

Published in final edited form as:

Stat Med. 2014 March 15; 33(6): 1001–1014. doi:10.1002/sim.5994.

Simulation from a known Cox MSM using standard parametric models for the g-formula

Jessica G. Young^{a,*} and Eric J. Tchetgen Tchetgen^{a,b}

^aDepartment of Epidemiology, Harvard School of Public Health, Boston, MA 02115

^bDepartment of Biostatistics, Harvard School of Public Health, Boston, MA 02115

Abstract

It is routinely argued that, unlike standard regression-based estimates, inverse probability weighted (IPW) estimates of the parameters of a correctly specified Cox marginal structural model (MSM) may remain unbiased in the presence of a time-varying confounder affected by prior treatment. Previously proposed methods for simulating from a known Cox MSM lack knowledge of the law of the observed outcome conditional on the measured past. While unbiased IPW estimation does not require this knowledge, standard regression-based estimates rely on correct specification of this law. Thus, in typical high-dimensional settings, such simulation methods cannot isolate bias due to complex time-varying confounding as it may be conflated with bias due to misspecification of the outcome regression model. In this paper, we describe an approach to Cox MSM data generation that allows for a comparison of the bias of IPW estimates versus that of standard regression-based estimates in the complete absence of model misspecification. This approach involves simulating data from a standard parametrization of the likelihood and solving for the underlying Cox MSM. We prove that solutions exist and computations are tractable under many data generating mechanisms. We show analytically and confirm in simulations that, in the absence of model misspecification, the bias of standard regression-based estimates for the parameters of a Cox MSM is indeed a function of the coefficients in observed data models quantifying the presence of a time-varying confounder affected by prior treatment. We discuss limitations of this approach including that implied by the “g-null paradox”.

Keywords

marginal structural models; simulation; g-formula; survival analysis; causal inference; g-null paradox

1. Introduction

Inverse probability weighted (IPW) estimation of Cox Marginal Structural Models (MSMs) [1, 2] is now a popular approach to estimating the causal effect of a time-varying treatment on survival in observational studies. A Cox MSM is a model for the hazard ratio at a given follow-up time comparing counterfactual time-varying treatment regimes. It is routinely argued that, unlike standard regression-based estimates, IPW estimates of the parameters of a correctly specified Cox MSM may remain unbiased in the presence of a time-varying confounder affected by prior treatment. For example, in observational studies of HIV-infected patients, CD4 cell count affects whether a patient will receive treatment and is

associated with future survival. It is also itself affected by whether treatment has been previously initiated.

Young *et al.* [3, 4] and Havercroft and Didelez [5] have proposed algorithms for simulating data under a known Cox MSM and known model for the treatment mechanism. Westreich *et al.* [6] recently applied a variant of one of these algorithms to compare the performance of IPW estimates and standard regression-based estimates of the true Cox MSM parameters under several simulation scenarios where time-varying confounding affected by prior treatment is present. As knowledge of the correct functional form of the Cox MSM and of the model for the treatment mechanism are required for unbiasedness of IPW estimation, these previously proposed approaches are reasonably useful for simulation studies of IPW estimator performance. In particular, under such data generating algorithms, the properties of IPW estimators in the complete absence of model misspecification may be studied.

These previously proposed simulation approaches, however, lack explicit knowledge of the law of the observed outcome at each time conditional on the measured past. Unlike IPW estimates, standard regression-based estimates rely on correct specification of this law. In settings most often of interest, where treatment and confounders are frequently updated over time and/or covariates are high-dimensional, regression-based estimates cannot be constructed non-parametrically and, typically, parametric models are used. It follows that, in such settings, these previous simulation methods will not be useful for studying the performance of standard regression-based estimates as bias due to time-varying confounding may be conflated with bias due to model misspecification.

Xiao *et al.* [7] suggested an alternative approach to simulating from a Cox MSM by generating according to standard parametric models for the joint distribution of the observed data. These authors argued, under a particular data generating mechanism and a rare disease assumption, that the parameters of the underlying Cox MSM may be derived analytically from the parameters of the specified observed data generating models. These include a regression model for the treatment mechanism used in the construction of IPW estimates. These also include a regression model for the law of the outcome conditional on the measured past used in the construction of standard regression-based estimates. In turn, this approach allows a comparison of IPW and standard regression-based estimates in the absence of model misspecification.

In this paper, we show more generally that the parameters of an underlying Cox MSM may be derived based on a particular parametrization of the observed data distribution. This derivation follows from the general relationship between a Cox MSM and Robins' g-formula [8]. We prove that solving for the true Cox MSM parameters is both possible and computationally tractable under many data generating models with or without the assumption of rare disease. Various examples are presented where follow-up time is arbitrary and standard parametric models for the observed data are imposed such that time-varying confounding affected by prior treatment is present. A large sample simulation study is also presented. We begin with a description of the observed data to be generated.

2. Observed data structure

We wish to generate samples of n i.i.d observations where each observation represents measurements on a subject in a hypothetical observational study. In this study, subjects are followed beginning at time t_0 (baseline) and the investigator takes measurements on each subject during frequent, regular intervals. Specifically, for $m = 0, \dots, K$: Let A_m be the value of a binary treatment measured during interval m defined by $[t_m, t_{m+1})$, with $A_m = 1$ if the subject is treated and $A_m = 0$ otherwise. Further, let L_m be a covariate measured at the start

of that same interval. Define $Y_{m+1} = I(T \leq t_{m+1})$ with T a subject's exact failure time which may be either continuous or discrete; equivalently, Y_{m+1} is an indicator of failure by t_{m+1} .

In general, we denote the history of a random variable using overbars; for example $\bar{m} = (A_0, \dots, A_m)$ is the observed treatment history through the end of interval m . By definition $Y_0 = 0$ (all subjects must be at risk for failure at baseline). For notational convenience we set L_{-1} and \bar{m}_{-1} to be identically 0. To simplify the presentation, we will assume no loss to follow-up.

3. Definition of a Cox MSM

Let $\bar{K} = (a_0, \dots, a_m, \dots, a_K)$ denote a treatment regime in $\bar{\mathcal{A}}_K$, the support of \bar{K} consisting of all possible treatment regimes. Examples include $\bar{1}$ (or "always treat") and $\bar{0}$ (or "never treat"). Treatment regimes of the form \bar{K} are known as *static* regimes in that the treatment received in every future interval by someone following that regime is deterministically known at baseline. By contrast, a *dynamic* treatment regime is one under which treatment at a future time may depend on the values of evolving time-dependent covariates. For example, see [9, 10, 11, 12, 13, 14, 15, 16]. In this paper we limit our attention to causal contrasts involving only static regimes.

