Multiply robust estimation of the causal effect of treatment in data subject to irregular observation times and confounding

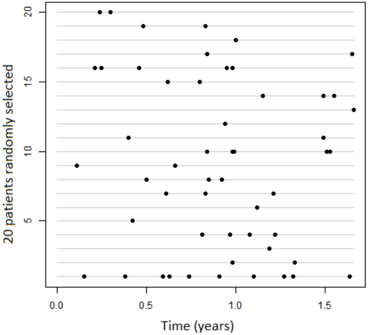
Janie Coulombe[1] and Shu Yang[2]

1. Department of Mathematics and Statistics, Université de Montréal, CA
2. Department of Statistics, North Carolina State University, US

Joint Statistical Meetings, August 2023, Toronto, Canada

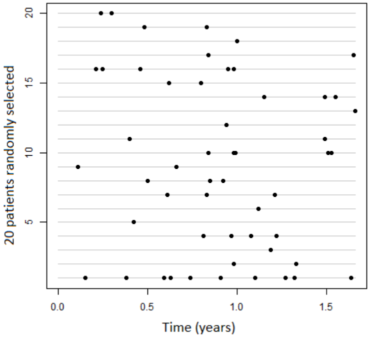
Setting

Suppose **repeated irregular measurements** of forced expiratory volumes (FEV), denoted $Y(t)$ at time t, from a cohort of size N over τ year(s) of follow-up.



Setting

Suppose **repeated irregular measurements** of forced expiratory volumes (FEV), denoted $Y(t)$ at time t, from a cohort of size N over τ year(s) of follow-up.



Question: What is the average causal effect of air pollution (variable $A(t) \in \{0, 1\}$) on the FEV?

Setting (2)

Denote by $Y^1(t)$ and $Y^0(t)$ the two potential outcomes corresponding to the different levels of air pollution. We assume an acute effect.

Setting (2)

Denote by $Y^1(t)$ and $Y^0(t)$ the two potential outcomes corresponding to the different levels of air pollution. We assume an acute effect.

Under the following marginal structural model

$$E[Y^a(t)] = \beta_0 + \beta_1 a,$$

the average treatment effect (ATE) is given by

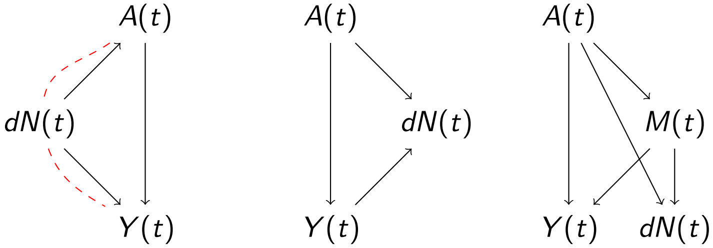
$$E[Y^1(t) - Y^0(t)] = \beta_1.$$

Why are irregular measurements problematic?

Denote a counting process for the outcome measurements by $N(t)$ and $dN(t)$ and indicator of jump in the process at t.

Why are irregular measurements problematic?

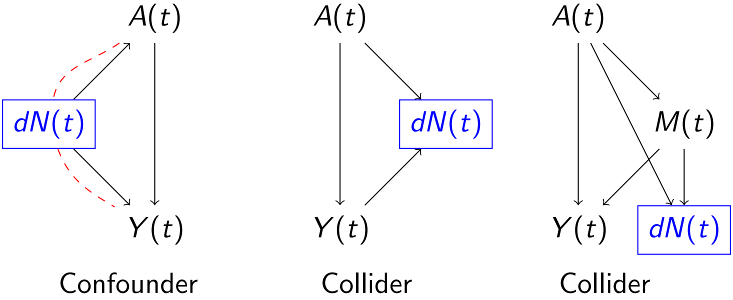
Denote a counting process for the outcome measurements by $N(t)$ and $dN(t)$ and indicator of jump in the process at t.



Confounder Collider Collider

Why are irregular measurements problematic?

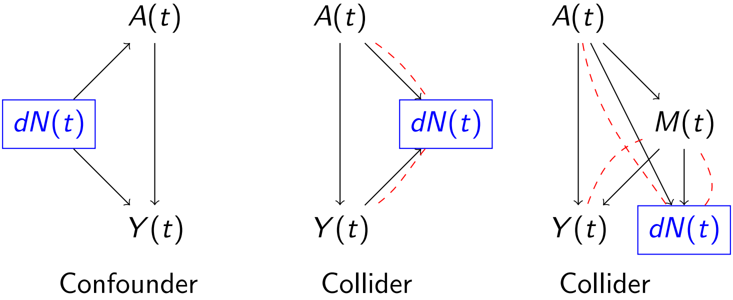
Denote a counting process for the outcome measurements by $N(t)$ and $dN(t)$ and indicator of jump in the process at t.



