# Data-adaptive longitudinal model selection in causal inference with collaborative targeted minimum loss-based estimation

**Mireille E. Schnitzer[1,\*], Joel Sango[2,3], Steve Ferreira Guerra[1], and Mark J. van der Laan[4]**

[1]Faculty of Pharmacy, Université de Montréal, Montréal, H3C 3J7, Canada,

[2]Statistics Canada, Ottawa, K1A 0T6, Canada,

[3]Department of Mathematics and Statistics, Université de Montréal, Montréal, H3T 1J4, Canada,

[4]School of Public Health, University of California, Berkeley, 94720, USA.

[\*]*email:* mireille.schnitzer@umontreal.ca

SUMMARY: Causal inference methods have been developed for longitudinal observational study designs where confounding is thought to occur over time. In particular, one may estimate and contrast the population mean counterfactual outcome under specific exposure patterns. In such contexts, confounders of the longitudinal treatment-outcome association are generally identified using domain-specific knowledge. However, this may leave an analyst with a large set of potential confounders that may hinder estimation. Previous approaches to data-adaptive model selection for this type of causal parameter were limited to the single time-point setting. We develop a longitudinal extension of a collaborative targeted minimum loss-based estimation (C-TMLE) algorithm that can be applied to perform variable selection in the models for the probability of treatment with the goal of improving the estimation of the population mean counterfactual outcome under a fixed exposure pattern. We investigate the properties of this method through a simulation study, comparing it to G-Computation and inverse probability of treatment weighting. We then apply the method in a real data example to evaluate the safety of trimester-specific exposure to inhaled corticosteroids during pregnancy in women with mild asthma. The data for this study were obtained from the linkage of electronic health databases in the province of Quebec, Canada. The C-TMLE covariate selection approach allowed for a reduction of the set of potential confounders, which included baseline and longitudinal variables.

KEY WORDS: Collaborative double robustness; saturated marginal structural model; TMLE.

This paper has been submitted for consideration for publication in *Biometrics*

# 1. Introduction

Causal inference methods have been developed for longitudinal observational study designs where confounding is thought to occur over time (Robins et al., 2000). In particular, several methods estimate the expected counterfactual outcome under a fixed exposure regime (a *saturated marginal structural model* Robins et al. 2000), including inverse probability of treatment weighting (IPTW; Robins et al. 2000) and targeted minimum loss-based estimation (TMLE; van der Laan and Rubin 2006; van der Laan and Gruber 2010). For a study design where information is only collected at discrete time points, identifiability of this parameter depends on the sequential ignorable treatment assignment assumption (Robins, 1998). Sequential ignorability is equivalent to the assumption that, at every time point, treatment is essentially randomly assigned conditional on past measured information.

In such contexts, variables thought to satisfy sequential ignorability are generally identified using domain-specific knowledge (Robins, 2001; Hernán et al., 2002). However, this may leave an analyst with a large set of potential confounders that may hinder estimation (VanderWeele and Shpitser, 2011; Schnitzer et al., 2016). Various methods have been developed to perform variable or model selection with the goal of improving estimation of the causal parameter of interest (e.g. Crainiceanu et al. 2008; van der Laan and Gruber 2010; De Luna et al. 2011; Wang et al. 2012; Vansteelandt et al. 2012; Wilson and Reich 2014; Belloni et al. 2014). However, these methods have largely focused on the setting where treatment is defined at a single time point. Machine learning methods have been used for the estimation of the treatment model(s) in both longitudinal and non-longitudinal settings (e.g. Neugebauer et al. 2007; Petersen et al. 2007; Ertefaie et al. 2017; Shortreed and Ertefaie 2017; Benkeser et al. 2017; Schnitzer and Cefalu 2018). Recently, Shi et al. (2018) proposed using the Dantzig selector based on the A-learning estimating function for high-dimensional covariate selection in dynamic treatment regimes.

Collaborative targeted minimum loss-based estimation (C-TMLE; van der Laan and Gruber 2010; Gruber and van der Laan 2011; Ju et al. 2017) is a data-adaptive procedure that, given an initial outcome model fit, constructs a targeted estimator by sequentially selecting covariates into the model for treatment. The selection is made based on a loss function criterion placed on the updated outcome model fit. Cross-validation is then used to determine at what step the sequential selection terminates. By design, this procedure prioritizes the selection of terms to reduce bias in the estimation of the target parameter while avoiding terms that disproportionately increase the estimation variance. In this paper, we demonstrate a novel extension of the single time point implementation of C-TMLE to the longitudinal case for the estimation of the expected counterfactual outcome under fixed exposure regimen. To ease understanding, we present the methodology in the context of a simplified example where treatment and covariate information is collected at two time points and the outcome measured at a third time point.

Our illustrative application involves drug safety during pregnancy. Asthma is a highly prevalent chronic condition affecting pregnant women that can impact the health of the mother and fetus (Murphy et al., 2011). The current standard of asthma treatment during pregnancy is to continue medication because poorly controlled symptoms pose an acute risk to the fetus and mother. However, many women with mild asthma stop taking medication entirely due to the fear of increased risk to the pregnancy. Such women are indicated to receive a low dose of inhaled corticosteroids (ICS; 0-250$\mu g$ in Fluticasone equivalent) (National Asthma Education and Prevention Program expert panel, 2005) versus none. Our study evaluates the safety of taking low dose ICS during pregnancy for women with mild asthma. While similar questions have been investigated using cohort data (Cossette et al., 2013), a formal causal approach has yet to be taken. A particular challenge in this context is to contrast treatment regimes in a population eligible for either option. In reality, treatment

can vary over time due to fluctuating symptom severity and pregnancy complications. Previous work did not properly account for time-varying confounders (variables that can affect future treatment and outcome) that are also influenced by previous treatment (Robins et al., 2000). We investigate whether performing model selection with our proposed method changes the study conclusion relative to several existing causal inference methods.

## 2. The delivery cohort dataset and structure

Our dataset is a cohort of singleton deliveries, defined as live or stillborn births occurring past 20 weeks of gestation. It was constructed from a linkage of the *Régie de l'assurance-maladie du Québec* and *Maintenance et Exploitation des Données pour l'Étude de la Clientèle Hospitalière* databases. The former resource contains information on all medical services used and medication prescriptions filled by Quebec residents on the public drug insurance plan. The latter resource includes data from acute care hospitalizations. The derived database contains all singleton deliveries in the province of Quebec, Canada between the years 1998-2008 for women $\leqslant 45$ years with at least one asthma diagnosis and at least one asthma medication prescribed in the year before or during pregnancy. For inclusion, these women also had to be covered by the Quebec public drug insurance plan in the year prior to and during pregnancy (Cossette et al., 2013). Women taking theophylline, cromoglycate, nedocromil, ketotifen, or long-acting $\beta_2$-agonists without an ICS were excluded. A subsetted cohort of women with mild asthma in the year prior to pregnancy was created (totaling 5881 deliveries). For simplicity, the cohort was restricted to each woman's first delivery ($n = 5048$).

For this investigation, we structured the data to contain two trimester-specific exposure times. The dataset contains the following information about each subject:

- Pre-pregnancy covariates ($L_0$) measured in the year prior to pregnancy, including maternal characteristics (age at delivery, receiving social assistance, rural residence, and usage of

beta-blockers), chronic maternal diseases (hypertension, diabetes, cystic fibrosis of the pancreas, antiphospholipid syndrome, and cyanotic heart disease), and asthma control-related variables (not controlled, use of short-acting $\beta_2$-agonist, use of leukotriene receptor antagonist, use of oral corticosteroids, use of intranasal corticosteroids, at least one hospitalization for asthma, and at least one emergency room visit for asthma),

- No exposure to ICS ($A_0 = 0$) or low-dose ICS ($\leqslant 250\mu g$ in Fluticasone equivalent daily; $A_0 = 1$) at the beginning of the first trimester,

- Covariates measured during the first trimester ($L_1$), including chronic maternal diseases, pathologies related to pregnancy (gestational diabetes, maternal infection, anemia, vaginal bleeding, placental complications, placental abruption, and usage of beta-blockers), and asthma control-related variables,

- Exposure status at the beginning of the second trimester ($A_1 = \{0, 1\}$), and

- Three binary fetal growth outcomes ($Y$), analyzed separately: low birth weight (LBW; < 2,500 g), premature birth (delivery before 37 weeks gestation), and small for gestational age (SGA; below the 10th percentile) (Cossette et al., 2013). These outcomes are measured at delivery, which can occur in the second or third trimester.

