The application of target trials with longitudinal targeted maximum likelihood estimation to assess the effect of alcohol consumption in adolescence on depressive symptoms in adulthood

Yan Liu *Université de Montréal*

Mireille E. Schnitzer *Université de Montréal*

Ronald Herrera *Université de Montréal; Bayer AG, Berlin, Germany*

Iván Díaz *NYU Langone Health*

Jennifer O'Loughlin *Université de Montréal*

Marie-Pierre Sylvestre *Université de Montréal*

Time-varying confounding is a common challenge for causal inference in observational studies with time-varying treatments, long follow-up periods, and participant dropout. Confounder adjustment using traditional approaches can also be limited by data sparsity, weight instability and computational issues. The Nicotine Dependence in Teens (NDIT) study is a prospective cohort study involving 24 data collection cycles to date, among 1,294 students recruited from 10 high schools in Montreal, Canada, including follow-up into adulthood. Our aim is to estimate associations between the timing of alcohol initiation and the cumulative duration of alcohol use on depression symptoms in adulthood. Based on the target trials framework, we define intention-to-treat and as-treated parameters in a marginal structural model with sex as a potential effect-modifier. We then use the observational data to emulate the trials. For estimation, we use pooled longitudinal target maximum likelihood estimation (LTMLE), a plug-in estimator with double robust and local efficiency properties. We describe strategies for dealing with high-dimensional potential drinking patterns and practical positivity violations due to a long follow-up time, including modifying the effect of interest by removing sparsely observed drinking patterns from the loss function and applying longitudinal modified treatment policies to represent the effect of discouraging drinking.

Keywords: Adolescent, alcohol consumption, depression, longitudinal targeted maximum likelihood estimation, marginal structural model, sequential G-computation, target trials

---

Estimation of the effect of time-varying exposures in observational studies becomes methodologically challenging in the presence of time-dependent confounding, requiring statistical methods beyond the standard approaches (1, 2). Robins (3) proposed marginal structural models (MSMs) which model the potential outcome under an assigned treatment history (or "pattern"). Hernán and Robins (4) proposed the target trials framework to define causal effects, in particular MSM parameters, by means of a mapping of the observational analysis onto an analysis of a hypothetical randomized

controlled trial. The parameters of an MSM can be estimated with inverse probability of treatment weighting (IPTW) (2), G-computation (5, 6), augmented-IPTW estimators (7, 8) and more recently estimators based on the targeted maximum likelihood framework (9), in particular using longitudinal targeted maximum likelihood estimation (LTMLE) (10, 11). LTMLE estimates the effects of time-varying exposures on outcomes with the advantage of double-robustness, meaning that the estimator is consistent if either the models for treatments or outcome are correctly specified. LTMLE is efficient in a semiparametric statistical model if all models are estimated consistently. In addition, LTMLE can readily incorporate machine learning in the process of generating the initial estimates while providing valid statistical inference (11), thus avoiding bias due to model misspecification.

Though LTMLE has been successfully applied in different contexts (12–16), there exist data sparsity and high-dimensionality challenges (17). One solution to these challenges lies in defining hypothetical longitudinal interventions that shift an individual's propensity score, making them more or less likely to be exposed, corresponding to exposure encouragement or discouragement. The intervention can also be applied conditional on the observed exposure, e.g. only discouraging exposure for those who actually were exposed (18, 19).

In this paper, we demonstrate the application of LTMLE in a complex substantive application. We consider an observational cohort study called Nicotine Dependence in Teens (NDIT) (20, 21). Associations between alcohol use in adolescence and later risk-taking behaviours have been established in longitudinal studies (22, 23). Regarding assessment of the longitudinal effect of alcohol use on depressive symptoms in early adulthood, to the best of our knowledge, no study has adjusted for time-varying confounding using causal inference methods (24–27). We aim to study the effect of time of initiation and the cumulative duration of drinking in adolescence on depression in young adulthood. We take into account time-varying confounders that can also be caused by drinking in adolescence, including depressive symptoms (24, 28), smoking (29, 30), stress (31, 32), and participation in team sports (33, 34). We define two target trials that recruit adolescents who have not yet initiated regular drinking. Using working MSMs, we correspondingly define the "intention-to-treat" (ITT) and "as-treated" (AT) effects respectively, and investigate effect modification by sex. The working MSM represents a projection of a true causal relationship between exposures and the outcome onto a low-dimensional linear model. We then estimate the parameters of the two MSMs using

G-computation and LTMLE. We describe high-dimensionality and sparsity challenges encountered when estimating the AT effect and explore ways to address them.

NDIT data

The NDIT is a prospective longitudinal investigation of 1,294 grade 7 students recruited from 10 Montréal-area (Canada) high schools in 1999-2000 (21). Self-report questionnaires were administered at each of the 10 schools every three months for a total of 20 cycles from 1999 to 2005 (i.e., during the five years of high school). Mail or in-person questionnaires were administered in 2007/2008 (cycle 21) when participants were age 20.4 years on average. The data collected include repeated measures of a wide range of socio-demographic, substance use, psychosocial, lifestyle, and physical and mental health variables. Participants in the NDIT study completed self-report questionnaire at school every three months from grade 7 to 11. Figure S1 in Appendix I presents the structure of the follow-ups in the NDIT study. Parents and legal guardians provided informed consent, and all students provided assent and then consent in adulthood. The study was approved by the Ethics Research Committee of the *Centre de recherche du Centre Hospitalier de l'Université de Montréal*.

Exposure

Participants were asked "During the past three months, how often did you drink alcohol (beer, wine, hard liquor)?" We considered a participant exposed to regular alcohol use if the participant answered "once or a couple of times a week" or "usually every day" (alternatives were "never," "a bit to try" or "once or a couple of times a month"). Therefore, "alcohol use" in this paper refers to "at least weekly use". We correspondingly denote the binary exposure over time as $A_t, t \in (0, \cdots, 19)$ with $A_t = 1$ representing exposure and $A_t = 0$ representing unexposed. In defining the population of interest, we excluded all participants reporting regular alcohol use at time zero. We then defined exposure initiation as the time when participants first became exposed.

Censoring

We denote the censoring indicators as $C_t, t \in (1, \cdots, 20)$. A participant was censored by time t, denoted $C_t = 1$, when they were lost to follow-up or when they skipped more than one entire year of follow-up; otherwise, $C_t = 0$.

Covariates

Baseline covariates As baseline variables, we included socio-demographic characteristics including sex, school indicator, mothers' education, whether the participant lived in a single-parent home, if the participant spoke French at home, and country of birth, which were assessed in the first data collection cycle. In addition, we also included as baseline covariates: self-esteem, impulsivity, and novelty-seeking, measured in the cycle 12, since they are considered as personal traits and unlikely to vary considerably over time. We denoted the baseline variables as $\boldsymbol{W}$.

