Supporting Information for "Data-adaptive longitudinal model selection in causal inference with collaborative targeted minimum loss-based estimation"

Mireille E. Schnitzer, Joel Sango, Steve Ferreira Guerra,
and Mark J. van der Laan

Web Appendix A: More details about LTMLE and solving the efficient influence function estimating equation

The efficient influence function for ψ in the nonparametric model space is given as

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

at a given $Q = \{Q_2(\overline{L}_1), Q_1(L_0)\}$ and $g = \{g_1(\overline{L}_1), g_0(L_0)\}$.

The direct result of the Longitudinal TMLE (LTMLE) procedure is that the updated estimates of the conditional expectations of the outcome now solve the empirical efficient influence function equation. Specifically, the empirical average (taken over all subjects) of the values of

$$\frac{I(\overline{A}_1 = \overline{a})}{g_{1,n}(\bar{l}_1)g_{0,n}(l_0)}\{y - Q^*_{2,n}(\bar{l}_1)\} + \frac{I(A_0 = a_0)}{g_{0,n}(l_0)}\{Q^*_{2,n}(\bar{l}_1) - Q^*_{1,n}(l_0)\} + Q^*_{1,n}(l_0) - \psi_n$$

is equal to zero (y is the value of Y for a sampled subject) where ψ_n is the LTMLE estimate. This is ensured by the procedure because the logistic regression update steps produce estimates of $\epsilon_t, t = 1, 0$, that solve the logistic regression score equations. The consequence of solving this influence function estimating equation is that, under some regularity conditions, ψ_n is a locally efficient and double robust estimator of ψ with a large-sample variance that can be approximated by $Var_n\{D(\psi_n, Q_n, g_n)(O)\}/n$ (van der Laan and Robins, 2003).

Consider data-generating distribution P belonging to the model space $\mathcal{M}$. We use the empirical process notation $Pf = \int f(z)dP(z)$ for the expectation that averages over

the randomness in Z but not the P-integrable function f. The $L_2(P)$ norm is given as $\|f\|^2 = Pf^2$. With respect to the required convergence rates of the so-called "nuisance" models, the single time-point setting requires that the product of the errors (i.e. the root-$L_2(P)$ norm of the difference between the estimate and the true function) for the probability of treatment and the conditional expectation of the outcome must converge at a rate of $O_p(n^{-1/2})$ to ensure $n^{-1/2}$ convergence of the doubly robust estimator for the expectation of the counterfactual outcome under a given treatment (Kennedy, 2016). This implies that if one error term converges to zero at a parametric rate or if both terms converge at $n^{-1/4}$ rates then this condition is satisfied. For the two treatment time setting, we evaluate

$$\begin{aligned} &P[D(\psi, Q_n, g_n) - D(\psi, Q, g)] = \\ &\quad P\left[\frac{g_0}{g_{1,n}g_{0,n}}\{g_1 - g_{1,n}\}\{Q_{2,n} - Q_2\}\right] + \qquad (2) \\ &\quad P\left[\frac{g_0 - g_{0,n}}{g_{0,n}}\{Q_{1,n} - Q_1\}\right]. \qquad (3) \end{aligned}$$

By the Cauchy-Schwartz inequality and the positivity assumptions, the absolute value of expectation (3) is bounded above by

$$\|g_0 - g_{0,n}\| \, \|Q_{1,n} - Q_1\| .$$

The absolute value of the expectation (2) is bounded above by

$$\left\| \frac{g_0}{g_{1,n}g_{0,n}} \right\| \|g_1 - g_{1,n}\| \, \|Q_{2,n} - Q_2\|$$

Assuming that the first term is greater than one (likely for reasonable estimation of g_1 and g_0), the convergence of expectation (2) is similarly bounded by the product of the estimation errors of the treatment probability and outcome expectation at the second time point. If both terms (2) and (3) converge to zero at $n^{-1/2}$ rates, the LTMLE will also converge to the true value of ψ at a $n^{-1/2}$ rate. Therefore if both models (for treatment and outcome) at both time points converge at rates of $n^{-1/4}$, this condition is satisfied. If only one of the models at each time point converges to the truth, then this rate must be $n^{-1/2}$. These rates of convergence are also required for the C-LTMLE estimator for the asymptotically selected $\tilde{g}^{(k)}$ and $\tilde{Q}^{*,(k)}$.