We are interested in estimating the mean outcome under exposure versus no exposure to low-dose ICS at the beginning of the first two trimesters. The algorithm developed in the next section aims to select covariates into a model predicting exposure with the goal of minimizing empirical risk based on a loss function that specifically relates to the target parameter.

## 3. Methods

### 3.1 *Identifiability of the target parameter*

We observe randomly sampled longitudinal data $O = (L_0, A_0, L_1, A_1, Y) \sim P_0$ where $L_t$ are covariates, $A_t$ is a treatment indicator measured at times $t = 0$ and 1, and $P_0$ is the

true data-generating distribution, lying in some model $\mathcal{M}$. Let $Y$ be the outcome of interest, measured at the end of pregnancy. Let $Y^{\bar{a}}$ denote the counterfactual outcome under the fixed treatment regime $\bar{a}$ and similarly let $L_1^{a_0}$ represent the intermediate outcome under treatment $a_0$ at the first time point. Our goal is to estimate $\psi = E(Y^{\bar{a}})$, the marginal mean outcome under treatment regime $\bar{a}$. For instance, if we consider the regime $\bar{a} = (a_0, a_1) = (1, 1)$, then $\psi$ represents the expected mean outcome in the population had all mothers received a low daily dose of ICS in the first and second trimesters.

Let $Q_2(\bar{l}_1) = E(Y^{\bar{a}} \mid \overline{L}_1 = \bar{l}_1)$ be the conditional expectation of the outcome under a two time point fixed treatment regime (with an implicit dependence on $\bar{a}$, the regime in question) evaluated at some value $l_1$. Let $Q_1(l_0) = E(Y^{\bar{a}} \mid L_0 = l_0)$ be the conditional expectation of the outcome after fixing exposure in the first time point to the treatment $a_0$ at the baseline covariate value $l_0$. In this setting, the *sequential ignorability assumption* (needed for identifiability) requires

$$Y^{\bar{a}} \perp\!\!\!\perp A_1 \mid A_0 = a_0, \overline{L}_1 \text{ and } Y^{\bar{a}} \perp\!\!\!\perp A_0 \mid L_0, \tag{1}$$

meaning that under a fixed regime, the counterfactual outcome is independent of past treatment conditional on the observed past prior to each treatment. This is often thought of as measuring all risk factors that affect subsequent treatment (Robins et al., 2000). In the example, such variables include indicators of asthma control during the first trimester (hospitalization and complementary controller mediation) which may directly affect fetal growth and subsequent treatment as well as being affected by the previous treatment taken.

We also require *positivity* which implies that for all values of $a_0$ that are being considered, $P(A_0 = a_0 \mid L_0) > 0$ almost surely (i.e. for all possible combinations of baseline covariates, all treatments must be possible at the first time point). Similarly, for all $\bar{a}$ of interest, we require that $P(A_1 = a_1 \mid \overline{L}_1, A_0 = a_0) > 0$ almost surely (Bang and Robins, 2005). For the asthma example, this essentially means that all subjects must be eligible to follow all

possible treatment regimens at all time points. Note that if these probabilities are estimated to be (close to) zero, then we have *(near) practical positivity violations*, which can hinder estimation. An additional assumption is *no interference*, such that one subject's treatment will not affect another's counterfactual outcome under a fixed treatment. No interference is credible in this case because asthma and fetal growth are not infectious, nor were patient interactions likely to otherwise affect outcomes. Finally, we must also assume *consistency*, that the intervention of assigning treatment at a given time point corresponds with the exposure classification. We can then write that we observe $L_1 = L_1^{a_0}$ when $A_0 = a_0$ and $Y = Y^{\overline{a}}$ when $\overline{A}_1 = \overline{a}$.

Hence, we have that $Q_2(\overline{l}_1) = E(Y \mid \overline{A}_1 = \overline{a}, \overline{L}_1 = \overline{l}_1)$ at the realization $\overline{l}_1$ and $Q_1(l_0) = E(Q_2(\overline{l}_1) \mid A_0 = a_0, L_0 = l_0)$ at $l_0$, so that both quantities are estimable from the data. And since $E(Y^{\overline{a}}) = E\{Q_1(L_0)\}$, the parameter of interest is also identifiable (Bang and Robins, 2005; Petersen et al., 2013). Throughout, a subscript $n$ will be used to denote an estimate of a quantity. If we sequentially estimate the quantities $Q_2(\overline{l}_1)$ and $Q_1(l_0)$, we can obtain a plug-in estimator for $E(Y^a)$ by taking the mean over the values $Q_{1,n}(l_0)$, i.e. the estimates of $Q_1(l_0)$ obtained for each subject (Bang and Robins, 2005).

Define $g_0(l_0) = P(A_0 = a_0 \mid L_0 = l_0)$, the conditional probability of taking treatment $a_0$ at the beginning of the first trimester. Similarly, define $g_1(\overline{l}_1) = P(A_1 = a_1 \mid \overline{L}_1 = \overline{l}_1, A_0 = a_0)$ as the conditional probability of taking $a_1$ at the beginning of the second trimester. An inverse probability of treatment weighted estimate for $\psi$ can be taken as the sample average of $YI(\overline{A}_1 = \overline{a})/\{g_{0,n}(L_0)g_{1,n}(\overline{L}_1)\}$ (Robins et al., 2000) where $I(\cdot)$ is the indicator function.

### 3.2 *Longitudinal targeted minimum loss-based estimation*

Longitudinal TMLE (LTMLE) (van der Laan and Gruber, 2012) is a semiparametric estimator that takes initial estimates of $Q_2(\overline{l}_1)$ and $Q_1(l_0)$ and updates them to reduce excess bias in the estimation of $\psi$ resulting from the use of misspecified models or flexible learning

algorithms with slower rates of convergence. The form of this update is designed to satisfy the efficient influence function estimating equation (van der Laan and Rubin, 2006), resulting in locally efficient and doubly robust estimation of $\psi$ (van der Laan and Robins, 2003) (details in the Web Appendix A). Below we give the algorithm for a binary outcome and fixed $\bar{a}$ though it has been described elsewhere (van der Laan and Gruber, 2012).

---

**Algorithm 1** LTMLE algorithm for two time points

---

1: Compute $Q_{2,n}(\bar{l}_1)$, the initial estimate of $Q_2(\bar{l}_1) = E(Y \mid \overline{A}_1 = \bar{a}, \overline{L}_1 = \bar{l}_1)$, for each subject at their realization $\bar{l}_1$.

2: Run intercept-free logistic regression $Y \sim \epsilon_2/\{g_{0,n}(L_0)g_{1,n}(\overline{L}_1)\}+offset[logit\{Q_{2,n}(\overline{L}_1)\}]$ using subjects with $\overline{A} = \bar{a}$. Let $\widehat{\epsilon}_2$ be the coefficient estimate.

3: Set $Q_{2,n}^*(\bar{l}_1) = expit\big[\widehat{\epsilon}_2/\{g_{0,n}(l_0)g_{1,n}(\bar{l}_1)\} + logit\{Q_{2,n}(\bar{l}_1)\}\big]$ for each subject.