Time-varying covariates The time-varying covariates $\boldsymbol{L}_t, t \in (1, \cdots, 20)$ were measured between exposures A_t and $A_{(t+1)}$ and included: current depressive symptoms (validated 6-item symptoms scale)(35–37); participation in team sports; family-related stress (validated 4-point scale) with higher values indicating more stress; other type stress (validated 4-point scale); worry about weight; and ever smoked.

Detailed information on all covariates is given in Appendix A.

Outcomes

The outcome Y, depression symptoms, was measured using the Major Depressive Inventory (MDI) in 2007/2008. This scale measures depression symptoms over the past two weeks with range 0-50, where greater values indicate more severe symptoms (36, 38). A detailed list of items included in the MDI is presented in Appendix B.

Data structure

Given the above, the following represents the observed data structure:

$$O = \{\boldsymbol{W}, \boldsymbol{L}_1, A_1, \boldsymbol{L}_2, C_2, A_2 \cdots, \boldsymbol{L}_{19}, C_{19}, A_{19}, \boldsymbol{L}_{20}, C_{20}, Y\}.$$

By design, $A_0 = 0$ and $C_1 = 0$ for all participants.

Definition of target parameter

Target trial

We define two target trials, with corresponding ITT and AT parameters of interest. Both trials recruit participants who had not initiated regular drinking at the beginning of grade 7. The first target trial randomizes drinking initiation to one of the first 19 follow-up time points. The second target trial randomizes drinking (yes/no) at each of the 19 time points during the follow-up. The depression score outcome is measured at the follow-up time in adulthood. The parameters of interest are the coefficients of a linear regression conditional on sex, drinking assignment and their interaction.

Parameter of interest

In the observational study, we characterize counterfactuals under the different types of hypothetical interventions. In the ITT trial, the intervention is initiation time where the analysis ignores changes in subsequent alcohol use. There are 19 possible initiation times denoted $\delta = 1, ..., 19$ or never. In contrast, the AT study randomizes drinking at each time, so that some pattern of drinking, denoted $\bar{\boldsymbol{a}}$, a vector of length 19 of zeros and ones, is assigned to each participant. Therefore, we have 2^{19} potential treatment patterns. We define a specific treatment pattern $\bar{\boldsymbol{a}}^d, \ for \ \bar{\boldsymbol{a}}^d \in \mathcal{D}$ where $\mathcal{D}$ is the set of all possible patterns.

Define $Y(\delta)$ and $Y(\bar{\boldsymbol{a}}^d)$ as the counterfactual outcomes that would have been observed under assigned initiation time δ or treatment pattern $\bar{\boldsymbol{a}}^d = (a_1^d, \cdots, a_{19}^d)$, respectively. The parameters of interest are defined through working MSMs to summarize how the mean counterfactual outcome varies as a function of different interventions, and the baseline covariate sex. Then the true causal quantities can be interpreted as the projection of the true function onto the linear working models.

The two working MSMs are:

$$\mathbb{E}[Y(\delta)|\text{sex}] = m(\text{sex}, \delta; \boldsymbol{\beta}^{\delta}) = \beta_0^{\delta} + \beta_1^{\delta}\text{sex} + \beta_2^{\delta}(20 - \delta) + \beta_3^{\delta}\{\text{sex} \times (20 - \delta)\} \quad (1)$$

$$\mathbb{E}[Y(\bar{\boldsymbol{a}}^d)|\text{sex}] = m(\text{sex}, \bar{\boldsymbol{a}}^d; \boldsymbol{\beta}^d) = \beta_0^d + \beta_1^d\text{sex} + \beta_2^d cum\ (\bar{\boldsymbol{a}}^d) + \beta_3^d\{\text{sex} \times cum(\bar{\boldsymbol{a}}^d)\} \quad (2)$$

where $cum((\bar{\boldsymbol{a}}^d)$ counts the number of exposed time-points in the pattern and $\mathbb{E}[Y(\cdot)|\text{sex}]$ represents the mean counterfactual outcome under some intervention in a sex subgroup such that sex=1 denotes female. The true parameter values of $\boldsymbol{\beta}^{\delta}$ and $\boldsymbol{\beta}^d$ minimize the risk under a squared error loss function, summing over the all patterns (See Appendix C), corresponding to the parameters estimated in the target trials. Equations (1) and (2) thus contrast the counterfactual mean depression score given different alcohol initiation times and usage patterns, respectively.

Assumptions

Several causal assumptions allow us to write the parameters in terms of distributions of the observed data (identifiability). The consistency assumption requires that the counterfactuals under the treatment pattern actually taken are equal to the observed outcomes. Conditional exchangeability (or no unmeasured confounders) for AT [ITT, respectively], requires that at each time point, exposure [initiation] is effectively randomized within each stratum defined by the participant history up to that point, among uncensored [and uninitiated] participants. Non-interference requires that one individual's drinking exposure does not impact others' drinking and others' depression outcomes given their respective exposures. Finally, the positivity assumption requires that all individuals must have a positive probability of initiating (ITT) or continuing to follow any drinking pattern (AT) given their covariate history. The latter implies that all participants must be able to consume alcohol or abstain at all time points. Even if theoretically satisfied, if these probabilities are estimated to be close to zero, this amounts to practical positivity violations (or sparsity) and the estimation relies on extrapolation or smoothing across covariate strata (39). Under these assumptions, we can identify the parameter of interest using Iterated Conditional Expectations (ICE) (1, 8) (See Appendix D).

Methods

In the application of the causal inference methods to obtain point estimates of the parameters of interest, we assumed independence between study participants. However, the variance estimation adjusted for clustering by school (12). Our handling of baseline and time-varying covariate missingness involved multiple imputations by chained equations (See Appendix A).

Sequential G-computation

We use the notation $\bar{\boldsymbol{L}}_t$ to denote the history of time-dependent covariates up to time t and likewise $\bar{A}_t$ represents the history of the exposure $A_1, \cdots, A_t$. Define $\bar{Q}_t(\bar{\boldsymbol{a}}^d)$ as the mean counterfactual outcome at time $t \in (21, \cdots, 1)$, had past exposures been set to $\bar{\boldsymbol{a}}_t^d$, given the covariate history. To apply G-computation, we first rescaled the continuous outcome Y to $(0, 1)$ and defined $\bar{Q}_{21}(\bar{\boldsymbol{a}}^d) = Y$, then sequentially, for each time t, we fit logistic regressions conditioning on exposure and covariate history using uncensored participants (Appendix E). Finally we obtained estimates of $\bar{Q}_1(\bar{\boldsymbol{a}}^{\boldsymbol{d}})$ for all participants by predicting from this fit under imputed exposure history corresponding to the given pattern $\bar{\boldsymbol{a}}^d$. We then stacked the vectors $\bar{Q}_1(\bar{\boldsymbol{a}}^d)$ for each pattern $\bar{\boldsymbol{a}}^d$ and regressed this vector on baseline covariates and summaries of regular drinking exposure using linear regression according to the MSM. Standard errors were then estimated by clustered bootstrap (5, 12). The detailed algorithm of sequential G-computation is given in the Appendix G. To estimate the variance, for each imputed dataset we bootstrapped 100 samples to estimate the standard errors then calculated the confidence intervals using Rubin's rule (40).