Web Appendix B: C-LTMLE algorithm for arbitrary treatment and censoring times

In the more general case (van der Laan and Gruber, 2012; Schnitzer et al., 2014), we observe longitudinal data $O = (L_0, A_0, L_1, A_1, ...L_{T-1}, A_{T-1}, Y)$ where L_t are a set of covariates measured at time $t = 0, ..., T-1$ and A_t is a possibly multivariate treatment and censoring node at time $t = 0, ..., T-1$. Let Y be the outcome of interest, measured at the end of the study after a fixed follow-up. Let $Y^{\overline{a}}$ denote the counterfactual outcome (Rubin, 1974) under the fixed treatment regime $\overline{a}$ and under no censoring, and

similarly let $L_t^{\overline{a}_{t-1}}$ represent an intermediate outcome under past treatment regime $\overline{a}_{t-1}$ and no censoring. Our goal is to estimate $\psi = E(Y^{\overline{a}})$, the marginal mean outcome under some fixed treatment regime $\overline{a}$. Throughout, a subscript n will be used to denote an estimate of a quantity.

Firstly, estimate each probability of treatment, i.e. $g_t(\bar{l}_t) = Pr(A_t = a_t \mid \overline{A}_{t-1} = \overline{a}_{t-1}, \overline{L}_t = \bar{l}_t)$. This may correspond to fitting one model for each time point's treatment and censoring nodes and using these models to predict each subject's probability of being uncensored and following the regimen at that time point conditional on being uncensored and following the regimen up to time $t-1$. We define the estimates of the predicted probability at time t as $g_{t,n}(\bar{l}_t)$.

Algorithm 1 LTMLE algorithm for the general case

1: Initialize $Q^*_{T+1} = Y$.
2: **for** t=T,...,1 **do**
3: Estimate $Q_t = E(Q_{t+1} \mid \overline{A}_{t-1} = \overline{a}_{t-1}, \overline{L}_{t-1} = \bar{l}_{t-1})$, the conditional expectation of the outcome under treatment regime $\overline{a}_{t-1}$ for observed data history $\bar{l}_{t-1}$. This is done by by regressing $Q^*_{t+1,n}$ on $\overline{A}_{t-1}$ and $\overline{L}_{t-1}$ and making a prediction for each subject setting $A_0 = a_0$. Denote this estimate as $Q_{t,n}$.
4: Run intercept-free logistic regression $Q^*_{t+1,n} \sim \hat{\epsilon}_t / \{\prod_{l=0}^{t-1} g_{l,n}\} + offset[logit\{Q_{t,n}\}]$ using subjects with $\overline{A}_{t-1} = \overline{a}_{t-1}$. Let $\hat{\epsilon}_t$ be the coefficient estimate.
5: Set $Q^*_{t,n} = expit\left[\overline{\epsilon}_t / \{\prod_{l=0}^{t-1} g_{l,n}\} + logit\{Q_{t,n}\}\right]$ for all subjects.
6: **end for**
7: Set $\psi_n = mean(Q^*_{1,n})$ to be the targeted estimate.

The C-LTMLE procedure as written in the main document applies immediately for the multiple time and multivariate treatment and censoring nodes setting. The first step is now again to estimate the initial $g_t^{(0)}(\bar{l}_t) = Pr(A_t = a_t \mid \overline{A}_{t-1} = \overline{a}_{t-1}, \overline{L}_t = \bar{l}_t)$ for all time points. The corresponding models may include some "forced-in" covariates or may be unconditional on covariates. The next step is to create initial estimates – that may be targeted – of each $Q_t, t = T, ..., 1$. Generally, one would start by running a regression to estimate $Q_T = E(Y \mid \overline{A}_{t-1} = \overline{a}_{t-1}, \overline{L}_{t-1} = \bar{l}_{t-1})$. The next step could involve performing the TMLE update step with respect to the estimates of $g_t^{(0)}(\bar{l}_t), t = 0, ..., T-1$. From these updated values, we run another regression to estimate $Q_{T-1} = E\{Q_T(\bar{l}_{t-1}) \mid \overline{A}_{t-2} = \overline{a}_{t-2}, \overline{L}_{t-2} = \bar{l}_{t-2}\}$, and so on. Thus we obtain our T n-vectors of initial estimates $Q_n^B = Q_n^{init}$.