There is almost always a condition on the measurement indicator $dN(t)$ (i.e., a box around it).

Why are irregular measurements problematic?

Denote a counting process for the outcome measurements by $N(t)$ and $dN(t)$ and indicator of jump in the process at t.



There is almost always a condition on the measurement indicator $dN(t)$ (i.e., a box around it).

Double missingness mechanism

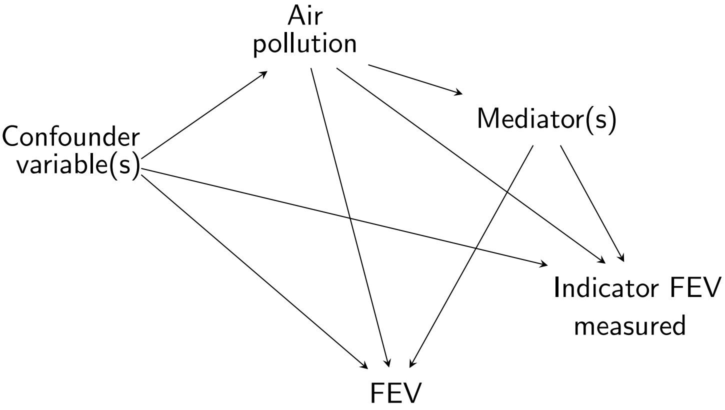
- We observe one of the two potential outcomes informatively, according to the treatment mechanism.
- We do not observe $Y^1(t)$ and $Y^0(t)$ at all times t but only at covariate-dependent measurement times.

Double missingness mechanism

- We observe one of the two potential outcomes informatively, according to the treatment mechanism.
- We do not observe $Y^1(t)$ and $Y^0(t)$ at all times t but only at covariate-dependent measurement times.
- Informative sampling and inverse weighting methods were discussed in the survey sampling (e.g., Horvitz and Thompson 1952) and causal inference (e.g., Robins, Rotnitzky, and Zhao 1994) literatures.
- In Coulombe et al. (2021), we proposed a doubly-weighted (DW) estimator for that setting.

Causal diagram in Coulombe et al. (2021)

We (again) assume the following causal diagram at time t:



We further suppose that air pollution $A(t)$, confounders of the pollution-FEV relationship, $\mathbf{K}(t)$, and predictors of the measurement of FEV, $\mathbf{V}(t)$ are updated and available daily.

Causal assumptions

- Outcome consistency: $Y(t) = A(t)Y^1(t) + \{1 - A(t)\} Y^0(t)$
- Positivity of treatment: $0 < P\{A(t) \mid \mathbf{K}(t)\} < 1$
- Positivity of observation: $0 < E[dN(t) \mid \mathbf{V}(t)] < 1$
- No unmeasured confounder (NUC): $\{Y^0(t), Y^1(t)\} \perp A(t) \mid \mathbf{K}(t)$
- Conditional exchangeability:
  $$\text{NUC} + \{A(t), Y^0(t), Y^1(t)\} \perp dN(t) \mid \mathbf{V}(t)$$

For consistency of the novel estimator only, we also require $\mathbf{K}(t) \subset \mathbf{V}(t)$.

Previous estimator

The DW estimator solved the following equations for β_0 and β_1 (the ATE):

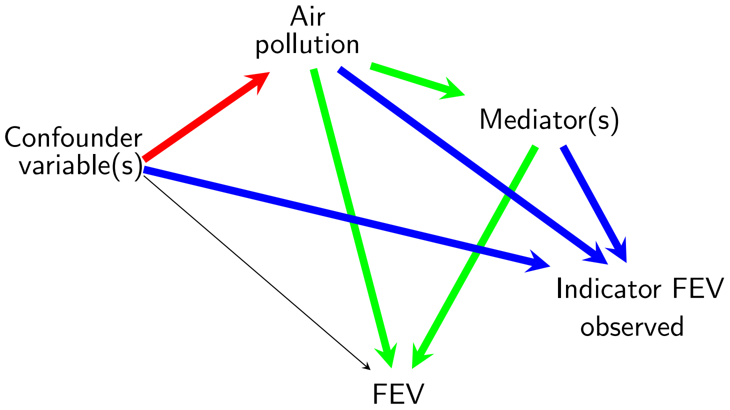
$$E\left[\int_0^{\tau} \frac{\frac{\mathbf{1}\{A(t)=a\}}{P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}} Y(t) - \beta_0 - \beta_1 a}{E[dN(t) \mid \mathbf{V}(t);\hat{\gamma}]} dN(t)\right] = 0,$$