LTMLE

In contrast to the sequential G-computation, LTMLE requires models for the treatment and censoring mechanisms to update the initial estimates of each $\bar{Q}_t(\bar{\boldsymbol{a}}^{\boldsymbol{d}})$ with the objective of satisfying the efficient influence function estimating equation (41). This results in double robustness and asymptotic local efficiency (42). First we used logistic regression models to estimate the treatment probabilities stratified by time conditional on the baseline and time varying covariates, and lagged exposure for uncensored participants. The probabilities of censoring were estimated using logistic regression models conditioning on all the covariate and exposure history (Appendix E). Define w_t^d

Table 1: Pooled LTMLE Algorithm

Pooled LTMLE Algorithm

1. Estimate every component of w_t^d for $t = 1, \cdots, 20$ to obtain the estimated weights $w_{t,n}^d$ for each treatment pattern $\bar{\boldsymbol{a}}^d$.
2. Define $\bar{Q}_{21,n}^{d*} = Y$, where Y is rescaled to $(0, 1)$.
3. For $t = 20, \cdots, 1$
   - 3.1 Initial estimate of $\bar{Q}_{t,n}^d$: Fit a logistic regression of $\bar{Q}_{t+1,n}^{d*}$ on uncensored participants given all the treatment and covariate history. When $t = 20$, predict outcomes by setting $\bar{\boldsymbol{A}}_{19} = \bar{\boldsymbol{a}}_{19}^d$ otherwise setting $\bar{\boldsymbol{A}}_t = \bar{\boldsymbol{a}}_t^d$ for each treatment pattern and each subject. Define $\bar{Q}_{t,n}^d$ as the stacked vector of predictions with length $n \times |\mathcal{D}|$.
   - 3.2 Construct a covariate matrix for each subject and each treatment pattern $\bar{\boldsymbol{a}}^d$, $\boldsymbol{h}_t(\bar{\boldsymbol{a}}^d, \text{sex}) = \mathbb{1}(\bar{\boldsymbol{A}}_t = \bar{\boldsymbol{a}}_t^d, C_t = 0) \times [\partial m(\boldsymbol{\beta}, \text{sex}, \bar{\boldsymbol{a}}^d)/\partial \boldsymbol{\beta}]$ where $\partial m(\boldsymbol{\beta}, \text{sex}, \bar{\boldsymbol{a}}^d)/\partial \boldsymbol{\beta}$ in our example equals $1, \text{sex}, cum(\bar{\boldsymbol{a}}_t^d), \text{sex} \times cum(\bar{\boldsymbol{a}}_t^d)$. For a treatment pattern $\bar{\boldsymbol{a}}^d$, the dimension of $\boldsymbol{h}_t(\bar{\boldsymbol{a}}^d, \text{sex})$ is the same as the dimension of $\boldsymbol{\beta}$. Thus, for all possible patterns, the $\boldsymbol{h}_t(\bar{\boldsymbol{a}}^d, \text{sex})$ is of dimension $(n \times |\mathcal{D}|) \times 4$.
   - 3.3 Update $\bar{Q}_{t,n}^d$ to $\bar{Q}_{t,n}^{d*}$ by fitting a intercept-free weighted pooled logistic regression of $\bar{Q}_{t+1,n}^{d*}$ on the covariate matrix produced from the previous step with offset $\text{logit}(\bar{Q}_{t,n}^d)$ and $w_{t,n}^d$ as weights for each $\bar{\boldsymbol{a}}^d$, $\text{logit}(\bar{Q}_{t,n}^{d*}) = \text{logit}(\bar{Q}_{t,n}^d) + \boldsymbol{\epsilon}\boldsymbol{h}_t(\bar{\boldsymbol{a}}^d, \text{sex})$.
   - 3.4 Generate $\bar{Q}_{t,n}^{d*}$ by making predictions for every subject under each regime $\bar{\boldsymbol{a}}^d$ using the logistic model fit in step 3.3 $\bar{Q}_{t,n}^{d*}$ is the stacked vector of predictions and has length $n \times |\mathcal{D}|$.
4. Rescale $\bar{Q}_{1,n}^{d*}$ (length $n \times |\mathcal{D}|$) to the original scaling of Y.
5. The coefficients are estimated by fitting a pooled linear regression of $\bar{Q}_{1,n}^{d*}$ on stacked covariates and all treatment patterns $\mathcal{D}$ corresponding to the MSM model in Equation 2.

as the cumulative weight which is the cumulative product of the inverse of treatment and censoring probabilities from time one to t under the specific treatment pattern $\bar{\boldsymbol{a}}^d$. Here, we used stabilized weights which results in a weaker positivity assumption (39) (see Appendix F). Because we observed practical positivity violations, we also truncated the weights at fixed values (1,000, 5,000, and 10,000) to improve the performance of the estimators.

LTMLE allows for the integration of machine learning to increase the chances of consistent estimation under regularity conditions (10). Super Learning (SL) is a methodology that uses V-fold cross validation to find an optimal convex combination of the predictions of a library of candidate algorithms defined by the user (43). We therefore used SL to estimate the $\bar{Q}_t(\bar{\boldsymbol{a}}^d)$s, and the exposure and censoring probabilities. In the AT analysis, we included the following methods in the SL library: generalized additive models (`gam` function), generalized linear models (GLM) (`glm` function), and LASSO (`glmnet` function). We customized these functions by adding interaction terms between treatment and sex in the SL wrappers. We present the pooled LTMLE algorithm to estimate the parameters of an MSM (11) in Table 1. The subscript n is used to denote an estimate of a quantity.

Since we hand-coded the pooled LTMLE algorithm for the clustered setting (12), we verified its correctness using two simulations with two time points and clustered observations and estimated ITT and AT parameters, described in Appendix H. We simulated 500 datasets where we generated $5,000$ participants in 50 clusters, with random intercepts in the outcome model. We verified the unbiasedness of the LTMLE and also compared standard error estimators assuming independence and clustering (respectively) using the influence function based sandwich estimator and clustered bootstrap, respectively, showing that the clustered versions are needed under random effects. The full data-generation and results are given in Table S2 and S3 in Appendix H.