The "allowable moves" now include any covariate (or nonlinear term) addition to any of the models that were used to make the predictions at any time point. For example, in the application of Section 5 in the main document, there were both treatment and censoring models at times 1 and 0. A move corresponded to the addition of a covariate to any of these four models. One must restrict the covariate additions to only allow treatment or censoring to be conditional on past covariates. Additional restrictions can be made by the user.

An important note is that when we update the $Q_n^{*,(k)}$, we do so at each time point separately. For example, in the simple two time point case, if we add a variable to the model for g_0, we obtain a new estimate of g_0, and so we must update the values of $Q_1^{*,(k)}$ and $Q_2^{*,(k)}$ in order to check the mean error corresponding to the move. We do so by directly applying lines 4 and 5 of Algorithm 1 to $Q_{1,n}^{*,(k)}$ and $Q_{2,n}^{*,(k)}$. We do not regress the new estimate of Q_2 on the covariate history to get a new estimate of Q_1. Instead, we directly apply the update steps to the given $Q_{1,n}^{*,(k)}$ and $Q_{2,n}^{*,(k)}$ separately. This allows us to guarantee that the error in $Q_n^{*,(k)}$ and $g_n^{(k)}$ is consistently improving over $k = 1, ..., K$.

As pointed out by a reviewer, a concern arises when there are not many subjects following each regime of interest, which is increasingly likely for larger numbers of time points at which treatment may change or censoring may occur. This may result in poor estimation of the initial Q_t (step 3) and/or an ineffective update (step 4).

Web Appendix C: An intuitive understanding of how C-TMLE prioritizes covariates

C-TMLE has been shown to reduce bias and variance and appropriately select confounders (Gruber and van der Laan, 2011; Porter et al., 2011; Schnitzer et al., 2016). Intuitively, if a variable added to g is not associated with the treatment (beyond the other variables already included in the model), the g_n estimate will not improve, the TMLE update step will not change the previously estimated $Q_n^{*,(k)}$, and thus the cross-validated loss-function-based error will not improve. Alternatively, if a variable added to g is associated with the treatment but not the outcome, then while the g_n estimate will change, the new g_n will not be better at explaining the residual bias in Q_n. Such an addition may also increase the variability of the TMLE (Brookhart et al., 2006; Schnitzer et al., 2016) and thus will worsen the cross-validated penalized error. However, if an added variable is associated with both treatment and outcome beyond the variables already included, its addition will modify the prediction of g, this new information will be associated with the residual bias in $Q_n^{*,(k)}$ (i.e. ϵ will be non-zero in the TMLE update), and in performing the TMLE update step, the cross-validated error will likely be reduced for the new updated $Q_n^{*,(k)}$ (depending on the increase in the variance penalty). The selection will therefore be more likely to select variables that are predictive of both the treatment and outcome unless they greatly inflate the variance, and be less likely to select variables that act conditionally like instruments. Note that by selecting covariates into the model for g that explain the residual error in Q_n, we are fulfilling the collaborative double robustness criterion.

Web Appendix D: Data-generation and additional results for the simulation studies

Web Appendix D.1: Simple setting

We independently sampled data $O = (L_0, A_0, L_1, A_1, Y)$ in sequence as follows:

$$
\begin{aligned}
IV_0 &\sim N(0,1); \quad R_0 \sim N(0,1); \quad W_0 \sim N(0,1) \\
A_0 &\sim Bern(p = expit(IV_0 + W_0)) \\
IV_1 &\sim N(0,1); \quad R_1 \sim N(1/2R_0, 1); \quad W_1 \sim N(1/2W_0 + 1/2A_0, 1) \\
A_1 &\sim Bern(p = expit(A_0 + IV_0 + IV_1 + W_1)) \\
Y &\sim Bern(p = expit(-2 + 1/2A_0 + 1/2A_1 - W_0 - W_1 + 1/2R_0 + 1/2R_1)).
\end{aligned}
$$

Table 1 gives the frequency of selection of each covariate into the models for g_0 and g_1. C-LTMLE successfully selected the variables W_0 and W_1 into the models for g_0 and g_1 respectively, with high frequencies at $n = 250$ increasing to 100% by $n = 1,000$. The variable W_0 was only selected into the model for g_1 with a frequency of 20% because W_0 is not a confounder of the relationship between A_1 and Y. Similarly, the instruments were only selected between $10 - 30\%$ of the time (consistent with previous results (Schnitzer et al., 2016)) and the pure risk factors were selected slightly more often.

Web Table 1: Frequency of selection, Q_n only adjusting for treatment.

