β_1) and others have followed this suggestion. (18–22). The weighted likelihood function is then written as (15):

$$L_w(\gamma) = \prod_{i=1}^{N} \prod_{t=0}^{T_i} \left(p_{it}^{D_{it}} (1 - p_{it})^{(1-D_{it})} \right)^{w_{it}}.$$

This approximate approach was suggested mainly because software available at that time was unable to handle patient-specific time-varying weights in a Cox model. It has been noted that this approximation approach is inadequate when the event is not rare (23). Subsequently Xiao et al. (24) suggested the direct use of the Cox model weighted by IPTC weights to overcome this limitation. Through simulation, these authors also showed that direct use of the Cox model weighted by IPTC weights instead of any approximate MSCM approach (10) considerably reduced the variability of the estimated treatment association, even when both methods use the same weights.

Web Appendix 4: Weight models

The stabilized inverse probability of treatment (IPT) weights for patient *i* at month *t* are expressed as:

$$sw_{it}^{T} = \prod_{j=0}^{t} \frac{pr(A_{ij} = a_{ij} \mid \bar{A}_{i(j-1)} = \bar{a}_{i(j-1)}, L_{i0} = l_{i0})}{pr(A_{ij} = a_{ij} \mid \bar{A}_{i(j-1)} = \bar{a}_{i(j-1)}, L_{i0} = l_{i0}, \bar{L}_{ij} = \bar{l}_{ij})}. \qquad (4)$$

The probability appearing in the numerator of sw^{T} is modeled using a pooled logistic model as follows:

$$logit[\, pr(A_{ij} \mid \bar{A}_{i(j-1)}, L_{i0})] = \alpha_0(j) + \alpha_1 A_{i(j-1)} + \alpha_2 L_{i0}, \qquad (5)$$

where treatment status at the previous time interval (A_{j-1}; A_{-1} = 0 for all patients), the baseline covariates (L_0; in our application, EDSS, age, disease duration, sex) and a restricted cubic spline of the follow-up month index are included as predictors. These covariates, as well as the time-varying confounder cumulative relapse (L_{ij}) and its interaction with prior treatment status are included in the denominator model:

$$logit[\, pr(A_{ij} \mid \bar{A}_{i(j-1)}, L_{i0}, \bar{L}_{ij})] = \alpha_0(j) + \alpha_1 A_{i(j-1)} + \alpha_2 L_{i0} + \alpha_3 L_{ij} + \alpha_{13} A_{i(j-1)} L_{ij}. \qquad (6)$$

The output of this fit is reported in Web Table 1.

Web Table 1: Estimated coefficients from the treatment model (denominator of sw_{it}^{T}) for patients with relapsing-onset multiple sclerosis (MS), British Columbia, Canada (1995–2008)

	Estimate	*Z* Value	*P* Value
β-IFN$_{j-1}$	9.78	102.92	< 0.001
EDSS[a]	0.12	4.31	< 0.001
Age[a,b]	-0.07	-1.70	0.09
Disease duration[a,b]	-0.17	-3.17	< 0.001
Sex[a]	-0.07	-0.96	0.34
Cumulative relapse	0.34	7.70	< 0.001
Cumulative relapse:β-IFN$_{j-1}$	-0.55	-10.83	< 0.001

EDSS, Expanded Disability Status Scale.

[*] Time index is also fitted with restricted cubic spline, but the corresponding coefficients are not reported in the table.

[a] Baseline covariates (L_0).

[b] Expressed in decades.

The predicted value from the (denominator) model (.6) yields the estimated probability of the patient's treatment status in that month *t*. Since the exposure status may vary from one time point to another, first we estimate the probability of the observed treatment status at each time point, and then obtain the probability of the observed exposure sequence of a given patient by multiplying the corresponding probabilities. The numerator of sw_{it}^{T} is estimated in a similar fashion from model (.5), where L_{ij} is not included as a predictor. Dividing the numerator model probabilities of the patient's observed treatment status a_{ij} (either 0 or 1) by the corresponding denominator model probabilities yields the estimated IPT weights sw_{it}^{T} that account for the confounding due to L_{ij}, given the required assumptions are met.