Challenges and strategies due to high-dimensionality and sparsity in the AT analysis

The main challenges in the AT analysis involved computational issues introduced by the very large number of potential treatment patterns. Recall that we have $|\mathcal{D}| = 2^{19} = 524,288$ potential treatment patterns which would thereby produce several very large stacked vectors and matrices when we perform the pooled TMLE procedure. The vectors are of length $|\mathcal{D}| \times n = 2^{19} \times 1,231 = 645,398,528$ for each time t in the update step (steps (3.2-3.4) of the pooled LTMLE algorithm in Section LTMLE). However, objects of this size cannot be stored in the `R` memory or easily manipulated using `R`.

The parameters of interest are defined in terms of the minimization of a risk calculated over all 2^{19} patterns (see Appendix D). However, in order to tackle this issue, we proposed a pragmatic strategy that redefines the parameters of interest by minimizing risk over the patterns that are most supported by the data. Let $\mathcal{D}_t$ be the set of patterns that were supported by data (i.e. observed to be followed by at least one individual up to time t) for each time $t = 1, \cdots, 19$. Table S4 in Appendix I gives the cardinality (size) of $\mathcal{D}_t$ at each time point. We firstly performed an analysis involving all supported patterns up to time 19, such that $|\mathcal{D}_{19}| = 227$ patterns were included. In order to test the sensitivity of the results to the number of time-points used, we ran analyses using the patterns in $\mathcal{D}_{18}$ (494 patterns; note that these include the patterns in $\mathcal{D}_{19}$); then in $\mathcal{D}_{17}$ (936 patterns) ; and finally in $\mathcal{D}_{16}$ (1,688 patterns). On our local computer, we could not realistically go further than $t = 16$ for the LTMLE analysis with GLMs. We focused on $\mathcal{D}_{17}$ (936 patterns) since this was the largest number of patterns that could be incorporated in the LTMLE analysis with SL.

Longitudinal modified treatment policies (LMTP)

Inspired by the hypothetical interventions based on the natural value of treatment first discussed by Robins et al. (44) then formalized by other researchers (45, 46), Díaz et al. (19) proposed longitudinal modified treatment policies (LMTPs). LMTP involves a hypothetical intervention at each time point which can be expressed as a deterministic or random function of the observed treatment and the unit's covariate history. In this paper, we apply an incremental propensity score intervention based on risk ratio scale (47, 48) that shifts the propensity scores to discourage alcohol use in adolescents. Specifically, the intervention assigns a new exposure whose likelihood can be determined hypothetically by the user-defined risk ratio value. Under this intervention effects can be identified and estimated under weak conditions on the propensity score (47); consequentially, sparsity less destabilizes the analysis. We used the `lmtp` R package (available at (49) and CRAN) which implements LTMLE to estimate effects under specified LMTPs. We choose five risk ratios of alcohol use in $[0.1, 0.3, 0.5, 0.7, 0.9]$ then estimated the mean MDI stratified by sex and compared them with the mean MDI without any intervention. We applied the same SuperLearner algorithms as Section LTMLE. Further details are given in Appendix J.

Results

The dataset included 1,294 participants. We excluded participants who only completed the first follow-up, skipped the first year of the study, or reported alcohol consumption at baseline, leaving 1,231 participants in the analysis (Figure S2 in Appendix I). Summaries of baseline and time-varying characteristics of the 1,231 participants are presented in Table 2. There was missingness in the baseline covariates. Table 3 reports the cumulative numbers and percentages of censoring, initiators and actual exposed participants at each time. Note that at $t = 1$, there was no censoring due to the exclusion criteria.

ITT analysis

The range of the cumulative stabilized weights lay within $[0.22, 72.38]$ with mean 1.02, so no truncations were applied. In Figure 1 and Table S5 in Appendix I, we see that all estimates agreed that female sex was associated with more severe depressive symptoms. LTMLE suggested a stronger sex

Table 2: Baseline characteristics and time-varying covariates at time $t = 1$ of the 1,231 participants in the analytic sample from the NDIT study.

Covariates	Variables	Class	Median	IQR	N.	%	Missing
Baseline	Sex	Female/Male			644/587	52.3/47.7	
	Single-parent	Yes/No			148/1,083	12/88	
	Speaking French	Yes/No			368/863	29.9/70.1	
	Country of birth	In/Outside Canada			1,132/99	92/8	
	Mother education	Less/Some university			523/419	55.5/44.5	289
	Self esteem[a]	Numeric	2.7	(2.2; 2,9)			276
	Impulsivity[a]	Numeric	2.1	(1.6; 2.9)			326
	Novelty-seeking[a]	Numeric	2.9	(2.3; 3.4)			324
$\boldsymbol{L}_1$	Weight worry	Yes/No			427/736	36.7/63.3	68
	Participate sports	Yes/No			750/437	63.2/36.8	44
	Ever smoked	Yes/No			365/862	29.7/70.3	4
	Current dep. sym.	Numeric	2.0	(1.7; 2.5)			55
	Family stress	Numeric	1.2	(1.0; 1.4)			59
	Other stress	Numeric	1.4	(1.2; 1.6)			52

[a] Note that these three covariates, considered time-invariant, were measured at cycle 12.

Table 3: Size, cumulative size and cumulative percentage of censoring, size of cumulative initiators, exposed participants, and corresponding percentages at 20 follow-up time points

Time	Censoring			Alcohol use			
	N.	Cum. N	Cum. %	Cum. Init.	Cum. Init. %	N. Expo.	Expo. %
0				0	0	0	0
1	0	0	0.0	30	2.4	30	2.4
2	9	9	0.7	56	4.5	40	3.2
3	20	29	2.4	68	5.5	43	3.5
4	85	114	9.3	101	8.2	60	4.9
5	7	121	9.8	151	12.3	100	8.1
6	5	126	10.2	188	15.3	111	9.0
7	10	136	11.0	210	17.1	103	8.4
8	81	217	17.6	233	18.9	116	9.4
9	10	227	18.4	257	20.9	113	9.2
10	2	229	18.6	276	22.4	113	9.2
11	7	236	19.2	284	23.1	104	8.4
12	87	323	26.2	289	23.5	119	9.7
13	8	331	26.9	317	25.8	136	11.0
14	2	333	27.1	340	27.6	124	10.1
15	3	336	27.3	358	29.1	134	10.9
16	39	375	30.5	377	30.6	170	13.8
17	2	377	30.6	394	32.0	171	13.9
18	5	382	31.0	411	33.4	167	13.6
19	7	389	31.6	425	34.5	184	14.9
20	169	558	45.3	425	34.5		

association than G-computation. In addition, LTMLE using GLM and SL indicated that earlier alcohol initiation time was associated with more severe depression symptoms in males but lower depression severity in female adults. G-computation estimated no association between alcohol initiation and depression.