where $\frac{\mathbf{1}\{A(t)=a\}}{P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}}$ are inverse probability of treatment (IPT)[1] weights and $\frac{dN(t)}{E[dN(t)|\mathbf{V}(t);\hat{\gamma}]}$ are inverse intensity of visit (IIV)[2] weights.

[1] Horvitz and Thompson, 1952; Rosenbaum and Rubin, 1983; Robins et al., 2000

[2] Lin, Scharfstein and Rosenheck, 2004

The DW tackled two nuisance models: the **treatment** and the **outcome measurement** models:



Blocking these associations leads to consistent estimation of the **average treatment effect (ATE).**

Two weaknesses of the DW

- Consistency relies on the **correct specification** of IPT and IIV weights
- Relatively **variable** estimator

Semiparametric efficiency

For a general estimand $E[AY]$ estimated with outcome data selected at random given A and $\mathbf{K}$, and π the conditional selection probability, the class of estimators

$$N\hat{T}_{diff}(\mu) = \sum A\mu(A, \mathbf{K}) + \sum_{\mathbf{1}\{A=a\}} A\left\{Y - \mu(A, \mathbf{K})\right\}/\pi$$

contains all the semiparametric estimators (Robins et al., 1994, 1998).

Semiparametric efficiency

For a general estimand $E[AY]$ estimated with outcome data selected at random given A and $\mathbf{K}$, and π the conditional selection probability, the class of estimators

$$N\hat{T}_{diff}(\mu) = \sum A\mu(A, \mathbf{K}) + \sum_{\mathbf{1}\{A=a\}} A\left\{Y - \mu(A, \mathbf{K})\right\}/\pi$$

contains all the semiparametric estimators (Robins et al., 1994, 1998).

The function $\mu_{eff}(a, \mathbf{k}) = E[Y \mid A = a, \mathbf{K} = \mathbf{k}]$ leads to the smallest possible variance (Robins et al., 1994, 1998).

Semiparametric efficiency

For a general estimand $E[AY]$ estimated with outcome data selected at random given A and $\mathbf{K}$, and π the conditional selection probability, the class of estimators

$$N\hat{T}_{diff}(\mu) = \sum A\mu(A, \mathbf{K}) + \sum_{\mathbf{1}\{A=a\}} A\left\{Y - \mu(A, \mathbf{K})\right\}/\pi$$

contains all the semiparametric estimators (Robins et al., 1994, 1998).

The function $\mu_{eff}(a, \mathbf{k}) = E[Y \mid A = a, \mathbf{K} = \mathbf{k}]$ leads to the smallest possible variance (Robins et al., 1994, 1998).

In parallel, similar results were found in the survey sampling literature (see e.g., the discussion in Robins et Rotnitzky (1998) on the paper of Firth and Benneth (1998). Firth and Benneth were interested in *model-based estimators* that are *design consistent*).

Example: The AIPW

The AIPW (see e.g., Robins et al. (1995), Scharfstein et al. (1999), Bang and Robins (2005)) is the most efficient estimator from its class for $E[Y^a]$.

Note we only see Y^a in those for whom $\mathbf{1}\{A = a\}, a \in \{0, 1\}$, and that selection depends on confounders $\mathbf{K}$.

The AIPW (see e.g., Robins et al. (1995), Scharfstein et al. (1999), Bang and Robins (2005)) is the most efficient estimator from its class for $E[Y^a]$.

Note we only see Y^a in those for whom $\mathbf{1}\{A = a\}, a \in \{0, 1\}$, and that selection depends on confounders $\mathbf{K}$.

The estimator for each potential outcome mean is given by

$$N\hat{T}_{diff}(\mu_a) = \sum \frac{Y\mathbf{1}\{A = a\}}{P(A = a \mid \mathbf{K})} - \sum E[Y \mid A = a, \mathbf{K}] \left(\frac{\mathbf{1}\{A = a\} - P(A = a \mid \mathbf{K})}{P(A = a \mid \mathbf{K})} \right)$$

for $\mu_a = E[Y \mid A = a, \mathbf{K}]$.