To estimate the IPTC weights $sw_{it} = sw_{it}^{T} \times sw_{it}^{C}$, the inverse probability of censoring (IPC) weights sw_{it}^{C} are estimated in the same fashion. In order to produce the normalized IPTC weights $sw^{(n)}$, each weight *sw* is divided by its risk set's mean weight.

Web Appendix 5: MSCM fitting in R

For time-dependent survival analysis, all person-time observations are pooled to make an augmented data set. Short intervals, such as months, are chosen so that the most recently observed changes of the time-varying variables can be updated in a new row in the data set to reflect the patient's time-varying status with respect to covariates, censoring and response. In the longitudinal analysis literature, this is referred to as the 'long' format.

Guidelines regarding IPTC weight calculations in R are available in the literature (21). These IPTC weights can be viewed as a generalization of the Horvitz-Thompson estimator (25–27). Recently, due to the availability of packages for the analysis of complex surveys in standard software (SAS, Stata, and R), it is possible to fit the time-dependent IPTC-weighted Cox model directly or via approximation, say, using the weighted pooled logistic model. In all the model choices, reliable SEs can be obtained from a reasonable number of patient-specific bootstrap samples.

- Most MSCM analyses in the literature use weighted pooled logistic regression to approximate the IPTC-weighted Cox model fit. In R, performing weighted pooled logistic regression using the `glm` function from the `base` package (with `log` link) is straightforward (21).
- Similarly, the `svyglm` function from the `survey` package can be used to implement the (weighted) pooled logistic model (26).
- With data organized in person-month format, to perform survival analysis using the weighted Cox model, we used the Andersen-Gill's counting process approach as implemented in the `svycoxph` function from the R package `survey` (28) with the `weights` option. Approximation via complementary-log-log and Poisson models can also be implemented using the same package. A sample code follows:

```
require(survey)
require(rms)
(weighted.design<-svydesign(id=~ID, data=long.format,
  weight=~normalized.stabilized.weight))
svycoxph(Surv(start, stop, event) ~ drug + covariate.list,

  design=weighted.design)
svyglm(event ~ drug + rcs(Time) + covariate.list,
  family=binomial(link=log),  design=weighted.design)
svyglm(event ~ drug + rcs(Time) + covariate.list,
  family=binomial(link=cloglog),  design=weighted.design)
svyglm(event ~ offset(log(stop-start))+ drug + rcs(Time) +

  covariate.list, family=poisson(),  design=weighted.design
)
```

- Alternatively, the `coxph` function from the `survival` package (29) can be used to fit the weighted Cox model (24). To handle correlated observations, the `cluster` option must be specified to identify the person-month observations from the same patient. Robust SEs are obtained by specifying the option `robust = TRUE`.

Web Appendix 6: Exclusion criteria and summary of selected cohorts

In total, 2,671 patients met the eligibility criteria to receive β-IFN treatment between July 1995 and December 2004 (31). 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 of the non-β-IFN treatment. Other 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 is provided in Web Table 2.

Web Table 2: Characteristics of the selected cohort of patients with relapsing-onset multiple sclerosis (MS), British Columbia, Canada (1995–2008).

Characteristics	β-IFN exposed patients	β-IFN unexposed patients
Frequency	868	829
Women, *n* (%)	660 (76.0)	637 (76.8)
Disease duration (at baseline)	5.8[a] (6.6[b])	8.3[a] (8.5[b])
Age (at baseline)	38.1[a] (9.2[b])	41.3[a] (10.0[b])
EDSS score (at baseline)	2.0[c] (0-6.5[d])	2.0[c] (0-6.5[d])
Relapse rate / year (over the 2 years prior to baseline)	0.5 [c] (0-1.2[e])	0.5 [c] (0-1.0[e])
Person-years exposed to β-IFN treatment	2,530	0
Person-years not exposed to β-IFN treatment	1,400	2,960

EDSS, Expanded Disability Status Scale.
[a] Mean.
[b] Standard deviation.
[c] Median.
[d] Range.
[e] Interquartile range (25th–75th percentiles).