Define $\bar{Y}_{K+1}^{\bar{a}}$ as the outcome history a subject would have had if, possibly contrary to fact, she followed regime \bar{K} with T her exact failure time. A discrete time Cox MSM $\gamma(m, \bar{m}, \psi)$ is defined as

$$\frac{\Pr[Y_{m+1}^{\bar{a}}=1|Y_m^{\bar{a}}=0]}{\Pr[Y_{m+1}^{\bar{0}}=1|Y_m^{\bar{0}}=0]} = \exp\{\gamma(m, \bar{a}_m, \psi)\} \quad (1)$$

where ψ is a constant parameter vector, and γ is a particular function of ψ , regime \bar{K} through m and m . The parameter ψ encodes the causal treatment effect of following a static regime compared with 0 up to any interval $m + 1$ such that $\psi = 0$ if and only if

$$\Pr[Y_{m+1}^{\bar{a}}=1|Y_m^{\bar{a}}=0] = \Pr[Y_{m+1}^{\bar{0}}=1|Y_m^{\bar{0}}=0] \text{ for all } \bar{m} \text{ in the support of } \bar{K}, m = 0, \dots, K.$$

Following D'agostino *et al.* [17], when T is continuous and measurement intervals are sufficiently small such that the event rate is negligible within each interval and over the follow-up, then the discrete time Cox MSM (1) may approximate the continuous time Cox MSM

$$\lambda_{T\bar{a}}(t) = \lambda_0(t) \exp\{\gamma(t, \bar{a}_t, \psi)\}$$

for $t \in [t_m, t_{m+1})$ where $\lambda_{T\bar{a}}(t)$ and $\lambda_0(t)$ are the counterfactual hazards at t under regimes \bar{K} and "never treat", respectively, for all \bar{K} and $m = 0, \dots, K$.

4. Identifying assumptions and the g-formula

Suppose the goal of this hypothetical study is to obtain an unbiased estimate of the parameter vector ψ under model (1). Note that, if $\gamma(m, \bar{m}, \psi)$ is a saturated model – i.e., the ratio on the LHS of (1) is allowed to differ for every possible \bar{m} and m – then the Cox MSM

is by definition correctly specified. As $\bar{Y}_{K+1}^{\bar{a}}$ for all \bar{K} in the support of $\bar{\mathcal{A}}_K$ are not observed for all study subjects, in order to identify ψ based only on measured variables, we require additional assumptions. For each $m = 0, \dots, K$ and each static regime \bar{K} , suppose that the following hold:

1. Consistency: If $m = m$ then $\bar{Y}_{m+1} = \bar{Y}_{m+1}^{\bar{a}}$ and $\bar{L}_m = \bar{L}_m^{\bar{a}}$ with $\bar{L}_m^{\bar{a}}$ the covariate history through m under \bar{a} .
2. Positivity: $f_{m-1, L_m, Y_m}(m-1, l_m, 0) > 0 \Rightarrow \Pr[A_m = a_m | L_m = l_m, m-1 = m-1, Y_m = 0] > 0$ w.p.1.
3. Exchangeability (no unmeasured confounding):

$$(Y_{m+1}^{\bar{a}}, \dots, Y_{K+1}^{\bar{a}}) \perp\!\!\!\perp A_m | \bar{L}_m = \bar{l}_m, \bar{A}_{m-1} = \bar{a}_{m-1}, Y_m = 0$$

where, in general, $A \perp\!\!\!\perp B/C$ denotes “A is independent of B given C”.

As stated in the appendix of [3], under the above three identifying assumptions for a given static regime \bar{a} , $\Pr[Y_{m+1}^{\bar{a}} = 1 | Y_m^{\bar{a}} = 0]$ is given by Robins’ g-formula [8]:

$$h(m, \bar{a}_m) = \sum_{\bar{l}_m} \Pr[Y_{m+1} = 1 | \bar{L}_m = \bar{l}_m, \bar{A}_m = \bar{a}_m, Y_m = 0] w_{\bar{a}}(m, \bar{l}_m) / \sum_{\bar{l}_m} w_{\bar{a}}(m, \bar{l}_m) \quad (2)$$

with

$$w_{\bar{a}}(m, \bar{l}_m) = \prod_{j=0}^m \Pr[Y_j = 0 | \bar{L}_{j-1} = \bar{l}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}, Y_{j-1} = 0] f(l_j | \bar{a}_{j-1}, \bar{l}_{j-1}, Y_j = 0)$$

for $m = 0, \dots, K$.

It follows that, given our identifying assumptions, the Cox MSM (1) will hold in our study population if the following relationship holds

$$\frac{h(m, \bar{a}_m)}{h(m, \bar{0}_m)} = \exp\{\gamma(m, \bar{a}_m, \psi)\} \quad (3)$$

for all m and $m = 0, \dots, K$. For simplicity, we have generally expressed the g-formula $h(m, \bar{a}_m)$ above in terms of a high-dimensional sum. However, when L_m contains continuously measured components, we may replace sums with integrals.

5. Parametric assumptions on the g-formula

Let us now additionally assume that the components of the g-formula $h(m, \bar{a}_m)$ may be characterized by standard parametric models. Under these additional restrictions, it follows that the Cox MSM (1) holds if

$$\frac{h(m, \bar{a}_m; \beta, \theta)}{h(m, \bar{0}_m; \beta, \theta)} = \exp\{\gamma(m, \bar{a}_m, \psi)\} \quad (4)$$

for all m and $m = 0, \dots, K$ where

$$h(m, \bar{a}_m; \theta, \beta) = \sum_{\bar{l}_m} \Pr[Y_{m+1} = 1 | \bar{L}_m = \bar{l}_m, \bar{A}_m = \bar{a}_m, Y_m = 0; \theta] w_{\bar{a}}(m, \bar{l}_m; \theta, \beta) / \sum_{\bar{l}_m} w_{\bar{a}}(m, \bar{l}_m; \theta, \beta) \quad (5)$$

with

$$w_{\bar{a}}(m, \bar{l}_m; \theta, \beta) = \prod_{j=0}^m \Pr[Y_j=0 | \bar{L}_{j-1}=\bar{l}_{j-1}, \bar{A}_{j-1}=\bar{a}_{j-1}, Y_{j-1}=0; \theta] f(l_j | \bar{a}_{j-1}, \bar{l}_{j-1}, Y_j=0; \beta)$$

such that (5) is a particular parametric version of (2).

We can now explicitly connect models for the observed data likelihood to a Cox MSM. Specifically, based on (4), when L_m contains only discrete components, one can, at least in theory, derive in closed form the underlying Cox MSM $\gamma(m, m, \psi)$ that holds for any choice of parametric models. However, this derivation may become computationally unwieldy in practice without additional restrictions (e.g. Markov assumptions) when the confounders may take on many levels and/or K is large.

When L_m contains continuously measured components, deriving the true Cox MSM based on a particular parametrization of the g-formula may require evaluating integrals with no closed form. In this case, whether there is a closed form solution will depend on the choice of parametrization and, possibly, whether additional restrictions are imposed (e.g. rare disease assumptions). In the following sections, we consider a variety of data generating assumptions under which we may tractably derive the true Cox MSM implied by a standard parametrization of the observed data likelihood.

6. Cox MSMs under Markov assumptions

Suppose the following Markov assumptions hold:

$$\Pr[Y_{m+1}=1 | \bar{L}_m=\bar{l}_m, \bar{A}_m=\bar{a}_m, Y_m=0] = g(l_m, a_m, a_{m-1}) \quad (6)$$

and

$$f(l_m | \bar{a}_{m-1}, \bar{l}_{m-1}, Y_m=0) = r(l_m, a_{m-1}) \quad (7)$$

$m = 0, \dots, K$ where g and r are any real-valued functions bounded between 0 and 1. We now have the following theorem.