4: Compute $Q_{1,n}(l_0)$, the initial estimate of $Q_1(l_0)$, by regressing $Q_{2,n}^*(\overline{L}_1)$ on $A_0$ and $L_0$ and making a prediction for each subject setting $A_0 = a_0$ .

5: Run intercept-free logistic regression $Q_{2,n}^*(\overline{L}_1) \sim \widehat{\epsilon}_1/\{g_{0,n}(L_0)\}+offset[logit\{Q_{1,n}(L_0)\}]$ using subjects with $A_0 = a_0$. Let $\widehat{\epsilon}_1$ be the coefficient estimate.

6: Set $Q_{1,n}^*(l_0) = expit\big[\widehat{\epsilon}_2/\{g_{0,n}(l_0)\} + logit\{Q_{1,n}(l_0)\}\big]$ for all subjects.

7: Let $\psi_n$, the targeted estimate, be the sample mean of $Q_{1,n}^*(L_0)$ over all subjects.

---

This procedure results in semiparametric efficient estimation of $\psi$ if both the estimation of $g$ and $Q$ are consistent where $g = (g_0, g_1)$ and $Q = (Q_2, Q_1)$ are defined with respect to a set of covariates that satisfies equation (1) (van der Laan and Gruber, 2012), i.e. sufficiently controls for confounding. This estimator is also doubly robust, in the sense that if the estimation of *either* $g$ or $Q$ is consistent, then the estimator is consistent for the target parameter. The collaborative double robustness result (van der Laan and Gruber, 2010; Schnitzer et al., 2016) shows that consistency can also be achieved when both sets of models are incorrectly specified in a compatible way. We explain further in Section 3.3.

There are several choices of loss functions to define the error (i.e. risk or lack of fit) of the TMLE built from this procedure. One example is the logistic log-likelihood loss function, defined as $\mathcal{L}(Q)(O) = \mathcal{L}_2(Q_2)(O) + \mathcal{L}_1(Q_1; Q_2)(O)$ where

$$\mathcal{L}_2(Q_2)(O) = - I(\overline{A}_1 = \overline{a}) \left[ Y \log\{Q_2(\overline{L}_1)\} + (1 - Y) \log\{1 - Q_2(\overline{L}_1)\} \right] \tag{2}$$

$$\mathcal{L}_1(Q_1; Q_2)(O) = - I(A_0 = a_0) \times$$

$$\left[ Q_2(\overline{L}_1) \log\{Q_1(L_0)\} + \{1 - Q_2(\overline{L}_1)\} \log\{1 - Q_1(L_0)\} \right]. \tag{3}$$

The empirical error is obtained by evaluating $\mathcal{L}(Q_n^*)(o)$ for each subject's observed values of $O$ and then taking the mean. The loss functions have a sequential validity in the sense that the expectation of $\mathcal{L}_2(Q_2)(O)$ is minimized at the true $Q_2(\overline{L}_1)$ and the expectation of $\mathcal{L}_1(Q_1; Q_2)(O)$ indexed by the true $Q_2$ is minimized at the true $Q_1(L_0)$. Both have a quadratic dissimilarity such that the distance between the errors at the true and candidate $Q$ is quadratic, with the second loss component indexed at a fixed value of $Q_2$ (van der Laan and Gruber, 2010, Section 2.2).

### 3.3 *Longitudinal collaborative double robustness*

The general collaborative double robustness result (van der Laan and Gruber, 2010) states that for doubly robust estimators with components $Q$ and $g$ and influence function $D(\psi, Q, g)$, the treatment model must only condition on the error in the outcome model predictions in order to obtain consistent estimation. Suppose that $Q_n$ and $g_n$ are estimators of $Q$ and $g$, respectively, but that $Q_n$ and $g_n$ converge componentwise in probability to some limits $\widetilde{Q}$ and $\widetilde{g}$. We have that $E\{D(\psi, \widetilde{Q}, \widetilde{g})(O)\} = 0$ when either $\widetilde{Q} = Q$ or $\widetilde{g} = g$. In the two time point case, given fixed but possibly misspecified estimates $Q_n = (Q_{2,n}, Q_{1,n})$, we are interested in determining the set of values of $g = (g_0, g_1)$ that will lead to consistent estimation of $\psi$. As in the general case, this corresponds to determining the form of estimates $(g_{0,n}, g_{1,n})$

which converge to $\widetilde{g} = (\widetilde{g}_0, \widetilde{g}_1)$ such that $E\{D(\psi, \widetilde{Q}, \widetilde{g})(O)\} = 0$. For more details about convergence, please refer to Web Appendix A.

Since $E\{D(\psi, Q, \widetilde{g})(O)\} = 0$ and by linearity of expectations, we can write $E\{D(\psi, \widetilde{Q}, \widetilde{g})(O)\} = E\{D(\psi, \widetilde{Q}, \widetilde{g})(O) - D(\psi, Q, \widetilde{g})(O)\}$ (van der Laan and Gruber, 2010). For the two time point case, we can rewrite the influence function (Web Appendix A, equation 1) as

$$D(\psi, Q, g)(O) = \left\{ \frac{I(\overline{A}_1 = \overline{a})}{g_1(\overline{L}_1)g_0(L_0)}Y - \psi \right\} - \left\{ \frac{I(\overline{A}_1 = \overline{a})}{g_1(\overline{L}_1)g_0(L_0)} - \frac{I(A_0 = a_0)}{g_0(L_0)} \right\} Q_2(\overline{L}_1)$$
$$- \left\{ \frac{I(A_0 = a_0)}{g_0(L_0)} - 1 \right\} Q_1(L_0)$$

Noting the linearity of $D(\psi, Q, g)$ in the components of $Q$, we can also write

$$D(\psi, \widetilde{Q}, \widetilde{g})(O) - D(\psi, Q, \widetilde{g})(O) = \frac{I(A_0 = a_0)}{\widetilde{g}_1(\overline{L}_1)\widetilde{g}_0(L_0)} \left\{ I(A_1 = a_1) - \widetilde{g}_1(\overline{L}_1) \right\} \left\{ \widetilde{Q}_2(\overline{L}_1) - Q_2(\overline{L}_1) \right\}$$
$$- \frac{1}{\widetilde{g}_0(L_0)} \left\{ I(A_0 = a_0) - \widetilde{g}_0(L_0) \right\} \left\{ \widetilde{Q}_1(L_0) - Q_1(L_0) \right\}$$

We have that the conditional expectation of the second component is zero when $\widetilde{g}_0(L_0) = P[A_0 = a_0 \mid \{\widetilde{Q}_1(L_0) - Q_1(L_0)\}]$. The conditional expectation of the first component is zero when we set $\widetilde{g}_1(\overline{L}_1) = P[A_1 = a_1 \mid \{\widetilde{Q}_2(\overline{L}_1) - Q_2(\overline{L}_1)\}/\widetilde{g}_0(L_0), I(A_0 = a_0)]$. Therefore, given an estimator $Q_n = (Q_{2,n}, Q_{1,n})$ consistent for $(\widetilde{Q}_2, \widetilde{Q}_1)$, if we knew the value of the errors $\widetilde{Q}_1(l_0) - Q_1(l_0)$ and $\widetilde{Q}_2(\overline{l}_1) - Q_2(\overline{l}_1)$, it would be sufficient to adjust for these errors in the respective treatment models. This would produce consistent inference for any doubly robust estimator. However, since this error is unknown, C-TMLE targets the reduction of the model risk (i.e. improves the fit of $Q_n$) by selecting variables into the treatment models that explain the residual bias $\widetilde{Q} - Q$. The longitudinal C-TMLE algorithm is presented in Section 3.4.