	$n = 250$		$n = 500$		$n = 1,000$	
	g_1	g_0	g_1	g_0	g_1	g_0
W_0	0.2	0.7	0.2	0.9	0.2	1.0
IV_0	0.3	0.3	0.2	0.3	0.1	0.3
R_0	0.2	0.3	0.3	0.4	0.3	0.4
W_1	0.6	-	0.9	-	1.0	-
IV_1	0.3	-	0.3	-	0.3	-
R_1	0.3	-	0.3	-	0.3	-

The data-generating function in R code is below.

```
datagen_xx<-function(n,seed=sample(1:100000,size=1)){

  set.seed(seed)
  In0=rnorm(n, 0, 1)
  L0=rnorm(n, 0, 1)
  R0=rnorm(n, 0, 1)
  In1=rnorm(n, 0, 1)
  #A0
  p_A0=plogis(In0+L0)
  A0=rbinom(n, 1, p_A0)
```

```
  #R1
  R1=rnorm(n,0.5*R0,1)
  #L1
  m_L1=0.5*L0+0.5*A0
  L1=rnorm(n,mean=m_L1,sd=1)
  #A1
  p_A1=plogis(A0+In1+In0+L1)
  A1=rbinom(n,1,p_A1)
  #Y
  p_Y=plogis(-2+0.5*A0+0.5*A1-L0-L1+0.5*R0+0.5*R1)
  Y=rbinom(n, 1, p_Y)

  truth<-  0.2858638

  #
  X=data.frame(In0, L0, In1, R0, A0, R1, L1 ,A1,Y)

 return(X)

}
```

Web Appendix D.2: Correlated covariates

We independently sampled data $O = (L_0, A_0, L_1, A_1, Y)$ in sequence as follows:

We generated a 40-dimensional multivariate normal random variable

$$L_0 \sim MVN(0, \Sigma)$$

where Σ is a 40×40 variance-covariance matrix with diagonal entries equal to 1 and off-diagonal entries equal to 0.2 and L_0 will be considered a random row vector.

We then generated treatment at the first time point as

$$A_0 \sim Bern(p = expit(L_0\beta_{A0}))$$

with column vector

$$\begin{aligned}\beta_{A0} = (&0.1, 0.1, 0.1, 0.1, 0.1, -0.1, -0.1, -0.1, -0.1, -0.1,\\ &0.1, 0.1, 0.1, 0.1, 0.1, -0.1, -0.1, -0.1, -0.1, -0.1,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^t.\end{aligned}$$

We then generated $L_1 = (W_1, IV_1, R_1)$ as ten independent normal random variables. Two true confounders W_1 were each generated independently according to $N(L_0\beta_{W1} +$

$0.5A_0, 0.5^2)$ with column vector

$$\begin{aligned}\beta_{W1} = (&0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2,\\ &0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^t\end{aligned}$$

and then bounded between -4 and 4. Two near-instruments IV_1 were generated independently according to $N(L_0\beta_{IV1}, 0.5^2)$ where

$$\begin{aligned}\beta_{IV1} = (&0, 0, 0, 0, 0, 0, 0, 0, 0, 0,\\ &0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^t\end{aligned}$$

and then bounded between -4 and 4. Three pure risk factors R_1 were generated independently according to $N(L_0\beta_{R1}, 0.5^2)$ where

$$\begin{aligned}\beta_{R1} = (&0, 0, 0, 0, 0, 0, 0, 0, 0, 0,\\ &0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,\\ &0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^t\end{aligned}$$

and then bounded between -5 and 5.

We generated treatment at the second time point as

$$A_1 \sim Bern(p = expit(L\beta_{A1}))$$

where $L = (L_0, L_1)$ and

$$\begin{aligned}\beta_{A1} = (&0.1, 0.1, 0.1, 0.1, 0.1, -0.1, -0.1, -0.1, -0.1, -0.1,\\ &0.1, 0.1, 0.1, 0.1, 0.1, -0.1, -0.1, -0.1, -0.1, -0.1,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0,\\ &0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0, 0, 0)^t.\end{aligned}$$

Finally, we generated the outcome

$$Y \sim Bern(p = expit(L\beta_Y + 0.5A_0 + 0.5A_1 + 0.1A_0L_0^{(1)} - 0.05A_1L_1^{(2)}))$$

where

$$\begin{aligned}\beta_Y = (&0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1,\\ &0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02,\\ &0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1,\\ &0, 0, 0, 0, 0, 0, 0, 0, 0, 0,\\ &0.1, 0.1, 0.1, 0.1, 0.1, 0.02, 0.02, 0.1, 0.1, 0.1)^t,\end{aligned}$$

$L_0^{(1)}$ represents the first element of L_0, and $L_1^{(2)}$ represents the second element of L_1.