Theorem 6.1 Assume that the restrictions (6) and (7) hold for all $m = 0, \dots, K$. Then, it

follows that the hazard ratio $\frac{h(m, \bar{a}_m)}{h(m, \bar{0}_m)}$ only depends on (m, m) through (a_m, a_{m-1}) with

$$\frac{h(m, \bar{a}_m)}{h(m, \bar{0}_m)} = \frac{\sum_{l_m} g(l_m, a_m, a_{m-1}) r(l_m, a_{m-1})}{\sum_{l_m} g(l_m, 0, 0) r(l_m, 0)} \quad (8)$$

for L_m discrete and

$$\frac{h(m, \bar{a}_m)}{h(m, \bar{0}_m)} = \frac{\int_{-\infty}^{\infty} g(l_m, a_m, a_{m-1}) r(l_m, a_{m-1}) dl_m}{\int_{-\infty}^{\infty} g(l_m, 0, 0) r(l_m, 0) dl_m} \quad (9)$$

for L_m continuous.

A proof of Theorem 6.1 is given in Appendix A. The following is a corollary of Theorem 6.1.

Corollary 6.2 Suppose the assumptions of Theorem 6.1 hold and denote $h(m, a_m) \equiv h(a_m, a_{m-1})$. Then, the Cox MSM $\gamma(m, a_m, \psi) = \psi_0 a_m + \psi_1 a_{m-1} + \psi_2 a_m a_{m-1}$ holds with

$$\exp(\psi_0) = \frac{h(1, 0)}{h(0, 0)} \quad (10)$$

$$\exp(\psi_1) = \frac{h(0, 1)}{h(0, 0)} \quad (11)$$

$$\exp(\psi_2) = \frac{h(1, 1)}{h(0, 0)} \times \frac{1}{\exp(\psi_0) \exp(\psi_1)} \quad (12)$$

7. Binary covariates

In this section, suppose L_m is binary. By Corollary 6.2, we obtain

$$\exp(\psi_0) = \frac{g(1, 1, 0)r(1, 0) + g(0, 1, 0)\{1 - r(1, 0)\}}{g(1, 0, 0)r(1, 0) + g(0, 0, 0)\{1 - r(1, 0)\}} \quad (13)$$

$$\exp(\psi_1) = \frac{g(1, 0, 1)r(1, 1) + g(0, 0, 1)\{1 - r(1, 1)\}}{g(1, 0, 0)r(1, 0) + g(0, 0, 0)\{1 - r(1, 0)\}} \quad (14)$$

$$\exp(\psi_2) = \frac{g(1, 1, 1)r(1, 1) + g(0, 1, 1)\{1 - r(1, 1)\}}{g(1, 1, 0)r(1, 0) + g(0, 1, 0)\{1 - r(1, 0)\}} \times \frac{g(1, 0, 0)r(1, 0) + g(0, 0, 0)\{1 - r(1, 0)\}}{g(1, 0, 1)r(1, 1) + g(0, 0, 1)\{1 - r(1, 1)\}} \quad (15)$$

Equations (13), (14) and (15) follow simply by plugging the RHS of (8) into (10), (11) and (12) for the appropriate (a_m, a_{m-1}) .

Standard parametric assumptions on $g(l_m, a_m, a_{m-1})$ and $r(l_m, a_{m-1})$ might be regression models with logit, probit or complementary log-log links [18]. To fix ideas, we work through an example under logistic regression models for both $g(l_m, a_m, a_{m-1})$ and $r(l_m, a_{m-1})$ such that

$$r(1, a_{m-1}; \beta) = \frac{\exp(\beta_1 a_{m-1})}{1 + \exp(\beta_1 a_{m-1})} \quad (16)$$

and

$$g(l_m, a_m, a_{m-1}; \theta) = \frac{\exp(\theta_0 + \theta_1 l_m + \theta_2 a_m + \theta_3 a_{m-1})}{1 + \exp(\theta_0 + \theta_1 l_m + \theta_2 a_m + \theta_3 a_{m-1})} \quad (17)$$

$m = 0, \dots, K$. To simplify the presentation, we have implicitly set the intercept in (16) to zero, however a non-zero intercept is easily added. Plugging these choices into (13), (14) and (15) we obtain the solutions

$$\exp(\psi_0) = \frac{\frac{\exp(\theta_1 + \theta_2)}{1 + \exp(\theta_0 + \theta_1 + \theta_2)} + \frac{\exp(\theta_2)}{1 + \exp(\theta_0 + \theta_2)}}{\frac{\exp(\theta_1)}{1 + \exp(\theta_0 + \theta_1)} + \frac{1}{1 + \exp(\theta_0)}} \quad (18)$$

$$\exp(\psi_1) = \frac{\frac{\exp(\theta_3)}{1+\exp(\beta_1)} \left\{ \frac{\exp(\theta_1+\beta_1)}{1+\exp(\theta_0+\theta_1+\theta_3)} + \frac{1}{1+\exp(\theta_0+\theta_3)} \right\}}{\frac{1}{2} \left\{ \frac{\exp(\theta_1)}{1+\exp(\theta_0+\theta_1)} + \frac{1}{1+\exp(\theta_0)} \right\}} \quad (19)$$

$$\exp(\psi_2) = \frac{\left\{ \frac{\exp(\theta_1+\beta_1)}{1+\exp(\theta_0+\theta_1+\theta_3)} + \frac{1}{1+\exp(\theta_0+\theta_2+\theta_3)} \right\} \left\{ \frac{\exp(\theta_1)}{1+\exp(\theta_0+\theta_1)} + \frac{1}{1+\exp(\theta_0)} \right\}}{\left\{ \frac{\exp(\theta_1)}{1+\exp(\theta_0+\theta_1+\theta_2)} + \frac{1}{1+\exp(\theta_0+\theta_2)} \right\} \left\{ \frac{\exp(\theta_1+\beta_1)}{1+\exp(\theta_0+\theta_1+\theta_3)} + \frac{1}{1+\exp(\theta_0+\theta_3)} \right\}} \quad (20)$$

Notably, if the disease were rare as in Xiao *et al.* [7] our solutions simplify considerably. Specifically, given the model (17) and rare disease within each measurement interval and history we have the approximation

$$g(l_m, a_m, a_{m-1}; \theta) \approx \exp(\theta_0 + \theta_1 l_m + \theta_2 a_m + \theta_3 a_{m-1}). \quad (21)$$

Plugging (21) into expressions (13), (14) and (15) in place of (17) we obtain the simplified approximate solutions

$$\psi_0 = \theta_2 \quad (22)$$

$$\psi_1 = \log \left[\frac{\exp(\theta_3) \left\{ \frac{\exp(\theta_1+\beta_1)}{1+\exp(\beta_1)} + \frac{1}{1+\exp(\beta_1)} \right\}}{\frac{1}{2} \{ \exp(\theta_1) + 1 \}} \right] \quad (23)$$

$$\psi_2 = 0 \quad (24)$$

Equations (22) and (23) establish that the parameters of a standard time-dependent Cox regression such as (21) do not generally match those of the Cox MSM. Specifically, the maximum likelihood estimates (MLEs) of θ_2 and θ_3 based on the correctly specified model (17) have bias approximately equal to $\theta_2 - \psi_0$ and $\theta_3 - \psi_1$ for ψ_0 and ψ_1 , respectively. Given the rare disease approximation (21), we can see by expression (22) that we have approximately $\theta_2 - \psi_0 = 0$ for any choice of θ_1 or β_1 . However, by expression (23), we will only have $\theta_3 = \psi_1$ if either θ_1 or β_1 is zero; that is, if L_m is either not a confounder or not itself affected by prior treatment. Without the rare disease assumption, by equations (18) and (19), the bias of the MLEs of θ_2 and θ_3 for ψ_0 and ψ_1 , respectively, depends not only on the values of θ_1 and β_1 but also on the other components of θ .