Web Appendix 7: Sensitivity analysis: impact of weight trimming

If the weights contain extreme values, one should be concerned about the positivity assumption. The MSCM approach is built on the counterfactual framework and it is necessary to assume patients could choose treatment exposure or non-exposure at any time point. If a group of patients with similar covariate history rarely or never receive treatment, then the estimated probability of being treated would be close to zero. Conversely, if a group of patients with similar covariate history almost always or always receive treatment, then the estimated probability of being treated would be close to one. Then the corresponding fitted probability will be close to zero or one resulting in a very large or small inverse probability weight respectively. This may produce unstable estimates from the MSCM.

As a sensitivity analysis, one could restrict the analysis to the subset of patients that have a probability of treatment and censoring that is reasonably removed from 0 and 1 at every time point. This procedure is known as trimming (30). As with truncation of the weights, systematically excluding such patients may produce a biased estimate. Also, the interpretation may lack generalizability due to this restriction. However, since the patients with extreme weights are removed, a relatively stable point estimate with a smaller CI would be expected.

After estimating the fitted probabilities from the weight models, if the probabilities are such that a few person-time observations are contributing too much in the pseudo-population, this may make the estimate of the causal association unstable. In our sensitivity analysis, we removed the patients with at least one fitted value either greater than 0.95 or less than 0.05 (represented more than 20 times in the pseudo-population). This left 1,603 patients, with 133 reaching the outcome. MSCM using $sw^{(n)}$ led to a HR estimate of 1.33 with a 95% bootstrap CI of 0.94 - 1.89. The conclusion regarding the treatment association between β-IFN and time to sustained EDSS 6 from these results remained the same.

Web Appendix 8: Sensitivity analysis: impact of more restrictive eligibility criteria

As another sensitivity analysis, a more restricted study sample was selected by defining active disease (two or more documented relapses during the two years prior to baseline) as part of the eligibility criteria, while also including all the previous criteria. This left 747 patients in the study with 3028 person-years of follow-up and 1460 person-years of β-IFN exposure. Only 52 of these patients reached the irreversible disease outcome.

The model fit is reported in Web Table 3. The regression coefficients and HR estimates were qualitatively similar to those reported in Table 2. The CIs from this restricted data set were wider due to the smaller sample size. Still, the conclusion regarding the treatment association between β-IFN and time to sustained EDSS 6 remained the same as before.

Web Table 3: The marginal structural Cox model (MSCM) fit with the normalized stabilized IPTC weights $sw^{(n)}$ for time to sustained EDSS 6 to estimate the causal association between β-IFN treatment for patients with relapsing-onset multiple sclerosis (MS), British Columbia, Canada (1995–2008) selected by more restrictive eligibility criteria. The model was also adjusted for baseline covariates EDSS, age, disease duration and sex.

Covariate	Estimate[a]	HR [b]	95% CI [c]
β-IFN	0.18	1.19	0.68 - 2.11
EDSS	0.40	1.48	1.24 - 1.77[d]
Disease duration[e]	-0.14	0.87	0.55 - 1.37
Age[e]	0.45	1.57	1.14 - 2.18[d]
Sex[f]	-0.33	0.72	0.38 - 1.35

CI, confidence interval; EDSS, Expanded Disability Status Scale; HR, hazard ratio.
[a] Estimated log HR: negative value is indicative of a beneficial association and positive value is indicative of a harmful association.
[b] HR, indicating the instantaneous risk of reaching sustained and confirmed EDSS 6.
[c] Based on 500 nonparametric bootstrap sample estimates.
[d] 95% CI that does not include 1.
[e] Expressed in decades.
[f] Referent: male.

Web Appendix 9: Sensitivity analysis: impact of the cumulative exposure to β-IFN

We also assessed the impact of the cumulative exposure to β-IFN (proportion of months exposed) over the last two years on time to sustained EDSS 6. The model fit is reported in Web Table 4. This analysis also failed to detect a significant association between the cumulative exposure to β-IFN and the hazard of reaching sustained EDSS 6. A similar finding was observed when the cumulative exposure was restricted to the past year only (data not shown).