AT analysis

We performed the AT analysis with 227, 494, 936 and 1,688 patterns, respectively, where the models in the pooled LTMLE procedure were fitted using GLM. We set fixed bounds at 1,000, 5,000 and 10,000 on the cumulative stabilized weights. Table 4 summarizes the median values and interquartile ranges of the cumulative stabilized weights and the percentages of truncated cumulative weights. There was no difference in the percentage of weights truncated over the different numbers of patterns. The bounding truncated $7-12\%$ of participants.

Figure 2 and Table S6 in Appendix I show the estimated coefficients using LTMLE with GLM at three levels of truncation. The estimated counterfactual mean of MDI in females was around 5 points higher than in males and was stable over different numbers of patterns. The cumulative duration of drinking over time was also associated with increased depressive levels but this estimate waned with a greater number of patterns and less restrictive bounds. The point estimates of the interaction term were stable with different numbers of treatment patterns but increased with less restrictive weights; the confidence interval widths also increased with less.

To better understand the impact of the weights, we compared the LTMLE results with GLM and SL, respectively, with the parametric sequential G-computation estimator (Figure 3 and Table S7 in Appendix I) using 936 patterns. LTMLE with GLM and SL produced similar results, though LTMLE with SL sometimes produced broader confidence intervals under the less restrictive bounds. G-computation produced much more narrow confidence intervals that contained the null for both the main term effect of cumulative duration and the interaction term. The weights had an important impact on the point-estimate for the coefficient of cumulative exposure, with LTMLE using GLM at three bounds and LTMLE using SL under the 1,000 bound suggesting an association. No method concluded that sex modified the effect of cumulative exposure.

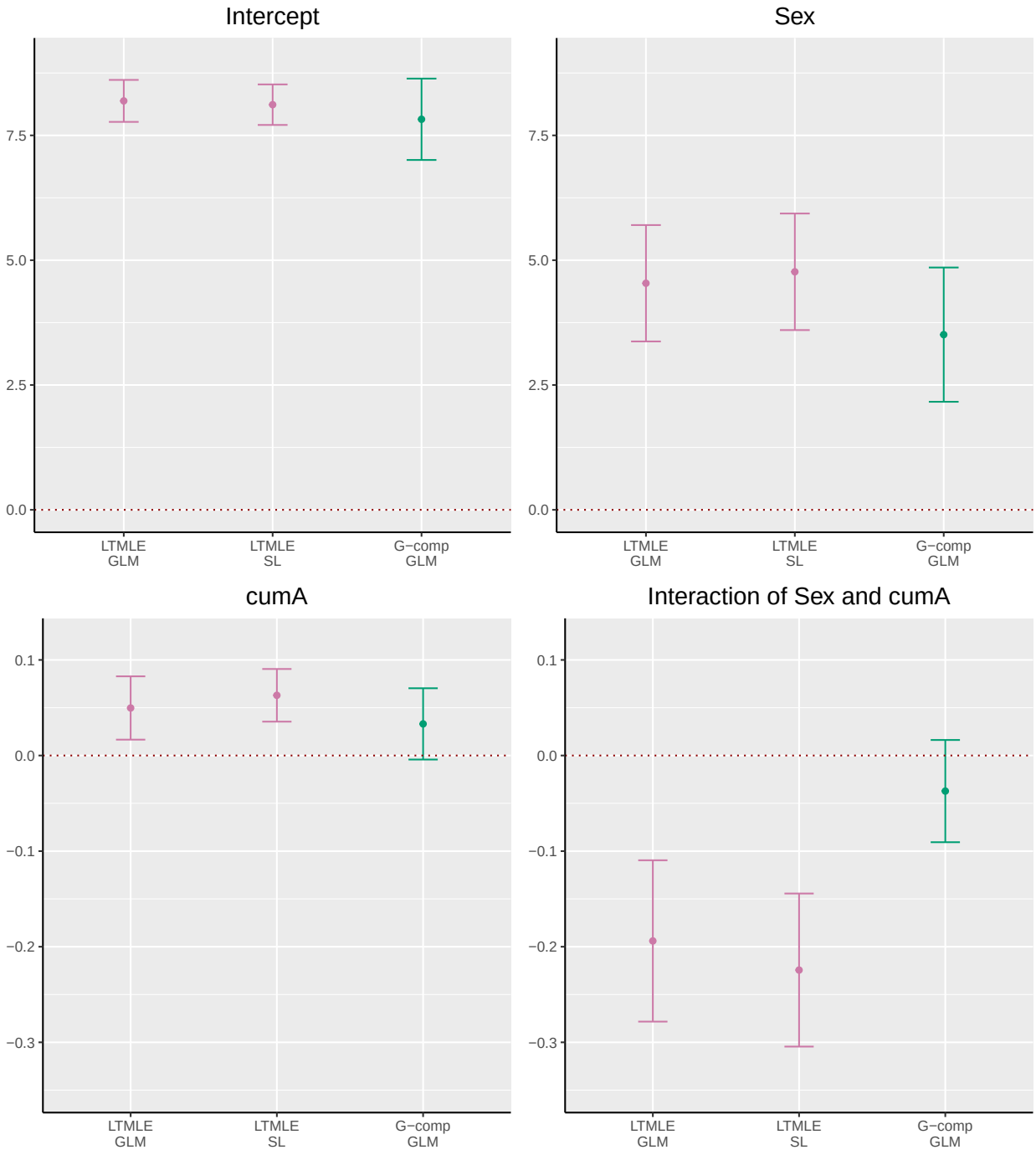

Figure 1: Estimated coefficients of G-computation and LTMLE with GLM and SL, and 95% confidence intervals in ITT analysis