By equation (24), we also see in this example that, given the rare disease approximation (21), absence of an interaction term between a_m and a_{m-1} in the model (17) also implies no interaction as quantified by ψ_2 . However, by equation (20), in the absence of rare disease, the presence of interaction as quantified by ψ_2 generally depends on β_1 and all components of θ , despite the absence of an interaction term in the model (17). For interested readers, we explicitly consider alternative solutions for ψ_2 in Appendix B when an interaction term between a_m and a_{m-1} is added to the model (17), both with and without a rare disease approximation. Note that solutions for ψ_0 and ψ_1 will remain unchanged (with or without rare disease) under a less restricted model for $g(l_m, a_m, a_{m-1})$ that allows interaction between a_m and a_{m-1} by equations (13) and (14), respectively.

8. Continuous covariates

In this section, suppose L_m is a continuous random variable $m = 0, \dots, K$. Given the assumptions of Theorem 6.1, whether a closed form solution exists for $\frac{h(a_m, a_{m-1})}{h(0, 0)}$ in this setting will now depend on the choice of g and r .

For example, as in Xiao *et al.* [7], assume that L_m is normally distributed given the past with

$$L_m | \bar{A}_{m-1}, \bar{L}_{m-1}, Y_m=0 \sim N(\beta_1 A_{m-1}, \sigma^2) \quad (25)$$

As in (16), an intercept is easily added to (25). With this choice of r , a closed form solution for (9) is not generally available for g the logistic regression model (17). If, however, along with this choice of r , we choose a probit link for g such that

$$g(l_m, a_m, a_{m-1}; \theta) = \Phi(\theta_0 + \theta_1 l_m + \theta_2 a_m + \theta_3 a_{m-1}) \quad (26)$$

with $\Phi(\cdot)$ the CDF of a standard normal, then, following Agresti [18], we have

$$h(m, \bar{a}_m; \beta, \theta) = \Phi\{c(\theta_0 + \theta_2 a_m + \theta_3 a_{m-1} + \theta_1 \beta_1 a_{m-1})\} \quad (27)$$

with $c = \left(\sqrt{1 + \theta_1^2 \sigma^2}\right)^{-1}$. By Corollary 6.2, we then also have that the Cox MSM $\gamma(m, m, \psi) = \psi_0 a_m + \psi_1 a_{m-1} + \psi_2 a_m a_{m-1}$ holds with specifically

$$\exp(\psi_0) = \frac{\Phi\{c(\theta_0 + \theta_2)\}}{\Phi(c\theta_0)}$$

$$\exp(\psi_1) = \frac{\Phi\{c(\theta_0 + \theta_3 + \theta_1 \beta_1)\}}{\Phi(c\theta_0)}$$

$$\exp(\psi_2) = \frac{\Phi\{c(\theta_0 + \theta_2 + \theta_3 + \theta_1 \beta_1)\}}{\Phi\{c(\theta_0 + \theta_2)\}} \times \frac{\Phi(c\theta_0)}{\Phi\{c(\theta_0 + \theta_3 + \theta_1 \beta_1)\}}$$

A similar result is available for the logistic regression model (17) provided r is alternatively defined in terms of the more complex bridge distribution function of Wang and Louis [19]. However, if we further assume rare disease, we may maintain assumption (25) for r along with the model (17) for g and obtain an approximate closed form solution for (9). Specifically, using the approximation (21) for g and model (25) for r we have

$$\begin{aligned}
 h(m, \bar{a}_m; \beta, \theta) &\approx \int_{-\infty}^{\infty} \exp(\theta_0 \\
 &\quad + \theta_1 l_m + \theta_2 a_m + \theta_3 a_{m-1}) r(l_m, a_{m-1}; \beta) dl_m \\
 &= \exp(\theta_0 \\
 &\quad + \theta_2 a_m + \theta_3 a_{m-1}) E[\exp(\theta_1 L_m) | A_{m-1} = a_{m-1}] = \exp(\theta_0 \\
 &\quad + \theta_2 a_m + \theta_3 a_{m-1}) \exp(\theta_1 \beta_1 a_{m-1}) \exp\left(\frac{1}{2} \sigma^2 \theta_1^2\right)
 \end{aligned} \tag{28}$$

with the last equality given by the moment generating function for the normal distribution. Note that, given (21), an analogous solution for $h(m, m; \beta, \theta)$ will exist for any choice of r provided the conditional distribution of L_m has homoscedastic errors by the general property of the moment generating function. By Corollary 6.2, plugging the approximation (28) into (10), (11) and (12) for the appropriate choices of (a_m, a_{m-1}) , the Cox MSM $\gamma(m, m, \psi) = \psi_0 a_m + \psi_1 a_{m-1} + \psi_2 a_m a_{m-1}$ holds with the approximate solutions $\psi_0 = \theta_2$, $\psi_1 = \theta_3 + \theta_1 \beta_1$ and $\psi_2 = 0$.

Analogous to the worked example for binary L_m under the rare disease assumption (21), we see that $\psi_0 = \theta_2$ regardless of the values of θ_1 and β_1 . By contrast, $\psi_1 = \theta_3$ only if θ_1 or β_1 is zero; that is, when L_m is not a confounder affected by prior treatment. As in the binary case, our probit example illustrates more generally that the discrepancy between ψ and the components of θ corresponding to treatment coefficients in the conditional outcome regression model, in addition to β_1 and θ_1 , may also depend on other components of θ .

Of the examples considered thus far, the data generating assumptions of our last example – with r defined by model (25) and g approximated by (21) – are closest to those of the data generating models given in Xiao *et al.* [7]. One key distinction, however, is that Xiao *et al.* [7] allowed the distribution of L_m also to depend on L_{m-1} . Under this weaker assumption, the resulting Cox MSM now depends on the entire treatment history m and not simply the two most recent values (a_m, a_{m-1}) as given by the following theorem.

Theorem 8.1 Assume the data generating mechanism (6) of Theorem 6.1 holds with g approximated by (21). Further assume

$$L_m | \bar{A}_{m-1}, \bar{L}_{m-1}, Y_m = 0 \sim N(\beta_1 A_{m-1} + \beta_2 L_{m-1}, \sigma^2) \tag{29}$$

$m = 0, \dots, K$. Data generated under these assumptions will approximately follow a Cox MSM of the form

$$\gamma(m, \bar{a}_m, \psi) = \psi_0 a_m + \psi_1 a_{m-1} + \sum_{s=1}^{m-1} \psi_{m-s+1} a_{s-1}$$

with $\psi_0 = \theta_2$, $\psi_1 = \theta_3 + \theta_1 \beta_1$ and $\psi_{m-s+1} = \theta_1 \beta_1 \beta_2^{m-s}$, $s = 1, \dots, m - 1$.

A proof of Theorem 8.1 is given in Appendix C. Note that Xiao *et al.* [7] concluded that, under their data generating models, the resulting Cox MSM should only depend on a_m and a_{m-1} for all $m = 0, \dots, K$. Theorem 8.1 appears to contradict this conclusion for $K > 1$. Under these data generating assumptions, IPW estimates constructed based on a Cox MSM that excludes the correct function of $m-2$ should theoretically incur some bias because such a Cox MSM will be misspecified. This is a particular problem when $|\beta_2| \neq 1$.