The code is below. An intuitive explanation can be found in the main manuscript.

```
library(MASS)


logit<-function(x){
 return(log(x/(1-x)))
}
expit<-function(x){
z<-1/(1+exp(-x))
return(z)
}
bound.ms<-function(x,min,max){
x[x>max]<-max
x[x<min]<-min
return(x)
}

datagen_corr<-function(n,seed=sample(1:100000,size=1)){

set.seed(seed)

#40 covariates
#covariance matrix
vcvmat0<-matrix(c(rep(c(1,rep(0.2,40)),39),1),nrow=40)
L0<-mvrnorm(n=n,mu=rep(0,40), Sigma=vcvmat0)
L0<-as.data.frame(bound.ms(L0,min=-3,max=3)) #bounding has no real impact
names(L0)<-paste("L0",1:40,sep="")

#10 confounders, 10 almost instruments, 10 pure risk factors,
#10 no effect on either

#A0
truecoefsA0<-c(rep(0.1,5),rep(-0.1,5),rep(0.1,5),rep(-0.1,5),
  rep(0.00,20))
pA0<-expit( as.matrix(L0)%*%truecoefsA0)
A0<-rbinom(n=n,size=1,p=pA0)

truecoefsL1<- rbind(c(rep(0.2,10),rep(0.05,10),rep(0,20)),
   c(rep(0,10),rep(0.2,10),rep(0,20)),
   c(rep(0,10),rep(0.05,10),rep(0.2,10),rep(0,10))
) #the three categories have different means, but same 3 means for
  #each subject

#5 confounders
```

```
muW1<-as.matrix(L0)%*%truecoefsL1[1,]+0.5*A0
W1<-cbind(rnorm(n=n,mean=muW1,sd=0.5),rnorm(n=n,mean=muW1,sd=0.5),
  rnorm(n=n,mean=muW1,sd=0.5),rnorm(n=n,mean=muW1,sd=0.5),
rnorm(n=n,mean=muW1,sd=0.5))
W1<-bound.ms(W1,min=-4,max=4)

#2 near-instruments
muIn1<-as.matrix(L0)%*%truecoefsL1[2,]
In1<-cbind(rnorm(n=n,mean=muIn1,sd=0.5),rnorm(n=n,mean=muIn1,sd=0.5))
In1<-bound.ms(In1,min=-4,max=4)

#3 pure risk factors
muR1<-as.matrix(L0)%*%truecoefsL1[3,]
R1<-cbind(rnorm(n=n,mean=muR1,sd=0.5),rnorm(n=n,mean=muR1,sd=0.5),
  rnorm(n=n,mean=muR1,sd=0.5))
R1<-bound.ms(R1,min=-5,max=5)

L1<-as.data.frame(cbind(W1,In1,R1))
names(L1)<-paste("L1",1:10,sep="")

#A1
truecoefsA1<-c(rep(0.1,5),rep(-0.1,5),rep(0.1,5),rep(-0.1,5),rep(0.00,20),
  rep(0.1,5),rep(0.1,2),rep(0,3))
pA1<-expit( as.matrix(cbind(L0,L1))%*%truecoefsA2)
A1<-rbinom(n=n,size=1,p=pA1)

#outcome 2 binary
truecoefsY<- c(rep(0.1,10),rep(0.02,10),rep(0.1,10),rep(0,10),rep(0.1,5),
  rep(0.02,2),rep(0.1,3))
pY<-expit((as.matrix(cbind(L0,L1))%*%truecoefsY+0.5*A0+0.5*A1+
  0.1*A0*L0[,1]-0.1*A1*L1[,2])/2)
Y<-rbinom(n=n,size=1,p=pY)

truth<-0.6220712

return(as.data.frame(cbind(L0,A0=A0,L1,A1=A1,Y=Y)))
}
```

Web Appendix D.3: Continuous outcome with potential practical positivity violations

We independently sampled data $O = (L_0, A_0, L_1, A_1, Y)$ in sequence as follows:

We generated five baseline covariates $L_0 = (W_0, IV_0, R_0)$, including two confounders