## 3.4 *Collaborative selection for the treatment models*

Suppose we have defined the error of estimates $Q_n$ and $g_n$ by specifying respective loss functions. The concept behind C-TMLE (van der Laan and Gruber, 2010, Section 2.4) is that the procedure must produce a sequence of targeted estimates represented by $\{Q_n^{*,(k)}, g_n^{(k)}; k = 1, ..., K\}$ such that the sequence is simultaneously decreasing in the empirical risk of both $Q_n^*$

and $g_n$. This is achieved by producing an initial estimate $Q_n^{init}$ and then gradually increasing the flexibility of the model for $g$, for instance by selecting covariates or non-linearities in a greedy stepwise manner, where the error is evaluated on the $Q_n^*$ resulting from the TMLE update. The step $k$ indexes the number of covariates or terms that have been added to either treatment model. $K$ is defined as the number of moves that results in all of the covariates or terms added to the models. We assume that at some minimal step $k_m \leqslant K$, we achieve consistent estimation where $g_n^{(k_m)}$ converges to $\widetilde{g}^{(k_m)}$ and $Q_n^{*,(k_m)}$ to $\widetilde{Q}^{*,(k_m)}$ such that $E[D\{\psi, \widetilde{Q}^{*,(k_m)}, \widetilde{g}^{(k_m)}\}(O)] = 0$ (which relaxes the typical assumption that we need to include all terms in the treatment models for consistent estimation). We also assume that models past $k_m$ also allow for consistent estimation (van der Laan and Gruber, 2010, Section 4). The C-TMLE procedure uses an empirical risk function, evaluated through cross-validation, to choose the index at which the estimator is selected. For this, we choose a penalized loss function (Gruber and van der Laan, 2011) $\mathcal{L}(Q)(O) = \mathcal{L}_2(Q_2)(O) + \mathcal{L}_1(Q_1; Q_2)(O) + Var\{D(\psi, Q, g)(O)\}/n$ where the penalty term is the variance of the efficient influence function divided by the sample size (see Section 3.4.2 for more intuition). This penalty term converges to zero, so that the loss function remains asymptotically valid.

3.4.1 *Algorithm.* A C-TMLE variable selection procedure in the setting with a single treatment point has been described in Gruber and van der Laan (2011). We describe a similar procedure for two treatment points for the estimation of $\psi = E(Y^{\overline{a}})$ (generalized for larger numbers of time points in the Web Appendix B) which we call Collaborative Longitudinal TMLE (C-LTMLE). We define the operation $update(Q_n, g_n)$ as the procedure that takes initial estimates $Q_n = (Q_{2,n}, Q_{1,n})$ and updates them separately using the LTMLE procedure with respect to $g_n = (g_{0,n}, g_{1,n})$, returning the updated estimates $Q_n^* = (Q_{2,n}^*, Q_{1,n}^*)$ as described in Section 3.2. Define "allowable moves" as the set of remaining additions of covariates or terms to the existing models for $g_0$ and $g_1$. We restrict such moves by only

allowing variables in $L_0$ to be selected for $g_0$, but any of $\overline{L}_1$ allowed into the model for $g_1$. Finally, define the notation $g_n^{(k-1)}(m)$ as the probability of treatment estimates using the models from step $k-1$ after additional move $m$ (that is, after adding a covariate or term to either model according to move $m$).

---

**Algorithm 2** C-LTMLE candidate building procedure

---

1: Initialize $g_n^{(0)}$ and $Q_n^B := Q_n^{(init)}$; and set $Q_n^{*,(0)} := update\{Q_n^{(init)}, g_n^{(0)}\}$; and $l_0 := mean(\mathcal{L}\{Q_n^{*,(0)}\})$.

2: **for** $k = 1, ..., K$ where $K$ is total number of allowable moves **do**

3:    Select move $m$ that results in the lowest error in the updated estimate $mean(\mathcal{L}[update\{Q_n^B, g_n^{(k-1)}(m)\}])$. Call this minimum error $l_k$.

4:    **if** $l_k \geqslant l_{k-1}$ **then**

5:       Set $Q_n^B := update\{Q_n^B, g_n^{(k-1)}\}$, the new baseline estimate.

6:       Select move $m$ that results in the lowest error in the updated estimate $mean(\mathcal{L}[update\{Q_n^B, g_n^{(k-1)}(m)\}])$. Call this minimum error $l_k$.

7:    **end if**

8:    Set $g_n^{(k)} := g_n^{(k-1)}(m)$; $Q_n^{*,(k)} := update\{Q_n^B, g_n^{(k)}(m)\}$.

9: **end for**

---

This procedure produces a sequence of targeted estimates $Q_n^{*,(k)} = (Q_{2,n}^{*,(k)}, Q_{1,n}^{*,(k)})$ with corresponding $g_n^{(k)}$ that are improving in fit by construction. Cross-validation of the full procedure is used to select the step $k_{m,n}$, dependent on sample size, at which the procedure should terminate. We thus obtain the final vector of predictions, $Q_{2,n}^{*,(k_{m,n})}$ which is used to produce an estimate of $Q_1$ which is then updated with the final selected $g_{0,n}^{k_{m,n}}$ to obtain $Q_{1,n}^{*,(k_{m,n})}$. The C-LTMLE estimate $\psi_n^{CLTMLE}$ is then computed as the mean of $Q_{1,n}^{*,(k_{m,n})}$ over all subjects. In the Web Appendix C, we provide intuition for why this procedure would be likely to select covariates in a way that improves the estimation of the target parameter.

3.4.2 *Properties.*  With a valid loss function (with corresponding risk minimized at the truth and possessing a quadratic dissimilarity), the cross-validation selector converges to an oracle selector (which picks the $Q_n^{*,(k)}$ closest to the true $Q$ with respect to the loss function) at a rate of $log(n)/n$ (van der Laan and Dudoit, 2003). A reduction in error of $Q_n$ implies a less biased plug-in estimate. In C-TMLE, because we are selecting amongst TMLEs targeting the same $\psi$ with the same initial estimate $Q_n^{(init)}$, the cross-validation selects the best update step indexed by $k$. Because overfit treatment models may result in increased variance, the penalized loss function may select an index prior to $k_m$ in finite samples, which allows the procedure to trade off between reduced bias and variance. The asymptotic linearity of C-TMLE, which is justified in van der Laan and Rose (2011, A.18), relies on the convergence of the cross-validation selector to some step $k$ between $k_m$ and $K$. We require that the candidates $g_n^{(k)}$ and $Q_n^{*,(k)}, k \geqslant k_m$ allow for TMLE convergence at the $\sqrt{n}$-rate as described for LTMLE in the Web Appendix A. The specific form of the influence function depends on the convergence of the selected $g_n^{(k)}$ to conditional distribution $\widetilde{g}^{(k)}$. If $\widetilde{g}^{(k)} = g$ the influence function of the C-LTMLE will not involve any contribution from the estimation of $Q$ and $D(\psi, Q_n^{*,(k)}, g_n^{(k)})$ will be $o_P(1/\sqrt{n})$ in the $L_2(P)$ norm. If not, the influence function may include a component related to $Q_n^{*,(k)}$ (van der Laan and Rose, 2011, A.18). In the simulation study and the example, we approximate the large-sample variance of C-LTMLE using the efficient influence function evaluated with the largest (i.e. unselected) model for $g$.

## 4. Simulation study

The performance of this estimator was assessed in a simple setting, a setting with a large number of correlated covariates, and a setting with nonlinear terms and the potential for practical positivity violations, respectively. For each data generation, 1000 independent draws of identically distributed longitudinal data with binary treatments were generated

with sample sizes $n = \{250, 500, 1000\}$. The full data-generation for the three settings is given in the Web Appendix D and summarized below.