Web Table 4: The marginal structural Cox model (MSCM) fit with the normalized stabilized IPTC weights $sw^{(n)}$ for time to sustained EDSS 6 to estimate the causal association of cumulative exposure to β-IFN over the last two years for patients with relapsing-onset multiple sclerosis (MS), British Columbia, Canada (1995–2008). The model was also adjusted for baseline covariates EDSS, age, disease duration and sex.

Covariate	Estimate	HR [b]	95% CI [c]
Cumulative β-IFN[a]	0.53	1.70	0.64 - 4.53
EDSS	0.54	1.71	1.53 - 1.91 [d]
Disease duration[e]	-0.20	0.82	0.66 - 1.10
Age[e]	0.30	1.34	1.10 - 1.63 [d]
Sex[f]	-0.23	0.79	0.55 - 1.15

CI, confidence interval; EDSS, Expanded Disability Status Scale; HR, hazard ratio.
[a] Expressed as proportion of months exposed over last two years.
[b] HR, indicating the instantaneous risk of reaching sustained and confirmed EDSS 6.
[c] Based on 500 nonparametric bootstrap sample estimates.
[d] 95% CI that does not include 1.
[e] Expressed in decades.
[f] Referent: male.

Web Appendix 10: Sensitivity analysis: impact of the cumulative number of relapses in the last year

We also assessed the impact of the exposure to β-IFN on time to sustained EDSS 6 while considering the cumulative number of relapses in the last year (instead of the last two years) as the time-varying confounder. The model fit is reported in Web Table 5. This analysis also failed to detect a significant association between the exposure to β-IFN and the hazard of reaching sustained EDSS 6.

Web Table 5: The marginal structural Cox model (MSCM) fit with the normalized stabilized IPTC weights $sw^{(n)}$ for time to sustained EDSS 6 to estimate the causal association of exposure to β-IFN for patients with relapsing-onset multiple sclerosis (MS), British Columbia, Canada (1995–2008) while considering the cumulative number of relapses in the last year as the time-varying confounder. The model was also adjusted for baseline covariates EDSS, age, disease duration and sex.

Covariate	Estimate	HR [a]	95% CI [b]
β-IFN	0.31	1.36	0.96 - 1.92
EDSS	0.54	1.72	1.54 - 1.92 [c]
Disease duration[d]	-0.18	0.82	0.66 - 1.04
Age[d]	0.28	1.32	1.10 - 1.60 [c]
Sex[e]	-0.22	0.80	0.55 - 1.16

CI, confidence interval; EDSS, Expanded Disability Status Scale; HR, hazard ratio.
[a] HR, indicating the instantaneous risk of reaching sustained and confirmed EDSS 6.
[b] Based on 500 nonparametric bootstrap sample estimates.
[c] 95% CI that does not include 1.
[d] Expressed in decades.
[e] Referent: male.

WEB REFERENCES

[1] INFB 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*, 43(4):655–661, 1993.

[2] Jacobs L.D., Cookfair D.L., Rudick R.A., Herndon R.M., Richert J.R., Salazar A.M., Fischer J.S., Goodkin D.E., Granger C.V., Simon J.H., et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Annals of Neurology*, 39(3):285–294, 1996.

[3] Simon JH, Jacobs LD, Campion M., Wende K., Simonian N., Cookfair DL, Rudick R., Herndon R., Richert J., Salazar A., et al. The Multiple Sclerosis Collaborative Research Group. Magnetic resonance studies of intramuscular interferon beta-1a for relapsing multiple sclerosis. *Annals of Neurology*, 43(1):79–87, 1998.

[4] Ebers, G.C. and 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. *The Lancet*, 352(9139):1498–1504, 1998.

[5] Freedman M. and the OWIMS Study Group. Evidence of interferon beta-1a dose response in relapsing–remitting MS: the OWIMS Study. *Neurology*, 53(4): 679–686, 1999.