9. Simulation Algorithm

The following general algorithm may be used to simulate a sample of n i.i.d. observations as described in §2 that follows a particular parametrization of the g-formula.

Let $\Pr[A_m = 1 | L_m = l_m, m-1 = m-1, Y_m = 0; \alpha]$ be a parametric model for the probability of receiving treatment in interval m given survival to m and history $(l_m, m-1)$. For each of $i = 1, \dots, n$ simulated observations, implicitly define $L_{-1,i} \equiv m-1, i \equiv Y_{0,i} = 0$. Then for each observation i :

For $m = 0, \dots, K$:

1. Draw $L_{m,i}$ from some choice of $f(L_m | m-1, L_{m-1}, Y_m = 0; \beta)$ evaluated at the previously generated $(m-1, i, L_{m-1, i})$.
2. Draw $A_{m,i}$ from some choice of $\Pr[A_m = 1 | L_m, m-1, Y_m = 0; \alpha]$ evaluated at previously generated $(m-1, i, L_{m, i})$.
3. Draw $Y_{m+1, i}$ from some choice of $\Pr[Y_{m+1} = 1 | L_m, m, Y_m = 0; \theta]$ evaluated at previously generated $(m, i, L_{m, i})$. If $Y_{m+1, i} = 1$ then this is the last record in the data set for observation i . Otherwise, generate another record for observation i (i.e., go to index $m + 1$).

The above algorithm may be used to confirm theoretical results under any of the data generating assumptions considered above. As a simple illustration, we performed a simulation study where 20,000 samples were generated according to the above algorithm, each with $n = 100,000$ observations and $K = 6$. Data were generated according to the restrictions of Theorem 6.1 with the covariate and outcome generated according to the logistic regression models (16) and (17), respectively. Treatment was generated according to the logistic regression model $\text{logit}[\Pr(A_m = 1 | L_m = l_m, m-1 = m-1, Y_m = 0; \alpha)] = \alpha_0 + \alpha_1 l_m$ for each $m = 0, \dots, K$.

Simulations were conducted under the following six different combinations of (β_1, θ_1) . In all scenarios, we fixed $\alpha_0 = 0.5$, $\alpha_1 = 0.5$, $\theta_0 = -7$, $\theta_2 = -0.8$ and $\theta_3 = 0$. As all components of θ were selected $\neq 0$, the rare disease approximation (21) holds under all six scenarios. Recall that by the analytic results of §7, under all simulation scenarios the Cox MSM $\gamma(m, m, \psi) = \psi_0 a_m + \psi_1 a_{m-1} + \psi_2 a_m a_{m-1}$ holds with approximate values of ψ_0 , ψ_1 and ψ_2 defined as in (22), (23) and (24), respectively.

Table 1 presents the bias of the IPW estimates of ψ_0 and ψ_1 for the true ψ_0 and ψ_1 , respectively, constructed under the correctly specified Cox MSM and model for treatment for the 20,000 runs. We see little bias at $n = 100,000$ in these estimates. See Appendix D for details of the IPW estimation procedure. Table 2 presents the bias of the MLE of θ_2 for $\theta_2 = \psi_0$ constructed under the correctly specified outcome regression model (17). Table 3 presents the bias of the MLE of θ_3 for θ_3 , along with the bias for ψ_1 , also under the correct model (17). Our simulations confirm the analytic results of §7. In particular, we see little bias of the MLE of θ_2 and θ_3 for $\theta_2 = \psi_0$ and θ_3 , respectively, regardless of the values of β_1 and θ_1 . However, we see that the bias of the MLE of θ_3 for ψ_1 approximates $\theta_3 - \psi_1$. As expected, this difference is approximately zero only when either β_1 or θ_1 is zero.

10. Relation to the g-null paradox

As discussed, a limitation of the proposed simulation approach is that, under some data generating assumptions, it may be intractable or impossible to solve for the true Cox MSM parameters. Interestingly, an additional limitation of the proposed simulation approach follows from previous arguments regarding the “the g-null paradox” [20]. These arguments

would suggest that given standard parametrizations of the observed data distribution consistent with the presence of a time-varying confounder affected by prior treatment, it is impossible for the null hypothesis of $\psi = 0$ to hold simultaneously. Our examples allow a careful consideration of this paradox in the current setting.

Consider the data generating assumptions of the simulation study described in §9. As before, under these assumptions, we approximately have $\psi_0 = 0$ if $\theta_2 = 0$ by equation (22). Further, by equation (23), we have $\psi_1 = 0$ if θ_3 is set to

$$\exp(\theta_3) = \frac{\frac{1}{2}\{\exp(\theta_1) + 1\}}{\frac{\exp(\theta_1 + \beta_1)}{1 + \exp(\beta_1)} + \frac{1}{1 + \exp(\beta_1)}}$$

regardless of the values of θ_1 and β_1 . Thus, we have at least one example illustrating that it is mathematically possible to generate data according to standard parametric models such that all components of ψ are zero and a time-varying confounder affected by prior treatment is present.

However, we do not expect such a scenario, where one coefficient is restricted to depend on a function of other coefficients of the data generating mechanism, to occur in nature. This is an example of the faithfulness assumption invoked when causal directed acyclic graphs are used to represent underlying data generating mechanisms [21, 22]. We therefore may be limited to simulation scenarios with the proposed algorithm to unrealistic settings if we wish simultaneously to generate data under the null.

11. Discussion

In this paper, we have illustrated how to derive a closed form Cox MSM given a set of parametric models for the observed data distribution. This gives an approach for simulating from a known Cox MSM using a standard parametrization of the likelihood. In contrast to previously proposed simulation methods, this approach allows a comparison of the performance of IPW and standard regression-based estimators of the effect of a time-varying treatment on survival in the complete absence of model misspecification. This, in turn, allows isolation of any particular source of bias in a simulation study. These sources may include finite sample bias, that due to (known) model misspecification and that due to complex time-varying confounding structures.

We used our analytic results to demonstrate and confirm in an example simulation study that the bias of standard regression-based estimates depends, at least in part, on the degree to which parameters quantifying the presence of a time-varying confounder affected by prior treatment are non-zero. Using analytic results, one may know, prior to undertaking a simulation study, how much large sample bias to expect in such a standard estimate in the absence of model misspecification. Confirmation of these expectations reduces the possibility of coding errors.

Our arguments highlight the importance of clearly defining the target population parameter of interest in any consideration of bias. As discussed, standard estimates will be approximately unbiased in large samples for the coefficients on treatment history in a correctly specified outcome regression model conditional on past treatment and confounders. However, following previous graphical arguments in largely model-free settings [9, 23, 24], these coefficients may fail to have a causal interpretation, even given the identifying assumptions of §4, when L_m is a time-varying confounder affected by prior treatment. Our arguments further highlight the need for careful consideration when imposing

a parsimonious Cox MSM. For example, as we showed, typical assumptions that restrict dependence of the causal hazard ratio on only the most recent values of treatment may be difficult to justify if one is unwilling to make potentially extreme Markov assumptions on the underlying observed data generating process.