W_0, two instruments IV_0, and one pure risk factor R_0 according to

$$\begin{aligned} W_0 &\sim MVN\left\{\begin{pmatrix}0.5\\1\end{pmatrix}, \begin{pmatrix}2 & 1\\1 & 1\end{pmatrix}\right\}, \\ IV_0 &\sim MVN\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}2 & 0\\0 & 1.9\end{pmatrix}\right\}, \\ R_0 &\sim N(1, 1.5^2). \end{aligned}$$

The first treatment A_0 was generated according to

$$A_0 \sim Bern\Big[p = expit\Big\{0.2 + W_0^{(1)} + 0.3IV_0^{(1)} + W_0^{(1)}IV_0^{(1)} - 0.02(W_0^{(2)} + IV_0^{(2)})^2\Big\}\Big]$$

where a vector with a bracketed superscript refers to the corresponding element in the vector, e.g. $W_0 = (W_0^{(1)}, W_0^{(2)})$.

At the next time point, the bivariate covariate L_1 was generated as

$$L_1 \sim MVN\left\{\begin{pmatrix}(W_0^{(1)} - W_0^{(2)})^2 - cos(R_0) - 0.5A_0\\0.5(cos(W_0^{(2)}) + W_0^{(1)}R_0) + 0.5A_0\end{pmatrix}, \begin{pmatrix}2 & 1\\1 & 1\end{pmatrix}\right\}.$$

The next treatment A_1 was generated as

$$A_1 \sim Bern\Big[p = expit\Big\{0.2 + 0.5A_0 + W_1^{(1)} + 0.3IV_0^{(1)} + W_1^{(1)}IV_0^{(1)} - 0.02(W_1^{(2)} + IV_0^{(2)})^2\Big\}\Big].$$

Finally, the outcome is generated as

$$Y \sim N(1 + A_0 + A_1 + 0.5L_0^{(1)} + L_0^{(2)} + 0.25L_1^{(1)} + 0.5L_1^{(2)} + 0.5(L_0^{(1)} + R_0)^2 + 0.2(L_1^{(1)} + R_0)^2, 1).$$

```
datagen_cts <- function(n,seed=sample(1:100000,size=1)){

sigma <- matrix(c(2, 1, 1, 1), ncol=2)
W <- matrix(rnorm(n*2), ncol = nrow(sigma)) %*% chol(sigma)
W <- W + matrix(rep(c(.5, 1),each=n), byrow=FALSE, ncol=2)
I1 <- rnorm(n,mean=1, sd=2)
I2 <- rnorm(n,mean=1, sd=1.9)
P1 <- rnorm(n,mean=1, sd=1.5)
```

```
L0 <- data.frame(W[,1],W[,2],I1,I2,P1,W[,1]^2,W[,2]^2,I1^2,I2^2,P1^2,
  W[,1]*W[,2],W[,1]*I1,W[,1]*I2,W[,1]*P1,W[,2]*I1,W[,2]*I2,W[,2]*P1,I1*I2,
  I1*P1,I2*P1)
colnames(L0) <- c("W11", "W12", "I11", "I12",  "P11", "W11sq","W12sq",
  "I11sq","I12sq","P11sq","W11W12","W11I11","W11I12","W11P11","W12I11",
  "W12I12","W12P11","I11I12","I11P11","I12P11")

pA0<-plogis(0.2 + L0[,"W11"] + 0.3*L0[,"I11"] + L0[,"W11"]*L0[,"I11"]
  -0.02*(L0[,"W12"] + L0[,"I12"])^2 )
A0 <- rbinom(n,size=1,p=pA0  )

W2 <- matrix(rnorm(n*2), ncol = nrow(sigma)) %*% chol(sigma)
meanW21<-((L0[,"W11"] - L0[,"W12"])^2 - cos(L0[,"P11"])) - 0.5*A0
meanW22<-(cos(L0[,"W12"]) + L0[,"W11"]*L0[,"P11"])/2 + 0.5*A0
W2 <- W2 + matrix(c(meanW21, meanW22), byrow=FALSE, ncol=2)
L1<-data.frame(W2[,1],W2[,2],W2[,1]^2,W2[,2]^2,W2[,1]*W2[,2],
  W2[,1]*W[,1],W2[,1]*W[,2],W2[,1]*I1,W2[,1]*I2,W2[,1]*P1,
  W2[,2]*W[,1],W2[,2]*W[,2],W2[,2]*I1,W2[,2]*I2,W2[,2]*P1)
colnames(L1)<-c("W21","W22","W21sq","W22sq","W21W22",
  "W21W11","W21W12","W21I11","W21I12","W21P11",
  "W22W11","W22W12","W22I11","W22I12","W22P11")

pA1<-plogis(0.2 + A0/2 + L1[,"W21"] + 0.3*L0[,"I11"] + L1[,"W21"]*L0[,"I11"]
  -0.02*(L1[,"W22"] + L0[,"I12"])^2 )
A1 <- rbinom(n,size=1,p=pA1  )

Y <-  1 + A0 + A1 + 0.5*L0[,"W11"] + L0[,"W12"] + 0.25*L1[,"W21"] +
  0.5*L1[,"W22"] + 0.5*(L0[,"W11"] + L0[,"P11"])^2 + 0.2*(L1[,"W21"] +
  L0[,"P11"])^2 + rnorm(n)
return(data.frame(L0, A0, L1, A1,Y))
}
```