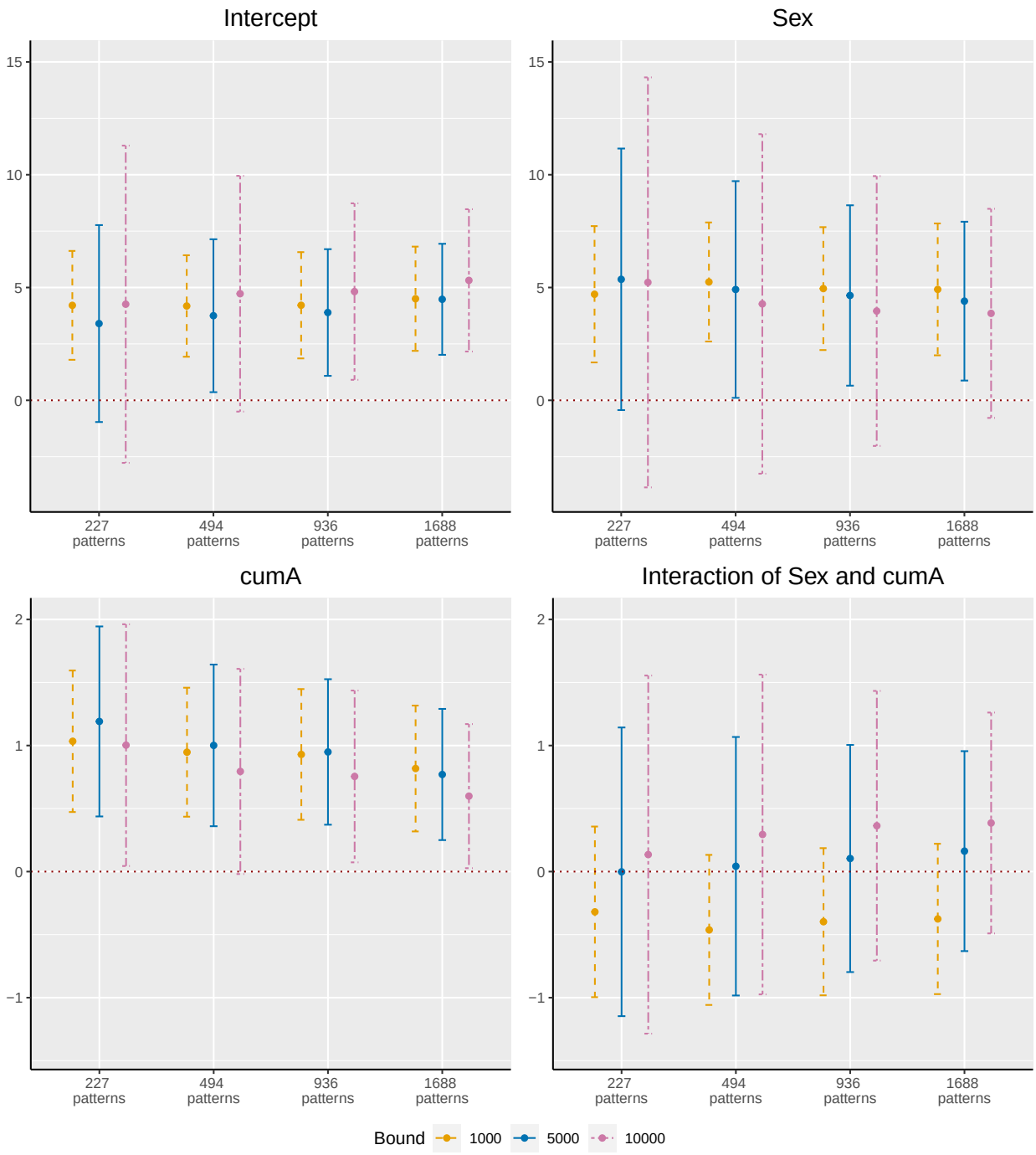

Figure 2: Estimated coefficients of LTMLE with GLM and 95% confidence intervals in AT analysis including 227, 494, 936 and 1,688 patterns under $1,000, 5,000$, and $10,000$ bounds of cumulative stabilized weights

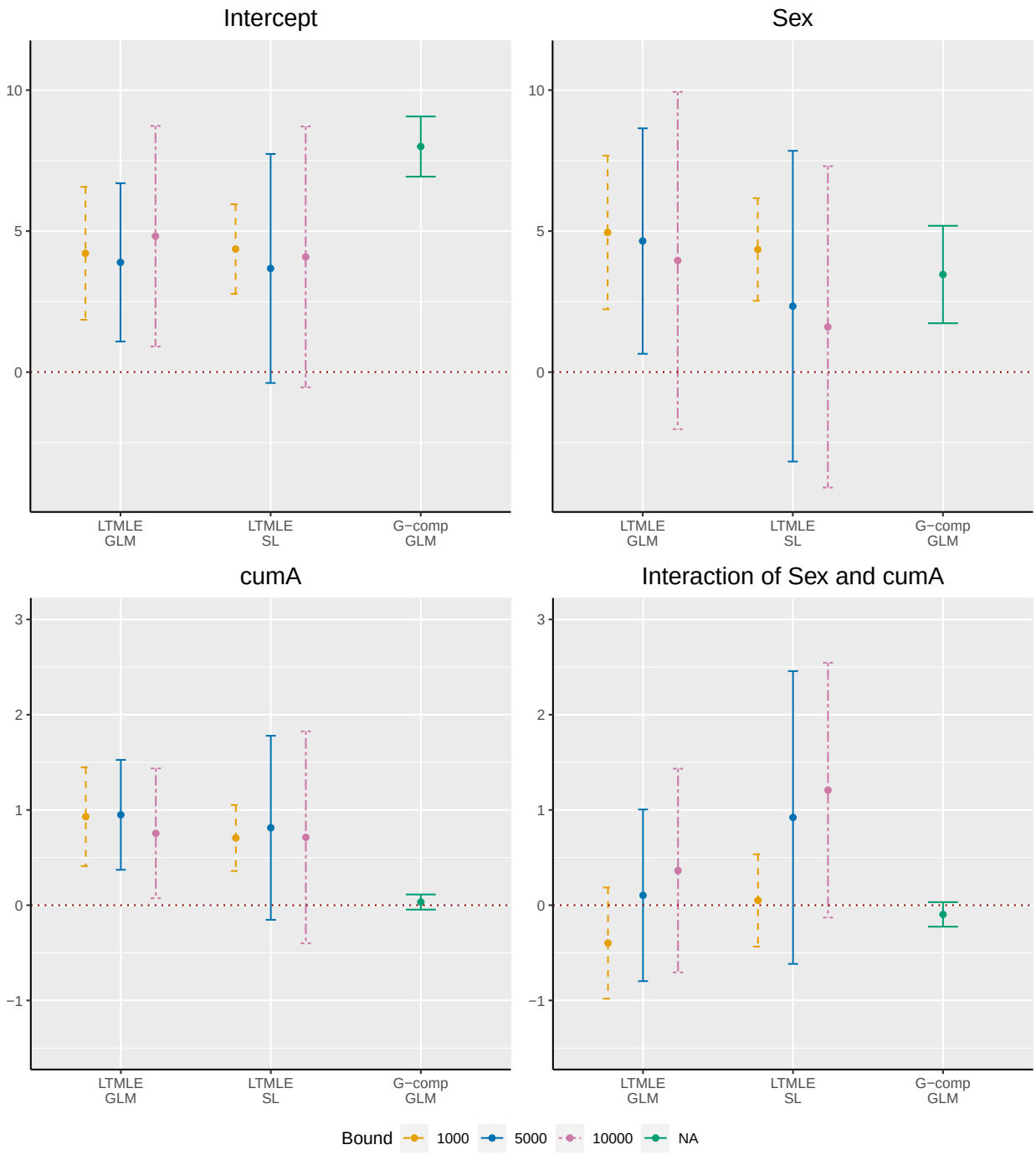

Figure 3: Estimated coefficients of LTMLE with GLM and SL and 95% confidence intervals in the AT analysis including 936 patterns under $1,000, 5,000$, and $10,000$ bounds of cumulative stabilized weights

Table 4: The median, IQR, and percentage of truncated cumulative weights with GLM for 227, 494, 936 and 1,688 patterns.