Finally, the utility of the proposed approach to known Cox MSM data generation is not limited to simulation-based comparisons of IPW performance versus that of standard regression-based estimators. This approach may also be useful in simulation studies aimed at comparing IPW with other estimators of correctly specified Cox MSM parameters that rely on a correctly specified conditional outcome regression model for optimal performance. These include parametric g-computation [8, 25, 26, 16, 27] as well as double-robust methods [28, 29, 30, 31].

Acknowledgments

The authors thank Miguel Hernán for helpful comments. This work was funded by NIH grants HL080644, R01AI104459-01A1, and 1R21ES019712-01.

Appendix A

We explicitly prove Theorem 6.1 under the discrete case as the proof under the continuous case is identical requiring integrals instead of sums. Incorporating assumptions (6) and (7) into the g-formula (2) we have

$$h(m, \bar{a}_m) = \frac{\sum_{l_m} g(l_m, a_m, a_{m-1}) r(l_m, a_{m-1}) \sum_{\bar{l}_{m-1}} wt(m-1, \bar{l}_{m-1}, \bar{a}_{m-1})}{\sum_{l_m} r(l_m, a_{m-1}) \sum_{\bar{l}_{m-1}} wt(m-1, \bar{l}_{m-1}, \bar{a}_{m-1})} \quad (30)$$

where

$$wt(m-1, \bar{l}_{m-1}, \bar{a}_{m-1}) = \{1 - g(l_{m-1}, a_{m-1}, a_{m-2})\} \prod_{j=0}^{m-1} \{1 - g(l_{j-1}, a_{j-1}, a_{j-2})\} r(l_j, a_{j-1}).$$

The result follows by noting that $\sum_{\bar{l}_{m-1}} wt(m-1, \bar{l}_{m-1}, \bar{a}_{m-1})$ cancels in the numerator and denominator of (30) and $\sum_{l_m} r(l_m, a_{m-1}) = 1$.

Appendix B

Consider the worked example of §7 but where the model (17) is replaced by

$$g(l_m, a_m, a_{m-1}; \theta) = \frac{\exp(\theta_0 + \theta_1 l_m + \theta_2 a_m + \theta_3 a_{m-1} + \theta_4 a_m a_{m-1})}{1 + \exp(\theta_0 + \theta_1 l_m + \theta_2 a_m + \theta_3 a_{m-1} + \theta_4 a_m a_{m-1})}$$

allowing interaction between a_m and a_{m-1} as quantified by θ_4 . Plugging this alternative choice of g into (15), along with the original model (16) for r , we have

$$\exp(\psi_2) = \exp(\theta_4) \frac{\left\{ \frac{\exp(\theta_1 + \beta_1)}{1 + \exp(\theta_0 + \theta_1 + \theta_2 + \theta_3 + \theta_4)} + \frac{1}{1 + \exp(\theta_0 + \theta_2 + \theta_3 + \theta_4)} \right\} \left\{ \frac{\exp(\theta_1)}{1 + \exp(\theta_0 + \theta_1)} + \frac{1}{1 + \exp(\theta_0)} \right\}}{\left\{ \frac{\exp(\theta_1)}{1 + \exp(\theta_0 + \theta_1 + \theta_2)} + \frac{1}{1 + \exp(\theta_0 + \theta_2)} \right\} \left\{ \frac{\exp(\theta_1 + \beta_1)}{1 + \exp(\theta_0 + \theta_1 + \theta_3)} + \frac{1}{1 + \exp(\theta_0 + \theta_3)} \right\}}$$

which is equivalent to (20) when $\theta_4 = 0$.

Given rare disease such that we have the approximation

$$g(l_m, a_m, a_{m-1}; \theta) \approx \exp(\theta_0 + \theta_1 l_m + \theta_2 a_m + \theta_3 a_{m-1} + \theta_4 a_m a_{m-1})$$

we obtain the simplified approximate solution

$$\psi_2 = \theta_4$$

which is equivalent to (24) when $\theta_4 = 0$.

Appendix C

For any $m = 0, \dots, K$ define

$$\frac{b^{\bar{a}}(m)}{b^0(m)} = \frac{E[\exp(\theta_1 L_m) | A_{m-1} = a_{m-1}, L_{m-1}]}{E[\exp(\theta_1 L_m) | A_{m-1} = 0, L_{m-1}]}$$

By (29) and the moment generating function for the normal distribution we have

$$\frac{b^{\bar{a}}(m)}{b^0(m)} = \exp(\theta_1 \beta_1 a_{m-1}) \frac{\exp(\frac{1}{2} \sigma^2 \theta_1^2) \exp(\theta_1 \beta_2 L_{m-1})}{\exp(\frac{1}{2} \sigma^2 \theta_1^2) \exp(\theta_1 \beta_2 L_{m-1})} = \exp(\theta_1 \beta_1 a_{m-1}) \frac{\exp(\theta_1 \beta_1 L_{m-1})}{\exp(\theta_1 \beta_2 L_{m-1})}$$

Next define

$$\frac{b^{\bar{a}}(m-1)}{b^0(m-1)} = \frac{E[b^{\bar{a}}(m) | A_{m-2} = a_{m-2}, L_{m-2}]}{E[b^0(m) | A_{m-2} = 0, L_{m-2}]}$$

By the fact that (29) holds for L_0, \dots, L_K and, again, using the moment generating function for the normal distribution we have

$$\begin{aligned} & \frac{E[b^{\bar{a}}(m) | A_{m-2} = a_{m-2}, L_{m-2}]}{E[b^0(m) | A_{m-2} = 0, L_{m-2}]} \\ &= \exp(\theta_1 \beta_1 a_{m-1}) \frac{E[\exp(\theta_1 \beta_2 L_{m-1}) | A_{m-2} = a_{m-2}, L_{m-2}]}{E[\exp(\theta_1 \beta_2 L_{m-1}) | A_{m-2} = 0, L_{m-2}]} \\ &= \exp(\theta_1 \beta_1 a_{m-1}) \frac{\exp(\frac{1}{2} \sigma^2 \theta_1^2 \beta_2^2) \exp\{\theta_1 \beta_2 (\beta_1 a_{m-2} + \beta_2 L_{m-2})\}}{\exp(\frac{1}{2} \sigma^2 \theta_1^2 \beta_2^2) \exp(\theta_1 \beta_2^2 L_{m-2})} = \exp(\theta_1 \beta_1 a_{m-1}) \exp(\theta_1 \beta_1 \beta_2 a_{m-2}) \frac{\exp(\theta_1 \beta_2^2 L_{m-2})}{\exp(\theta_1 \beta_2^2 L_{m-2})} \end{aligned}$$

Analogously define

$$\frac{b^{\bar{a}}(m-2)}{b^0(m-2)} = \frac{E[b^{\bar{a}}(m-1) | A_{m-3} = a_{m-3}, L_{m-3}]}{E[b^0(m-1) | A_{m-3} = 0, L_{m-3}]} = \exp(\theta_1 \beta_1 a_{m-1}) \exp(\theta_1 \beta_1 \beta_2 a_{m-2}) \exp(\theta_1 \beta_1 \beta_2^2 a_{m-3}) \frac{\exp(\theta_1 \beta_2^3 L_{m-3})}{\exp(\theta_1 \beta_2^3 L_{m-3})}$$