In all simulations, we contrast G-Computation (a plug-in estimator that uses the initial estimates of $Q_2$ and $Q_1$ without TMLE updates) (Bang and Robins, 2005), IPTW (Robins et al., 2000), Longitudinal TMLE, and C-LTMLE (Section 3.4) implemented with logistic regressions for simplicity. We also apply what we refer to as C-IPTW, which is the C-LTMLE algorithm applied to a completely misspecified initial estimate $Q_{2,n} = E_n(Y|A_0 = a_0, A_1 = a_1)$ from a model that does not adjust for any additional covariates. In this version of the algorithm, the model estimating $Q_{1,n} = E_n(Q_{2,n} \mid A_0 = a_0)$ is also fit without covariates aside from treatment $A_0$. C-IPTW was included to investigate how C-LTMLE might correct for a severely misspecified outcome model. We present the root mean squared errors (rMSE), mean bias, and Monte-Carlo standard errors (MC SE) for each estimator.

For variance estimation for C-IPTW and the TMLE methods, we applied the full-model influence curve sandwich estimator and a targeted variance estimator (Tran, 2016) implemented in the ltmle package, version 1.0-1 (Lendle et al., 2017). For G-Computation we applied the nonparametric bootstrap. We computed the average of the estimated standard errors $SE_{IC}$, $SE_{TE}$, and $SE_B$, respectively. For the first two approaches, the 95% confidence intervals were computed as the point estimate $\pm 1.96 \times SE$ and for the latter as the 5th and 95th percentiles of the bootstrap-resampled estimates. We present the percentage of interval coverage across the simulations (%Cov) for each method. All computations were carried out using R statistical software version 3.2.0 (R Development Core Team, 2011).

### 4.1 *Simple setting*

The twofold purpose of this first setting is to verify that the C-LTMLE algorithm is selecting covariates in a manner consistent with the known variable selection recommendations in

causal inference (face validity), and to contrast the performance of the alternative algorithms in a familiar setting (Brookhart et al., 2006).

We independently sampled data $O = (L_0, A_0, L_1, A_1, Y)$ in sequence. The independent Gaussian baseline covariates $L_0 = (IV_0, R_0, W_0)$ included an instrument $IV_0$ (that only affected binary treatment choices $A_0$ and $A_1$), a variable $R_0$ that is not a confounder but affects the binary outcome $Y$, and a baseline confounder $W_0$ (directly affecting $A_0$, $L_1$, and $Y$). The Gaussian covariates $L_1 = (IV_1, R_1, W_1)$ at the next time point included an instrument $IV_1$ (only affecting treatment choice $A_1$), a variable $R_1$ that only affects the outcome, and a time-dependent confounder $W_1$ (affecting $A_1$ and $Y$) influenced by previous treatment $A_0$. The target parameter is the mean outcome under treatment, $E\{Y^{1,1}\} = 0.286$. By design there is no potential for practical positivity violations, so one would expect an optimal method to select the confounders and omit the instruments.

All of the models were specified on the full set of covariate histories. Table 1 gives the results from each of the five evaluated methods. G-Computation, which was implemented with a regression conditional on all covariates, was unbiased and outperformed all other methods in terms of variance. IPTW conditional on all covariates maintained some bias and had the largest MC SE. C-IPTW had half of the variance of IPTW and smaller bias for the larger sample sizes. LTMLE was unbiased and had smaller variance than C-IPTW. C-LTMLE had a MC SE that rivaled G-Computation and was also unbiased, outperforming LTMLE for the smaller sample sizes. In terms of standard error estimation, the bootstrap performed well for G-Computation. The influence curve method performed well for C-LTMLE but underestimated the MC SE for LTMLE, and C-IPTW at smaller sample sizes. The targeted estimator consistently overestimated the standard errors for C-IPTW, LTMLE, and C-LTMLE, leading to confidence interval coverage around 99%, and underestimated for IPTW

at $n = 1,000$. In the Web Table 1, we show that the C-LTMLE selected the confounders 100% of the time for the larger sample size and non-confounders between 10-30% of the time.

[Table 1 about here.]

### 4.2 *Correlated covariates*

This scenario evaluates the performance of C-LTMLE with a large number of correlated covariates. We generated 40 baseline multivariate Gaussian covariates with pairwise covariances equal to 0.2 and variances equal to one. Ten of these covariates were confounders, ten were also confounders but more strongly related to the treatment than the outcome, ten were purely risk factors, and ten had no direct effects on either (noise). An additional ten intermediate Gaussian covariates were then generated conditional on the baseline and treatment $A_0$; five of these were confounders, two were more strongly related to treatment, and three only caused the outcome. We generated a binary outcome conditional on the treatments and non-noise baseline and intermediate covariates. The true value of the target parameter was $E(Y^{1,1}) = 0.622$.

Table 2 gives the results of the five methods. All outcome models were specified using logistic regression conditional on all 50 covariates. G-Computation had the least variance overall but the greatest bias, resulting in poor coverage. IPTW was unbiased but had a much greater MC SE than G-Computation. C-IPTW performed similarly to IPTW, though its MC SE was smaller for $n = 250$. LTMLE and C-LTMLE were unbiased and C-LTMLE had lower MC SE for the smaller samples. In terms of standard error estimation, the influence curve method performed well for LTMLE and C-LTMLE, resulting in good coverage for the larger sample sizes, though it underestimated the true standard error for C-IPTW. The targeted method performed well for IPTW and overcovered slightly for LTMLE and C-LTMLE.

[Table 2 about here.]

4.3 *Continuous outcome with potential practical positivity violations*

We extended the simulation study of (Bahamyirou et al., 2018) to the two time point setting. The concept of this scenario is that there are interaction and squared terms between covariates in the data-generating functions for the two treatments, but these nonlinearities are not confounding the relationship between treatments and the continuous outcome. Furthermore, the probabilities of treatment are close to zero when conditional on the true interactions and squared terms, and far from zero when excluding the nonlinearities. Thus, excluding these features should not affect consistency and will avoid practical positivity violations. We consider that the analyst intends to use C-TMLE to select amongst main terms and nonlinearities in ignorance of the consistency under a much smaller (main terms only) treatment model. One would expect optimal performance using a method that a priori excludes these nonlinearities. C-LTMLE would ideally be able to adaptively exclude them and perform similarly.

We generated five baseline covariates: two instruments, one pure cause of the outcome and intermediate confounders, and two confounders. At the second time point, we generated two additional confounders. The treatment and outcome data-generating functions included main terms, interactions, and squared terms though no nonlinearities were common between the two. We implemented G-Computation, IPTW, LTMLE, and C-LTMLE including all main, squared, and interaction terms (20 in $L_0$ and 15 in $L_1$). We also implemented IPTW and LTMLE using just the main terms. C-IPTW and C-LTMLE selected from the full set of possible terms (55 selection steps). To coincide with current practice, each probability of treatment was truncated at 0.01 (Lendle et al., 2017).

Table 3 gives the results of each method in this scenario. The G-Computation fit using a logistic regression that included all possible nonlinearities maintained a low bias and variance for all sample sizes. In contrast, the IPTW that included the same terms had bias and MC

SE that were one magnitude larger. When the IPTW included only the main terms, the bias was nearly eliminated and the variance greatly reduced. LTMLE outperformed the corresponding IPTW estimators in terms of variance when including all terms and just main terms, respectively. The C-IPTW improved upon the estimation of IPTW, roughly quartering the variance but maintaining the magnitude of bias. The C-LTMLE performed slightly better in terms of bias and MC SE than the LTMLE with main terms for $n = 250$ but maintained its bias as the sample size increased. The influence curve approach underestimated the standard error of LTMLE with the full model, and overestimated that of LTMLE with main terms only, resulting in good coverage for the latter. The targeted variance estimation method performed well for IPTW with main terms. The bootstrap produced reasonable standard error estimates for G-Computation. For other models, these standard error estimates did not correspond to the MC SE on average. Confidence interval coverage was low overall due to this issue and to estimation bias for all methods except LTMLE with main terms.

[Table 3 about here.]