Web Appendix D.4: Box plots of simulation study with continuous outcome with potential practical positivity violations

Web Figure 1: Box plots of the estimates in the simulation study for each method, corresponding to the results in Table 3 of the main document.

Web Appendix E: Limitations of the Application

As in most observational studies, the credibility of the results presented here depends on the validity of the sequential ignorability (no unmeasured baseline or time-dependent confounders) assumption. As noted previously (Cossette et al., 2013), this dataset lacks information on smoking. Over 10% of Canadian women smoke during pregnancy (Al-Sahab et al., 2010), with possibly greater rates in Quebec (Takser et al., 2004), and as smoking may increase the usage of asthma medications and the risk of poor pregnancy outcomes, the exclusion of this variable may bias the analysis. Another concern involves the veracity of the consistency assumption. While we have divided up the study timeline into three time intervals (with endpoints corresponding with one year pre-pregnancy, the beginning of first trimester, the beginning of second trimester, and the time of delivery), subjects grouped under the same exposure categories may have, in reality, had different true exposure to ICS that do not correspond with a coherent intervention strategy. This may be due to different frequencies of usage, different doses of the medication, or breaks in usage that were not recorded in the electronic health data. If these differences in exposure (within exposure categories) were to have an effect on the outcome, the counterfactuals would no longer be well-defined. In addition, changes in exposure within the time-intervals caused by time-varying confounders at a finer scale than our chosen discretization can also bias the analysis.

References

B Al-Sahab, M Saqib, G Hauser, and H Tamim. Prevalence of smoking during pregnancy and associated risk factors among canadian women: a national survey. *BMC Pregnancy and Childbirth*, 10(1), 2010.

M A Brookhart, S Schneeweiss, K J Rothman, R J Glynn, J Avorn, and T Stürmer. Variable selection for propensity score models. *American Journal of Epidemiology*, 163(12):1149–1156, 2006.

B Cossette, A Forget, M Beauchesne, É Rey, C Lemière, P Larivée, M-C Battista, and L Blais. Impact of maternal use of asthma-controller therapy on perinatal outcomes. *Thorax*, 68(8):724–730, 2013.

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

E H Kennedy. Semiparametric theory and empirical processes in causal inference. Technical report, Cornell University Library, 2016. URL `https://arxiv.org/abs/1510.04740`.

K E Porter, S Gruber, M J van der Laan, and J S Sekhon. The relative performance of targeted maximum likelihood estimators. *The International Journal of Biostatistics*, 7(1):1–34, 2011.

D B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

M E Schnitzer, M J van der Laan, E E M Moodie, and R W Platt. Effect of breastfeeding on gastrointestinal infection in infants: A targeted maximum likelihood approach for clustered longitudinal data. *Annals of Applied Statistics*, 8(2):703–725, 2014.

M E Schnitzer, J J Lok, and S Gruber. Variable selection for confounder control, flexible modeling and collaborative targeted minimum loss-based estimation in causal inference. *International Journal of Biostatistics*, 12(1):97–115, 2016.

L Takser, J Lafond, M Bouchard, G St-Amour, and D Mergler. Manganese levels during pregnancy and at birth: relation to environmental factors and smoking in a southwest quebec population. *Environmental Research*, 95(2):119–125, 2004.

M J van der Laan and S Gruber. Targeted minimum loss based estimation of an intervention specific mean outcome. *The International Journal of Biostatistics*, 8(1): Article 9, 2012.

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