Arguing recursively for any $j = 1, \dots, m-1$

$$\frac{b^{\bar{a}}(j)}{b^{\bar{0}}(j)} = \exp\left(\sum_{s=j}^m \theta_1 \beta_1 \beta_2^{m-s} a_{s-1}\right) \frac{\exp(\theta_1 \beta_2^{m-j+1} L_{j-1})}{\exp(\theta_1 \beta_2^{m-j+1} L_{j-1})}$$

Setting $j = 1$ we have

$$\frac{b^{\bar{a}}(1)}{b^{\bar{0}}(1)} = \exp\left(\sum_{s=1}^m \theta_1 \beta_1 \beta_2^{m-s} a_{s-1}\right) \frac{\exp(\theta_1 \beta_2^m L_0)}{\exp(\theta_1 \beta_2^m L_0)}$$

Further, by $\bar{a}_{-1} \equiv L_{-1} \equiv 0$, we have

$$\frac{E[b^{\bar{a}}(1)|A_{-1}=a_{-1}, L_{-1}]}{E[b^{\bar{0}}(1)|A_{-1}=0, L_{-1}]} = \exp\left(\sum_{s=1}^m \theta_1 \beta_1 \beta_2^{m-s} a_{s-1}\right)$$

By (21) we have that

$$\frac{h(m, \bar{a}_m; \beta, \theta)}{h(m, \bar{0}_m; \beta, \theta)} \approx \exp(\theta_2 a_m + \theta_3 a_{m-1}) \frac{E[b^{\bar{a}}(1)|A_{-1}=a_{-1}, L_{-1}]}{E[b^{\bar{0}}(1)|A_{-1}=0, L_{-1}]} \quad (31)$$

Our result follows by noting that the RHS of (31) is equivalent to exp

$$(\psi_0 a_m + \psi_1 a_{m-1} + \sum_{s=1}^{m-1} \psi_{m-s+1} a_{s-1}) \text{ where } \psi_0 = \theta_2, \psi_1 = \theta_3 + \theta_1 \beta_1 \text{ and } \psi_{m-s+1} = \theta_1 \beta_1 \beta_2^{m-s}, s = 1, \dots, m - 1.$$

Appendix D

A typical implementation of IPW estimation is as follows. Let $(\hat{\phi}, \hat{\psi})$ be the solution to the estimating equation

$$\sum_{\bar{a}_m}^K \sum_{m=0}^n \sum_{i=1}^n U_{i,m}^{\bar{a}}(\phi', \psi', \hat{\alpha}) = 0, \quad (32)$$

with respect to (ϕ', ψ') . Following the appendix of [15] and, again, suppressing the i subscript, define

$$U_m^{\bar{a}}(\phi', \psi', \hat{\alpha}) = [Y_{m+1} - \text{expit}\{w(m; \phi') + \gamma(m, \bar{a}_m, \psi')\}] (1 - Y_m) W_m^{\bar{a}}(\hat{\alpha}) q(m, \bar{a}_m) \prod_{j=0}^m I(A_j = a_j)$$

where $\text{expit}(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)}$, $w(m; \phi')$ is a function of interval m and a parameter vector ϕ' , $q(m, \bar{a}_m)$ is a user-selected vector function of (m, \bar{a}_m) and

$$W_m^{\bar{a}}(\hat{\alpha}) = \frac{1}{\prod_{j=0}^m \Pr[A_j = a_j | \bar{L}_j, \bar{A}_{j-1} = \bar{a}_{j-1}, Y_j = 0; \hat{\alpha}]}, \quad (33)$$

with $\hat{\alpha}$ the MLE of α given the treatment model $\Pr[A_m = a_m | L_m, \bar{a}_{m-1}, Y_m = 0; \alpha]$.

Assume the following holds for all m and m :

1. The Cox MSM (3) holds,
2. $h(m, 0_m) = \exp\{w(m; \phi')\}$ when evaluated at $\phi' = \phi$,
3. the treatment model $\Pr[A_m = a_m | L_{m-1}, Y_m = 0; \alpha]$ is correctly specified and
4. $h(m, m) \approx 0$ such that (given assumptions 1 and 2)

$$h(m, \bar{a}_m) \approx \text{expit}\{w(m; \phi) + \gamma(m, \bar{a}_m, \psi)\}. \quad (34)$$

Given these four assumptions we have

$$E[U_m^{\bar{a}}(\phi, \psi, \alpha)] \approx 0 \quad (35)$$

for all m and m and the IPW estimator $\hat{\psi}$ approximately consistent for ψ and asymptotically normal. The choice of $q(m, m)$ in large samples affects only the efficiency of $\hat{\psi}$.

Note that the approximation (35) follows under these assumptions by the equivalence between the g-formula (2) and the ratio of expectations

$$\frac{E[Y_{m+1}(1 - Y_m)W_m^{\bar{a}} \prod_{j=0}^m I(A_j = a_j)]}{E[(1 - Y_m)W_m^{\bar{a}} \prod_{j=0}^m I(A_j = a_j)]}$$

where $W_m^{\bar{a}}$ is equivalent to $W_m^{\bar{a}}(\hat{a})$ but with the true conditional probability of treatment $\Pr[A_m = a_m | L_{m-1} = L_{m-1}, Y_m = 0]$ replacing its MLE.

A convenient choice of $q(m, m)$ is $\left\{ \frac{\partial w(m; \phi')}{\partial \phi'^T}, \frac{\partial \gamma(m, \bar{a}_m; \psi')}{\partial \psi'^T} \right\}^T$. With this choice an approximate solution to (32) may be obtained with off-the-shelf software by fitting a weighted logistic regression model in a person-time data set of the structure described in §9. This approach was used to construct IPW estimates in the example simulation study also described in §9. Specifically, a logistic regression model was fit in SAS using the LOGISTIC procedure with dependent variable Y_{m+1} and independent variables A_m and A_{m-1} along with a completely flexible function of $m = 0, \dots, 6$. The WEIGHT option was used with stabilized weights to increase efficiency. That is, the weight for observation i at time m was defined by expression (33) under the model used to generate treatment and then

multiplied by an estimate of $\prod_{j=0}^m \Pr[A_j = a_j | \bar{A}_{j-1} = \bar{a}_{j-1}, Y_j = 0]$ with m selected as that observation's treatment history through m . The weight numerator can be considered an implicit component of $q(m, m)$. It is straightforward to confirm that the data generating parameters under all simulation scenarios maintain assumption (34) for all m and m . Code is available upon request.