## 5. Estimating the effect of inhaled corticosteroids during pregnancy

The C-LTMLE method described above was applied to estimate the effect of taking low daily doses of ICS during pregnancy in women categorized with mild asthma in the year prior to pregnancy. Mild asthma was categorized using an algorithm previously validated in this population (Cossette et al., 2013). Women were considered exposed to low ICS daily doses ($< 250\mu g$ in Fluticasone equivalent) at the beginning of the first trimester ($A_0 = 1$) if they had begun a prescription by the beginning of pregnancy. As summarized in Section 2, there are 18 baseline and 19 intermediate measured variables, leading to a total of 110 possible covariate additions to the four treatment models. All three of the outcomes investigated (LBW, premature birth, and SGA) are binary perinatal risk indicators. Therefore, we seek

to estimate the marginal risk ratio under exposure to low-dose ICS at the beginning of both trimesters versus neither, $E(Y^{1,1})/E(Y^{0,0})$. This parameter is greater than 1 if adherence to low daily doses of ICS increases the probability of the outcome and is therefore hazardous to the fetus.

An additional complication in this dataset that was not previously mentioned is the presence of censoring. We considered all women in the cohort who were not taking either treatment category of interest (for instance, those taking higher daily doses) to be censored. We indicate $C_0 = 1$ if a woman is censored at the beginning of the first trimester (and $C_0 = 0$ otherwise), and $C_1 = 1$ if a woman is censored at the beginning of the second trimester ($C_1 = 0$ otherwise). In particular, if $C_0 = 1$ then $C_1 = 1$ for the same individual. We therefore have the full observed dataset $O = (L_0, C_0, A_0, L_1, C_1, A_1, Y)$ where once censored, all subsequent values of $L, A$ and $Y$ are considered to be missing. In order to adjust for potentially informative censoring in addition to the confounded treatment, we must break $g$ into two components. In particular, we redefine $g_0 = g_{A,0} \times g_{C,0}$ and $g_1 = g_{A,1} \times g_{C,1}$ where now $g_{A,0}(l_0) = P(A_0 = a_0 \mid L_0 = l_0, C_0 = 0)$ and $g_{A,1}(\bar{l}_1) = P(A_1 = a_1 \mid \bar{L}_1 = \bar{l}_1, A_0 = a_0, C_1 = 0)$ are the treatment probabilities and $g_{C,0}(l_0) = P(C_0 = 0 \mid L_0 = l_0)$ and $g_{C,1}(\bar{l}_1) = P(C_1 = 0 \mid \bar{L}_1 = \bar{l}_1, A_0 = a_0, C_0 = 0)$ are the censoring probabilities. In the C-LTMLE algorithm, this allows for the possibility of adding the eligible covariates into both the treatment and censoring models separately, increasing the number of possible additions at every step. Otherwise, the procedure as described above remains unchanged.

We contrasted the results of G-Computation, IPTW, LTMLE, and C-LTMLE, run separately for $\bar{a} = (1, 1)$ and $\bar{a} = (0, 0)$, for the estimation of the marginal risk ratio. For each of the three outcomes, G-Computation, LTMLE and C-LTMLE used the same outcome model, fit with a main terms logistic regression. Treatment probabilities were truncated to lie between (0.01,0.99) in all cases. The variances of LTMLE and C-LTMLE were calculated

using the standard sandwich estimator which uses the efficient influence function with the full (unselected) treatment models and the functional delta method (van der Laan and Gruber, 2012). At the time of writing, the targeted variance estimation method was not available for saturated contrasts in the ltmle package. The variances of G-Computation and IPTW were calculated using the nonparametric bootstrap with 1,000 resamples. All calculations were carried out in R statistical software version 3.1.3 (R Development Core Team, 2011).

The results for the three outcomes and four methods are presented in Table 4. The risk ratio estimates, 95% confidence intervals, and standard errors (SE) are reported. The number of variables selected by C-LTMLE are reported for each exposure category. Truncation was found to be necessary for IPTW and LTMLE, and likely beneficial for C-LTMLE, due to the otherwise large weights. Similarly to related analyses of low dose versus no ICS (Cossette et al., 2013), all three outcomes had risk ratios that were close to one with confidence intervals that contained the null. The usage of C-TMLE only slightly altered the risk ratios compared to LTMLE. In terms of computation time on a local server, G-Computation (with bootstrapping) took 720 seconds, IPTW (with bootstrapping) took 1450 seconds, LTMLE took 1 second, and the C-LTMLE cross-validation procedure took 8670 seconds. We comment on the credibility of the causal assumptions in the Web Appendix E.

[Table 4 about here.]

## 6. Discussion

In this paper, we developed a collaborative longitudinal targeted minimum loss-based estimation (C-LTMLE) approach for time-varying treatments. This method sequentially selects terms to be included in the treatment (and/or censoring) models that are used in the TMLE update steps. This selection is decided by a penalized loss function for the updated nested conditional expectations of the outcome. Related work for dynamic treatment regimes by

Shi et al. (2018) involves a high-dimensional covariate selection method constrained by a doubly-robust estimating equation that involves the product of the errors of the estimated time-specific probability of treatment and nested outcome expectation. Rather than using cross-validation to select the tuning parameter as we do, their method uses a doubly robust Bayesian information criterion. Another difference is that TMLE methods do not require any assumed model structure for the treatment(s) or outcome, though they still require regularity conditions and minimal convergence rates.

Our procedure can be adapted for non-saturated marginal structural models using Pooled Longitudinal TMLE (Petersen et al., 2013). Since selection steps in the given procedure require testing all possible additions to each treatment model, the computational complexity is a concern for a large number of time points. One may potentially adapt the a priori variable ordering in Ju et al. (2017) to the longitudinal setting to reduce computational time.

In the simulation study, C-LTMLE performed better than other semiparametric methods, but worse than the fully parametric (slightly misspecified) G-Computation for $n < 1,000$, in terms of standard error and bias in the three scenarios. In the first simulation, we observed that C-LTMLE selected covariates in a way that roughly corresponded to common knowledge in causal inference, selecting true confounders 100% of the time for larger sample sizes and including instruments and pure causes of the outcome at similar rates around 10-30%. More empirical investigation may be required to better understand how C-TMLE prioritizes covariates in practice. In the third scenario, C-LTMLE was able to greatly reduce the bias and variance compared to the LTMLE that included all unnecessary nonlinear terms. However, the residual bias did not decrease as the sample size increased, though the Monte Carlo variance did. Variance estimation for C-LTMLE was done using both the standard influence curve approach and a robust targeted method (Tran, 2016), evaluated with the full model for $g$. The influence curve approach performed well for C-LTMLE under the two larger

sample sizes investigated and the targeted approach overestimated in the simple setting with few covariates (as it did for LTMLE). Neither approach performed well in the presence of positivity violations. For C-LTMLE, this was likely because the cross-validation selector omitted true terms in the $g$-model, resulting in an influence function that depended on the estimation of $Q$. Future work is needed to investigate post-selection issues in C-TMLE and variance estimation under misspecification.

The application involved the estimation of the perinatal safety of taking a low dose of inhaled corticosteroids (ICS) in the first two trimesters of pregnancy in women with mild asthma. Missing outcome data required that censoring models be added to the C-LTMLE procedure. No methods suggested that there is an effect on low birth weight, having a preterm birth, or delivering a baby with a weight in the lowest 10th percentile. The C-LTMLE slightly shifted the point estimates for the two latter outcomes but this did not change the interpretation. The sandwich variance estimate for C-LTMLE was greater than the bootstrapped estimate for the G-Computation and much less than that of IPTW.

By introducing our Longitudinal C-TMLE approach, we hope to stimulate development and discussion of variable selection methods for saturated and unsaturated marginal structural models. Without variable or model selection approaches, fitting the required models may prove challenging in areas where the number of potential time-varying confounders is large. The improvement of such techniques is likely to facilitate the adoption of longitudinal causal inference methods in substantive research areas.

## References

Bahamyirou, A., Blais, L., Forget, A., and Schnitzer, M. E. (2018). Understanding and diagnosing the potential for bias when using machine learning methods with doubly robust causal estimators. *Statistical Methods in Medical Research* **0,** in press.