[6] Tremlett H., Zhao Y., Joseph J., and Devonshire V. Relapses in multiple sclerosis are age- and time-dependent. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(12):1368–1374, 2008.

[7] Tremlett H., Yousefi M., Devonshire V., Rieckmann P., and Zhao Y. Impact of multiple sclerosis relapses on progression diminishes with time. *Neurology*, 73(20):1616–1623, 2009.

[8] Coles A. Multiple sclerosis: The bare essentials. *Neurology in Practice*, 9(2): 118–126, 2009.

[9] Gill R.D. Understanding Cox's regression model: a martingale approach. *Journal of the American Statistical Association*, 79(386):441–447, 1984.

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

[11] Therneau T.M. Extending the Cox Model. Technical report, Section of Biostatistics, Mayo Clinic, Rochester, 1998. URL http://mayoresearch.mayo.edu/mayo/research/biostat/upload/58.pdf.

[12] Cole S.R., Hudgens M.G., Tien P.C., Anastos K., Kingsley L., Chmiel J.S., and Jacobson L.P. Marginal structural models for case-cohort study designs to estimate the association of antiretroviral therapy initiation with incident AIDS or death. *American Journal of Epidemiology*, 175(5):381–390, 2012.

[13] Howe C.J., Cole S.R., Mehta S.H., and Kirk G.D. Estimating the effects of multiple time-varying exposures using joint marginal structural models: alcohol consumption, injection drug use, and HIV acquisition. *Epidemiology*, 23(4): 574–582, 2012.

[14] Harrell F.E. *Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, 2001.

[15] Cole S.R., Jacobson L.P., Tien P.C., Kingsley L., Chmiel J.S., and Anastos K. Using marginal structural measurement-error models to estimate the long-term effect of antiretroviral therapy on incident AIDS or death. *American Journal of Epidemiology*, 171(1):113–122, 2010.

[16] Thompson Jr W.A. On the treatment of grouped observations in life studies. *Biometrics*, 33(3):463–470, 1977.

[17] D'Agostino R.B., Lee M.L., Belanger A.J., Cupples L.A., Anderson K., and Kannel W.B. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Statistics in Medicine*, 9(12): 1501–1515, 1990.

[18] Choi H.K., Hernán M.A., Seeger J.D., Robins J.M., and Wolfe F. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. *The Lancet*, 359(9313):1173–1177, 2002.

[19] Cole S.R., Hernán M.A., Robins J.M., Anastos K., Chmiel J., Detels R., Ervin C., Feldman J., Greenblatt R., Kingsley L., Lai S., Young M., Cohen M., and Muñoz A. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *American Journal of Epidemiology*, 158(7): 687–694, 2003.

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

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

[22] Sterne J.A.C., Hernán M.A., Ledergerber B., Tilling K., Weber R., Sendi P., Rickenbach M., Robins J.M., and Egger M. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *The Lancet*, 366(9483):378–384, 2005.

[23] Young J.G., Hernán M.A., Picciotto S., and Robins J.M. 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.

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

[25] Horvitz D.G. and Thompson D.J. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

[26] Coffman D.L., Caldwell L.L., and Smith E.A. Introducing the at-risk average causal effect with application to HealthWise South Africa. *Prevention Science*, 13(4):437–447, 2012.

[27] Coffman D.L. and Zhong W. Assessing mediation using marginal structural models in the presence of confounding and moderation. *Psychological Methods*, 17(4):642–664, 2012.

[28] Lumley T. *Survey: Analysis of Complex Survey Samples.* 2011. R package version 3.26.

[29] Therneau T. and Lumley T. *survival: Survival analysis, including penalised likelihood.*, 2011. R package version 2.36-5.

[30] Platt R.W., Delaney J.A.C., and Suissa S. The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding. *European Journal of Epidemiology*, 27(2):77–83, 2012.

[31] Shirani A., Zhao Y., Karim M.E., Evans C., Kingwell E., van der Kop M., Oger J., Gustafson P., Petkau J., and Tremlett H. 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.