References

1. Robins, JM. Statistical Models in Epidemiology. New York: Springer; 1999. Marginal structural models versus structural nested models as tools for causal inference; p. 95-133.
2. 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. [PubMed: 10955409]

3. Young, JG.; Hernán, MA.; Picciotto, S.; Robins, JM. JSM Proceedings, Section on Statistics in Epidemiology. Alexandria, VA: American Statistical Association; 2008. Simulation from structural survival models under complex time-varying data structures.
4. Young JG, Hernán MA, Picciotto S, Robins JM. Equivalence between structural models for the effect of a time-varying exposure on survival. *Lifetime Data Analysis*. 2010; 16(1):71–84. [PubMed: 19894116]
5. Havercroft WG, Didelez V. Simulating from marginal structural models with time-dependent confounding. *Statistics in Medicine*. 2012; 31(30):4190–4206. [PubMed: 22826156]
6. Westreich D, Cole SR, Schisterman EF, Platt RW. A simulation study of finite-sample properties of marginal structural cox proportional hazards models. *Statistics in Medicine*. 2012; 31(19):2098–2109. [PubMed: 22492660]
7. Xiao Y, Abrahamowicz M, Moodie EE. Accuracy of conventional and marginal structural cox model estimators: a simulation study. *International Journal of Biostatistics*. 2010; 6(2) Article 13.
8. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period: application to the healthy worker survivor effect. *Mathematical Modelling*. 1986; 7:1393–1512. [Errata (1987) in *Computers and Mathematics with Applications* 14, 917–921. Addendum (1987) in *Computers and Mathematics with Applications* 14, 923–945. Errata (1987) to addendum in *Computers and Mathematics with Applications* 18, 477.].
9. Robins, JM. Causal inference from complex longitudinal data. In: Berkane, M., editor. *Latent Variable Modeling and Applications to Causality*. Lecture notes in statistics 120. Springer-Verlag; 1997. p. 69-117.
10. Murphy SA, van der Laan MJ, Robins JM. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*. 2001; 96(456):1410–1423. [PubMed: 20019887]
11. van der Laan MJ, Petersen ML, Joffe MM. History-adjusted marginal structural models and statically-optimal dynamic treatment regimens. *International Journal of Biostatistics*. 2005; 1(1) Article 4.
12. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & Clinical Pharmacology & Toxicology*. 2006; 98:237–242. [PubMed: 16611197]
13. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: Main Content. *International Journal of Biostatistics*. 2010a; 6 Article 7.
14. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part II: Proofs and Additional Results. *International Journal of Biostatistics*. 2010b; 6 Article 8.
15. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *International Journal of Biostatistics*. 2010; 6 Article 18.
16. Young JG, Cain LE, Robins JM, O'Reilly EJ, Hernán MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in Biosciences*. 2011
17. D'agostino RB, Lee M, Belanger AJ. Relation of pooled logistic regression to time-dependent Cox regression analysis: the Framingham Heart Study. *Statistics in Medicine*. 1990; 9:1501–1515. [PubMed: 2281238]
18. Agresti, A. *Categorical Data Analysis*. Connecticut, USA: Wiley Series in Probability and Statistics; 2012.
19. Wang Z, Louis TA. Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika*. 2003; 90(3):765–775.
20. Robins, JM.; Wasserman, L. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In: Geiger, D.; Shenoy, P., editors. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann; 1997. p. 409-420.
21. Spirtes, P.; Glymour, C.; Scheines, R. *Causation, Prediction and Search*. New York: Springer-Verlag; 1993.
22. Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995; 82:669–710.

23. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004; 15:615–625. [PubMed: 15308962]
24. Robins, JM.; Hernán, MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G., editors. *Advances in Longitudinal Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC Press; 2009. p. 553-599.
25. Robins, JM.; Hernán, MA.; Siebert, U. Effects of multiple interventions. In: Ezzati, M.; Lopez, AD.; Rodgers, A.; Murray, CJL., editors. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Geneva: World Health Organization; 2004.
26. Taubman SL, Robins JM, Mittleman MA, Hernán MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *International Journal of Epidemiology*. 2009; 38(6):1599–1611. [PubMed: 19389875]
27. Westreich D, Cole SR, Young JG, Palella F, Tien PC, Kingsley L, Gange SJ, Hernán MA. The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. *Statistics in Medicine*. 2012
28. van der Laan, MJ.; Robins, JM. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer; 2002.
29. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*. 2005; 61:692–972. [PubMed: 16135020]
30. van der Laan MJ. Targeted maximum likelihood based causal inference: Part I. *International Journal of Biostatistics*. 2010; 6(2) Article 2.
31. van der Laan MJ. Targeted maximum likelihood based causal inference: Part II. *International Journal of Biostatistics*. 2010; 6(2) Article 3.

Table 1

Bias of IPW estimates under the six choices of (β_1, θ_1) for $n = 100,000$ and $K = 6$. ψ_j is the IPW estimate of ψ_j , $E[\hat{\psi}_j]$ is the mean of the estimates $\hat{\psi}_j$ over the 20,000 simulation runs and $Bias(\hat{\psi}_j; \psi_j) = E[\hat{\psi}_j] - \psi_j, j = 0, 1$.

β_1	θ_1	ψ_0	$E[\hat{\psi}_0]$	$Bias(\hat{\psi}_0; \psi_0)$	ψ_1	$E[\hat{\psi}_1]$	$Bias(\hat{\psi}_1; \psi_1)$
-2.0	-2.0	-0.8	-0.8056	-0.0056	0.4574	0.4511	-0.0063
-0.5	-0.5	-0.8	-0.8021	-0.0021	0.0583	0.0586	0.0003
0.0	-0.5	-0.8	-0.8011	-0.0011	0	0.0004	0.0004
-0.5	0.0	-0.8	-0.8004	-0.0004	0	0.0002	0.0002
0.5	-2.0	-0.8	-0.7973	0.0027	-0.2064	-0.2047	0.0017
2.0	-2.0	-0.8	-0.7772	0.0228	-0.8676	-0.8709	-0.0033

Table 2

Bias of the MLE of θ_2 for $\theta_2 = \psi_0$ under the six choices of (β_1, θ_1) for $n = 100,000$ and $K = 6$. $\hat{\theta}_2$ is the MLE of θ_2 , $E[\hat{\theta}_2]$ is the mean of the estimates $\hat{\theta}_2$ over the 20,000 simulation runs and $Bias(\hat{\theta}_2, \theta_2) = E[\hat{\theta}_2] - \theta_2$.

β_1	θ_1	θ_2	$E[\hat{\theta}_2]$	$Bias(\hat{\theta}_2, \theta_2)$
-2.0	-2.0	-0.8	-0.8006	-0.0006
-0.5	-0.5	-0.8	-0.8006	-0.0006
0.0	-0.5	-0.8	-0.8006	-0.0006
-0.5	0.0	-0.8	-0.8004	-0.0004
0.5	-2.0	-0.8	-0.8004	-0.0004
2.0	-2.0	-0.8	-0.7992	0.0008

Table 3

Bias of the MLE of θ_3 for both θ_3 and ψ_1 under the six choices of (β_1, θ_1) for $n = 100,000$ and $K = 6$. $\hat{\theta}_3$ is the MLE of θ_3 , $E[\hat{\theta}_3]$ is the mean of the estimates $\hat{\theta}_3$ over the 20,000 simulation runs, $Bias(\hat{\theta}_3, \theta_3) = E[\hat{\theta}_3] - \theta_3$ and $Bias(\hat{\theta}_3, \psi_1) = E[\hat{\theta}_3] - \psi_1$.

β_1	θ_1	$E[\hat{\theta}_3]$	θ_3	$Bias(\hat{\theta}_3, \theta_3)$	ψ_1	$Bias(\hat{\theta}_3, \psi_1)$
-2.0	-2.0	0.0008	0	0.0008	0.4574	-0.4566
-0.5	-0.5	-0.0005	0	-0.0005	0.0583	-0.0588
0.0	-0.5	-0.0007	0	-0.0007	0	-0.0007
-0.5	0.0	-0.0006	0	-0.0006	0	-0.0006
0.5	-2.0	-0.0013	0	-0.0013	-0.2064	0.2051
2.0	-2.0	-0.0060	0	-0.0060	-0.8676	0.8616