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61,** 962–972.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives* **28,** 29–50.

Benkeser, D., Carone, M., Van der Laan, M. J., and Gilbert, P. B. (2017). Doubly robust nonparametric inference on the average treatment effect. *Biometrika* **104,** 863–880.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., and Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology* **163,** 1149–1156.

Cossette, B., Forget, A., Beauchesne, M., Rey, E., Lemière, C., Larivée, P., Battista, M.-C., and Blais, L. (2013). Impact of maternal use of asthma-controller therapy on perinatal outcomes. *Thorax* **68,** 724–730.

Crainiceanu, C. M., Dominici, F., and Parmigiani, G. (2008). Adjustment uncertainty in effect estimation. *Biometrika* **95,** 635–651.

De Luna, X., Waernbaum, I., and Richardson, T. S. (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika* **98,** 861–875.

Ertefaie, A., Asgharian, M., and Stephens, D. A. (2017). Variable selection in causal inference using a simultaneous penalization method. *Journal of Causal Inference* **6,**.

Gruber, S. and van der Laan, M. J. (2011). C-tmle of an additive point treatment effect. In van der Laan, M. J. and Rose, S., editors, *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer Series in Statistics.

Hernán, M. A., Hernández-Diaz, S., Werler, M. M., and Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *American Journal of Epidemiology* **155,** 176–184.

Ju, C., Gruber, S., Lendle, S. D., Franklin, J. M., Wyss, R., Schneeweiss, S., and van der Laan, M. J. (2017). Scalable collaborative targeted learning for large scale and high-dimensional data. *Statistical Methods in Medical Research* **0,** in press.

Lendle, S. D., Schwab, J., Petersen, M. L., and van der Laan, M. J. (2017). ltmle: An R package implementing targeted minimum loss-based estimation for longitudinal data. *Journal of Statistical Software* **81,** 1–21.

Murphy, V. E., Namazy, J. A., Powell, H., Schatz, M., Chambers, C., Attia, J., and Gibson, P. G. (2011). A meta-analysis of adverse perinatal outcomes in women with asthma. *BJOG: An International Journal of Obstetrics and Gynaecology* **118,** 1314–1323.

National Asthma Education and Prevention Program expert panel (2005). Managing asthma during pregnancy: recommendations for pharmacologic treatment – update 2004. *J Allergy Clin Immunol* **115,** 34–46.

Neugebauer, R., van der Laan, M. J., Joffe, M. M., and Tager, I. B. (2007). Causal inference in longitudinal studies with history-restricted marginal structural models. *Electronic Journal of Statistics* **1,** 119–154.

Petersen, M., Schwab, J., Gruber, S., Blaser, N., Schomaker, M., and van der Laan, M. J. (2013). Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *Journal of Causal Inference* **2,** 147–185.

Petersen, M. L., Wang, Y., van der Laan, M. J., Guzman, D., Riley, E., and Bangsberg, D. R.

(2007). Pillbox organizers are associated with improved adherence to hiv antiretroviral therapy and viral suppression: a marginal structural model analysis. *Clinical Infectious Diseases* **45,** 908–915.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Robins, J. M. (1998). Marginal structural models. *1997 Proceedings of the American Statistical Association. Section on Bayesian Statistical Science* pages 1–10.

Robins, J. M. (2001). Data, design, and background knowledge in etiologic inference. *Epidemiology* **11,** 313–320.

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

Schnitzer, M. E. and Cefalu, M. (2018). Collaborative targeted learning using regression shrinkage. *Statistics in Medicine* **37,** 530–543.

Schnitzer, M. E., Lok, J. J., and Gruber, S. (2016). Variable selection for confounder control, flexible modeling and collaborative targeted minimum loss-based estimation in causal inference. *International Journal of Biostatistics* **12,** 97–115.

Shi, C., Fan, A., Song, R., and W, L. (2018). High-dimensional A-learning for optimal dynamic treatment regimes. *Annals of Statistics* **46,** 925–957.

Shortreed, S. M. and Ertefaie, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics* **73,** 1111–1122.

Tran, L. M. (2016). *Comparative Causal Effect Estimation and Robust Variance for Longitudinal Data Structures with Applications to Observational HIV Treatment.* PhD thesis, U.C. Berkeley.

van der Laan, M. J. and Dudoit, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: Finite

sample oracle inequalities and examples. *U.C. Berkeley Division of Biostatistics Working Paper Series* **Working Paper 130,**.

van der Laan, M. J. and Gruber, S. (2010). Collaborative double robust targeted maximum likelihood estimation. *The International Journal of Biostatistics* **6,** Article 17.

van der Laan, M. J. and Gruber, S. (2012). Targeted minimum loss based estimation of an intervention specific mean outcome. *The International Journal of Biostatistics* **8,** Article 9.

van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality.* Springer Series in Statistics. Springer Verlag: New York, NY.

van der Laan, M. J. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer Series in Statistics. Springer, New York, NY.

van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics* **2,** Article 11.

VanderWeele, T. J. and Shpitser, I. (2011). A new criterion for confounder selection. *Biometrics* **67,** 1406–1413.

Vansteelandt, S., Bekaert, M., and Claeskens, G. (2012). On model selection and model misspecification in causal inference. *Statistical Methods in Medical Research* **21,** 7–30.

Wang, C., Parmigiani, G., and Dominici, F. (2012). Bayesian effect estimation accounting for adjustment uncertainty. *Biometrics* **68,** 661–671.

Wilson, A. and Reich, B. J. (2014). Confounder selection via penalized credible regions. *Biometrics* **70,** 852–861.

SUPPORTING INFORMATION

Web Appendices referenced in Sections 3, 4, and 5, including the data-generating code for the simulation studies, are available with this paper at the Biometrics website on Wiley Online Library.

**Table 1**

*Simple scenario simulation study results. Root mean squared error, mean bias, Monte Carlo standard error, and estimated standard error and coverage. $SE_{IC}$ is calculated from the influence curve; $SE_{TE}$ is the TMLE estimate of the standard error from the ltmle package. All models are specified on the full sets of covariate histories. True value $E(Y^{1,1}) = 0.286$.*

|  | GCOMP | IPTW | C-IPTW | LTMLE | C-LTMLE |
|---|---|---|---|---|---|
| $n = 250$ |  |  |  |  |  |
| rMSE | 0.045 | 0.165 | 0.090 | 0.076 | 0.055 |
| Mean Bias | -0.002 | -0.019 | -0.036 | -0.007 | -0.005 |
| MC SE | 0.045 | 0.164 | 0.083 | 0.076 | 0.055 |
| $SE_{IC}$(%Cov) | - | - | 0.069(78) | 0.057(78) | 0.057(91) |
| $SE_{TE}$(%Cov) | - | 0.213(93) | 0.213(98) | 0.160(99) | 0.160(99) |
| $SE_{B}$(%Cov) | 0.044(94) | - | - | - | - |
|  |  |  |  |  |  |
| $n = 500$ |  |  |  |  |  |
| rMSE | 0.031 | 0.125 | 0.059 | 0.055 | 0.038 |
| Mean Bias | 0.001 | -0.023 | -0.012 | -0.001 | -0.001 |
| MC SE | 0.031 | 0.123 | 0.058 | 0.055 | 0.038 |
| $SE_{IC}$(%Cov) | - | - | 0.052(87) | 0.043(82) | 0.043(95) |
| $SE_{TE}$(%Cov) | - | 0.122(91) | 0.122(99) | 0.090(99) | 0.090(100) |
| $SE_{B}$(%Cov) | 0.031(95) | - | - | - | - |
|  |  |  |  |  |  |
| $n = 1,000$ |  |  |  |  |  |
| rMSE | 0.022 | 0.107 | 0.037 | 0.041 | 0.026 |
| Mean Bias | -0.001 | -0.018 | -0.007 | -0.003 | -0.002 |
| MC SE | 0.022 | 0.106 | 0.037 | 0.040 | 0.026 |
| $SE_{IC}$(%Cov) | - | - | 0.040(93) | 0.033(83) | 0.033(97) |
| $SE_{TE}$(%Cov) | - | 0.085(84) | 0.085(99) | 0.061(99) | 0.061(100) |
| $SE_{B}$(%Cov) | 0.021(95) | - | - | - | - |