N. patterns	Median	IQR	% truncated		
			1,000 bound	5,000 bound	10,000 bound
227	1.45	(0.91; 18.50)	0.11	0.08	0.07
494	1.70	(0.93; 23.64)	0.11	0.08	0.07
936	1.87	(0.94; 27.21)	0.11	0.08	0.07
1,688	2.09	(0.95; 32.94)	0.12	0.08	0.07

LMTP estimates

Using LTMLE, we estimated mean MDI and the 95% confidence intervals in males and females separately under five policies of LMTP discouraging alcohol use at each time-point and no intervention in the Appendix J. For both males and females, comparisons between each intervention that would have discouraged alcohol use versus no intervention did not suggest any difference.

Discussion

Our study demonstrates how to apply target trials and modified treatment policies to define causal effects in a challenging longitudinal problem, using LTMLE for estimation. Our analysis involved detailed information on alcohol initiation and use in adolescents and depression in adulthood, with 21 follow-up time points and many baseline and time-dependent confounders. Censoring led to gradually decreasing sample size over time. Analytical challenges in the AT analysis included highly variable weights induced by data sparsity (17) and high-dimensional potential exposure patterns. To tackle these challenges, we used two approaches to modify the target parameter: an ad hoc approach to remove patterns with less data-support from the loss function; and through defining longitudinal interventions shifting the propensity scores to discourage drinking in adolescents.

Our LTMLE with GLM and SL analysis suggested that earlier drinking initiation time was associated with increased depression in males and reduced symptoms in females. LTMLE also indicated that cumulative drinking duration was associated with increased depression in males and females. However, this last analysis was hampered by sparsity and was sometimes sensitive to the weight bounds, though less sensitive to the number of patterns included in the loss function. For better insight, we employed sequential G-computation which uses the same estimation procedure as LTMLE

without the weighting component and again noted the large difference in the point estimates and standard errors. Because LTMLE is doubly robust, influential weights suggest misspecified models for the outcome. But it is not clear to what point the instability of the weights inserted bias into the analysis. Under the LMTP intervention discouraging alcohol use to various degrees, no effect was apparent. Similar LMTP parameters have been proposed elsewhere (18).

Gender and sex differences in the relationship between alcohol use and depression have been investigated in recent years. Fergusson et al. (50) found no gender differences of the association between alcohol use and depressive symptoms using a 25-year longitudinal study. Edwards et al. (51) observed the association between harmful alcohol use and depression in females, but not in males based on a prospective population based cohort data. Inconsistencies between these results and our own may be explained to some extent by different populations under study that vary in age composition, the social norms, the degree of depressive symptoms, the estimation methods, and the confounding adjustment involved.

Limitations in our analysis include the interference assumption likely being violated to some extent (e.g. within school classes) since drinking behavior is transmissible in adolescents as perceived peer norms have a direct effect on alcohol use (52). Second, though we adjusted for many relevant confounders, the no unmeasured confounders assumption is likely unmet as we do not have a complete profile on personal circumstances that would affect the timing of drinking in adolescence and depression in adulthood. Third, due to the limited computational ability of our local computers, we included at most 1,688 treatment patterns in the AT analysis which changed the parameter of interest, potentially leading to bias. However, we did not see much deviation in the statistical conclusions when we varied the set of treatment patterns, suggesting that the bias may not be substantial.

Our study contributes to a growing literature on the application of robust longitudinal causal inference methods. While these methods have many important theoretical properties, data sparsity is a common challenge. We thus encourage epidemiologists and applied statisticians to explore recently proposed parameter definitions and estimation methods that reduce positivity constraints, leading to more robust results.

Acknowledgments

For cycles 1-21, this work was supported by the Canadian Cancer Research Society (Grant 010271 and 017435). MES is supported by a Canada Research Chair Tier 2. MPS is supported by a Chercheur-Boursier J2 career award from the Fonds de Recherche du Quebec-Santé (FRQS).

References

1. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. 2000.
2. Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000:561–570.
3. Robins JM. Marginal structural models. *1997 Proceedings of the American Statistics Association, Section on Bayesian Statistical Science.* 1997:1–10.
4. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology.* 2016;183(8):758–764.
5. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *American Journal of Epidemiology.* 2011;173(7):731–738.
6. Vansteelandt S, Keiding N. Invited commentary: G-computation–lost in translation?. *American Journal of Epidemiology.* 2011;173(7):739–742.
7. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association.* 1999;94(448):1096-1120.
8. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics.* 2005;61(4):962-973.
9. Van der Laan MJ, Rose S. *Targeted learning in data science: causal inference for complex longitudinal studies.* Springer 2018.
10. Van der Laan MJ, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The International Journal of Biostatistics.* 2012;8(1).
11. Petersen M, Schwab J, Gruber S, et al. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *Journal of causal inference.* 2014;2(2):147-185.
12. Schnitzer ME, Van der Laan MJ, Moodie EE, et al. Effect of breastfeeding on gastrointestinal infection in infants: a targeted maximum likelihood approach for clustered longitudinal data. *The Annals of Applied Statistics.* 2014;8(2):703.
13. Kreif N, Tran L, Grieve R, et al. Estimating the comparative effectiveness of feeding interventions in the pediatric intensive care unit: a demonstration of longitudinal targeted maximum likelihood estimation. *American Journal of Epidemiology.* 2017;186(12):1370–1379.
14. Brooks JC, Van der Laan MJ, Singer DE, et al. Targeted minimum loss-based estimation of causal effects in right-censored survival data with time-dependent covariates: Warfarin, stroke, and death in atrial fibrillation. *Journal of Causal Inference.* 2013;1(2):235–254.
15. Tran L, Yiannoutsos CT, Musick BS, et al. Evaluating the impact of a HIV low-risk express care task-shifting program: a case study of the targeted learning roadmap. *Epidemiologic Methods.* 2016;5(1):69–91.