**Table 2**
*Correlated covariates simulation study results. Root mean squared error, mean bias, Monte Carlo standard error,
and estimated standard error and coverage. $SE_{IC}$ is calculated from the influence curve; $SE_{TE}$ is the TMLE
estimate of the standard error from the ltmle package. True value $E(Y^{1,1}) = 0.622$.*

|  | GCOMP | IPTW | C-IPTW | LTMLE | C-LTMLE |
|---|---|---|---|---|---|
| $n = 250$ |  |  |  |  |  |
| rMSE | 0.051 | 0.158 | 0.095 | 0.086 | 0.069 |
| Mean Bias | -0.025 | -0.010 | 0.029 | 0.002 | 0.001 |
| MC SE | 0.044 | 0.158 | 0.090 | 0.086 | 0.069 |
| $SE_{IC}$(%Cov) | - | - | 0.078(84) | 0.064(79) | 0.064(89) |
| $SE_{TE}$(%Cov) | - | 0.185(94) | 0.185(99) | 0.152(98) | 0.152(98) |
| $SE_B$(%Cov) | 0.047(94) | - | - | - | - |
| |  |  |  |  |  |
| $n = 500$ |  |  |  |  |  |
| rMSE | 0.039 | 0.065 | 0.073 | 0.051 | 0.045 |
| Mean Bias | -0.025 | -0.004 | 0.007 | -0.003 | -0.004 |
| MC SE | 0.030 | 0.065 | 0.072 | 0.051 | 0.045 |
| $SE_{IC}$(%Cov) | - | - | 0.053(83) | 0.047(91) | 0.047(95) |
| $SE_{TE}$(%Cov) | - | 0.066(95) | 0.066(93) | 0.056(96) | 0.056(98) |
| $SE_B$(%Cov) | 0.031(87) | - | - | - | - |
| |  |  |  |  |  |
| $n = 1,000$ |  |  |  |  |  |
| rMSE | 0.031 | 0.037 | 0.051 | 0.033 | 0.030 |
| Mean Bias | -0.022 | 0.000 | -0.006 | 0.003 | 0.000 |
| MC SE | 0.021 | 0.037 | 0.050 | 0.033 | 0.030 |
| $SE_{IC}$(%Cov) | - | - | 0.036(82) | 0.033(94) | 0.033(95) |
| $SE_{TE}$(%Cov) | - | 0.040(96) | 0.040(86) | 0.035(97) | 0.035(97) |
| $SE_B$(%Cov) | 0.021(83) | - | - | - | - |

**Table 3**

*Continuous outcome with potential practical positivity violations simulation study results. All models are fit on the full data, including main terms, first-order interactions, and squared terms with the exception of $IPTW_{MT}$ and $LTMLE_{MT}$ which include models fit on the main terms only. The collaborative methods select from all main terms, first-order interactions, and squared terms. Root mean squared error, mean bias, Monte Carlo standard error, and mean estimated standard error were exceptionally calculated with 10% trimmed means due to outliers. $SE_{IC}$ is calculated from the influence curve; $SE_{TE}$ is the TMLE estimate of the standard error from the ltmle package. True value $E(Y^{1,1}) = 10.39$.*

|  | GCOMP | IPTW | $IPTW_{MT}$ | C-IPTW | LTMLE | $LTMLE_{MT}$ | C-LTMLE |
|---|---|---|---|---|---|---|---|
| $n = 250$ |  |  |  |  |  |  |  |
| rMSE | 0.53 | 4.55 | 1.00 | 2.22 | 2.53 | 0.78 | 0.65 |
| Mean Bias | -0.32 | -2.27 | -0.15 | 2.13 | 1.32 | -0.22 | 0.21 |
| MC SE | 0.44 | 4.26 | 0.99 | 0.77 | 2.37 | 0.74 | 0.66 |
| $SE_{IC}$(%Cov) | - | - | - | 0.99(43) | 0.72(34) | 0.98(92) | 0.72(86) |
| $SE_{TE}$(%Cov) | - | 0.85(25) | 1.04(90) | 0.87(35) | 0.56(32) | 0.94(91) | 0.56(82) |
| $SE_B$(%Cov) | 0.53(90) | - | - | - | - | - | - |
| $n = 500$ |  |  |  |  |  |  |  |
| rMSE | 0.45 | 4.42 | 0.74 | 2.15 | 1.62 | 0.53 | 0.56 |
| Mean Bias | -0.32 | -2.90 | -0.12 | 2.11 | 0.89 | -0.10 | 0.26 |
| MC SE | 0.35 | 3.72 | 0.73 | 0.54 | 1.51 | 0.52 | 0.54 |
| $SE_{IC}$(%Cov) | - | - | - | 0.84(31) | 0.50(43) | 0.84(96) | 0.50(82) |
| $SE_{TE}$(%Cov) | - | 0.71(21) | 0.78(91) | 0.71(22) | 0.42(42) | 0.80(93) | 0.42(77) |
| $SE_B$(%Cov) | 0.38(82) | - | - | - | - | - | - |
| $n = 1,000$ |  |  |  |  |  |  |  |
| rMSE | 0.36 | 4.12 | 0.55 | 2.13 | 1.20 | 0.40 | 0.50 |
| Mean Bias | -0.30 | -3.29 | -0.08 | 2.11 | 0.68 | -0.09 | 0.27 |
| MC SE | 0.23 | 2.88 | 0.54 | 0.38 | 1.10 | 0.39 | 0.47 |
| $SE_{IC}$(%Cov) | - | - | - | 0.75(17) | 0.37(47) | 0.67(96) | 0.37(80) |
| $SE_{TE}$(%Cov) | - | 0.59(20) | 0.58(91) | 0.58(10) | 0.31(45) | 0.64(94) | 0.31(75) |
| $SE_B$(%Cov) | 0.27(79) | - | - | - | - | - | - |

**Table 4**
*Application results for longitudinal C-LTMLE, LTMLE, IPTW, and G-Computation, implemented with logistic regressions. The parameter estimated is $E(Y^{1,1})/E(Y^{0,0})$. $n = 5,048$*

| Outcome | | GCOMP | IPTW | LTMLE | C-LTMLE |
|---|---|---|---|---|---|
| SGA | | | | | |
| | Estimate | 1.02 | 1.02 | 1.05 | 1.02 |
| | $SE_{IC}$ | 0.09 | 0.16 | 0.10 | 0.10 |
| | 95% CI | (0.87,1.21) | (0.71,1.30) | (0.86,1.25) | (0.83,1.20) |
| | n. vars selected $\{(1,1),(0,0)\}$ | - | - | - | (0,0) |
| Premature | | | | | |
| | Estimate | 1.00 | 0.91 | 0.94 | 0.97 |
| | $SE_{IC}$ | 0.10 | 0.16 | 0.11 | 0.12 |
| | 95% CI | (0.81,1.22) | (0.76,1.38) | (0.72,1.16) | (0.73,1.20) |
| | n. vars selected $\{(1,1),(0,0)\}$ | - | - | - | (35,35) |
| LBW | | | | | |
| | Estimate | 1.11 | 1.07 | 1.10 | 1.13 |
| | $SE_{IC}$ | 0.13 | 0.18 | 0.14 | 0.14 |
| | (95% CI) | (0.90,1.41) | (0.81,1.55) | (0.83,1.38) | (0.86,1.41) |
| | n. vars selected $\{(1,1),(0,0)\}$ | - | - | - | (31,0) |