16. Tran L, Yiannoutsos C, Wools-Kaloustian K, et al. Double robust efficient estimators of longitudinal treatment effects: comparative performance in simulations and a case study. *The International Journal of Biostatistics.* 2019;15(2).

17. Schomaker M, Luque-Fernandez MA, Leroy V, et al. Using longitudinal targeted maximum likelihood estimation in complex settings with dynamic interventions. *Statistics in Medicine.* 2019;38(24):4888–4911.

18. Kennedy EH. Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association.* 2019;114(526):645–656.

19. Díaz I, Williams N, Hoffman KL, et al. Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association.* 2021:1–16.

20. O'Loughlin J, Karp I, Koulis T, et al. Determinants of first puff and daily cigarette smoking in adolescents. *American Journal of Epidemiology.* 2009;170(5):585–597.

21. O'Loughlin J, Dugas EN, Brunet J, et al. Cohort profile: the nicotine dependence in teens (NDIT) study. *International Journal of Epidemiology.* 2015;44(5):1537–1546.

22. Grant BF, Dawson DA. Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: results from the National Longitudinal Alcohol Epidemiologic Survey. *Journal of Substance Abuse.* 1997;9:103–110.

23. Torres JM, Rudolph KE, Sofrygin O, et al. Longitudinal associations between having an adult child migrant and depressive symptoms among older adults in the Mexican Health and Aging Study. *International Journal of Epidemiology.* 2018;47(5):1432–1442.

24. Mason WA, Kosterman R, Haggerty KP, et al. Dimensions of adolescent alcohol involvement as predictors of young-adult major depression. *Journal of Studies on Alcohol and Drugs.* 2008;69(2):275–285.

25. Rohde P, Lewinsohn PM, Kahler CW, et al. Natural course of alcohol use disorders from adolescence to young adulthood. *Journal of the American Academy of Child & Adolescent Psychiatry.* 2001;40(1):83–90.

26. Marmorstein NR. Longitudinal associations between alcohol problems and depressive symptoms: early adolescence through early adulthood. *Alcoholism: Clinical and Experimental Research.* 2009;33(1):49–59.

27. Boden J, Blair S, Newton-Howes G. Alcohol use in adolescents and adult psychopathology and social outcomes: Findings from a 35-year cohort study. *Australian & New Zealand Journal of Psychiatry.* 2020;54(9):909–918.

28. Edwards AC, Heron J, Dick DM, et al. Adolescent alcohol use is positively associated with later depression in a population-based UK cohort. *Journal of Studies on Alcohol and Drugs.* 2014;75(5):758–765.

29. Tercyak KP, Rodriguez D, Audrain-McGovern J. High school seniors' smoking initiation and progression 1 year after graduation. *American Journal of Public Health.* 2007;97(8):1397–1398.

30. Jackson KM, Sher KJ, Cooper ML, et al. Adolescent alcohol and tobacco use: onset, persistence and trajectories of use across two samples. *Addiction.* 2002;97(5):517–531.

31. Goldbach JT, Berger Cardoso J, Cervantes RC, et al. The relation between stress and alcohol use among Hispanic adolescents.. *Psychology of Addictive Behaviors.* 2015;29(4):960.

32. Meca A, Zamboanga BL, Lui PP, et al. Alcohol initiation among recently immigrated Hispanic adolescents: Roles of acculturation and sociocultural stress.. *American Journal of Orthopsychiatry.* 2019;89(5):569.

33. Lorente FO, Souville M, Griffet J, et al. Participation in sports and alcohol consumption among French adolescents. *Addictive Behaviors.* 2004;29(5):941–946.

34. Mays D, DePadilla L, Thompson NJ, et al. Sports participation and problem alcohol use: A multi-wave national sample of adolescents. *American Journal of Preventive Medicine.* 2010;38(5):491–498.

35. Brunet J, Sabiston CM, Chaiton M, et al. Measurement invariance of the depressive symptoms scale during adolescence. *BMC Psychiatry.* 2014;14(1):1–10.

36. Chaiton M, Contreras G, Brunet J, et al. Heterogeneity of depressive symptom trajectories through adolescence: Predicting outcomes in young adulthood. *Journal of the Canadian Academy of Child and Adolescent Psychiatry.* 2013;22(2):96.

37. Escobedo LG, Kirch DG, Anda RF. Depression and smoking initiation among US Latinos. *Addiction.* 1996;91(1):113–119.

38. Bech P, Timmerby N, Martiny K, et al. Psychometric evaluation of the Major Depression Inventory (MDI) as depression severity scale using the LEAD (Longitudinal Expert Assessment of All Data) as index of validity. *BMC Psychiatry.* 2015;15(1):1–7.

39. Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research.* 2012;21(1):31–54.

40. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Statistics in Medicine.* 2018;37(14):2252–2266.

41. Van der Laan MJ, Rubin D. Targeted maximum likelihood learning. *The International Journal of Biostatistics.* 2006;2(1).

42. Van der Laan MJ, Laan M, Robins JM. *Unified methods for censored longitudinal data and causality.* Springer Science & Business Media 2003.

43. Van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Statistical Applications in Genetics and Molecular Biology.* 2007;6(1).

44. Robins JM, Hernán MA, Siebert U. Effects of multiple interventions. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors.* 2004;1:2191–2230.

45. Richardson TS, Robins JM. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper.* 2013;128(30):2013.

46. Young JG, Hernán MA, Robins JM. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic Methods.* 2014;3(1):1–19.

47. Díaz I, Hoffman KL, Hejazi NS. Causal survival analysis under competing risks using longitudinal modified treatment policies. *arXiv preprint arXiv:2202.03513.* 2022.

48. Wen L, Marcus J, Young J. Intervention treatment distributions that depend on the observed treatment process and model double robustness in causal survival analysis. *arXiv preprint arXiv:2112.00807.* 2021.

49. Williams N, Díaz I. lmtp: Non-parametric Causal Effects of Feasible Interventions Based on Modified Treatment Policies. https://github.com/nt-williams/lmtp 2021. R package version 1.0.0.

50. Fergusson DM, Boden JM, Horwood LJ. Tests of causal links between alcohol abuse or dependence and major depression. *Archives of General Psychiatry.* 2009;66(3):260–266.

51. Edwards AC, Joinson C, Dick DM, et al. The association between depressive symptoms from early to late adolescence and later use and harmful use of alcohol. *European Child & Adolescent Psychiatry.* 2014;23(12):1219–1230.

52. Scheier LM, Botvin GJ. Expectancies as mediators of the effects of social influences and alcohol knowledge on adolescent alcohol use: A prospective analysis.. *Psychology of Addictive Behaviors.* 1997;11(1):48.