Construction of the novel estimator

Now suppose instead that we have a selection on both $\mathbf{1}\{A(t) = a\}$ and $\mathbf{1}\{dN(t) = 1\}$. We can start from the AIPW equations:

$$\eta(t) = \quad \frac{\mathbf{1}\{A(t)=a\}}{P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}} Y(t) - \frac{\mathbf{1}\{A(t)=a\} - P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}}{P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}} \mu_a\{\mathbf{K}(t);\hat{\boldsymbol{\alpha}}_K\} - \beta_0 - \beta_1 A(t)$$

Construction of the novel estimator

Now suppose instead that we have a selection on both $\mathbf{1}\{A(t) = a\}$ and $\mathbf{1}\{dN(t) = 1\}$. We can start from the AIPW equations:

$$\eta(t) = \frac{\mathbf{1}\{A(t)=a\}}{P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}} Y(t) - \frac{\mathbf{1}\{A(t)=a\} - P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}}{P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}} \mu_a\{\mathbf{K}(t);\hat{\boldsymbol{\alpha}}_K\} - \beta_0 - \beta_1 A(t)$$

and take the selection on $dN(t)$:

$$E\left[\int_0^\tau \frac{\eta(t)}{E[dN(t) \mid \mathbf{V}(t);\hat{\gamma}]} dN(t)\right] - E\left[\int_0^\tau \left(\frac{dN(t) - E[dN(t) \mid \mathbf{V}(t);\hat{\gamma}]}{E[dN(t) \mid \mathbf{V}(t);\hat{\gamma}]}\right) E[\eta(t) \mid A(t) = a, dN(t) = 1, \mathbf{V}(t);\hat{\boldsymbol{\alpha}}_V]\right] = 0$$

Construction of the novel estimator

Now suppose instead that we have a selection on both $\mathbf{1}\{A(t)=a\}$ and $\mathbf{1}\{dN(t)=1\}$. We can start from the AIPW equations:

$$\eta(t) = \frac{\mathbf{1}\{A(t)=a\}}{P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}} Y(t) - \frac{\mathbf{1}\{A(t)=a\} - P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}}{P\{A(t)=a|\mathbf{K}(t);\hat{\psi}\}} \mu_a\{\mathbf{K}(t);\hat{\boldsymbol{\alpha}}_K\} - \beta_0 - \beta_1 A(t)$$

and take the selection on $dN(t)$:

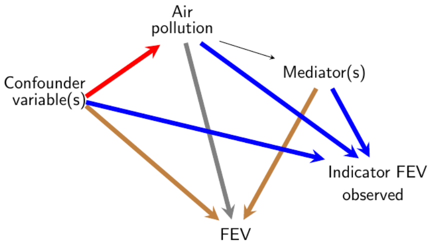
$$E\left[\int_0^\tau \frac{\eta(t)}{E[dN(t) \mid \mathbf{V}(t);\hat{\gamma}]} dN(t)\right] - E\left[\int_0^\tau \left(\frac{dN(t) - E[dN(t) \mid \mathbf{V}(t);\hat{\gamma}]}{E[dN(t) \mid \mathbf{V}(t);\hat{\gamma}]}\right) E[\eta(t) \mid A(t)=a, dN(t)=1, \mathbf{V}(t);\hat{\boldsymbol{\alpha}}_V]\right] = 0$$

which corresponds to projecting the influence function of the AIPW onto spaces orthogonal to residuals from IIV weights and to the projection of the outcome onto $\mathbf{V}(t)$. The novel estimator is called the AAIIW.

Novel estimator

The AAIIW is consistent when at least one of the "treatment-related" models and one of the "measurement-related" models are correctly specified, among:

- $P\{A(t) = a \mid \mathbf{K}(t); \psi\}$
- $\mu_a\{\mathbf{K}(t); \boldsymbol{\alpha}_K\} = E[Y(t) \mid A(t) = a, \mathbf{K}(t); \boldsymbol{\alpha}_K]$ (augmented term)
- $E[dN(t) \mid \mathbf{V}(t); \gamma]$
- $\mu_a\{\mathbf{V}(t); \boldsymbol{\alpha}_V\} = E[Y(t) \mid A(t) = a, \mathbf{V}(t); \boldsymbol{\alpha}_V]$ (augmented term)



Relative asymptotic efficiency

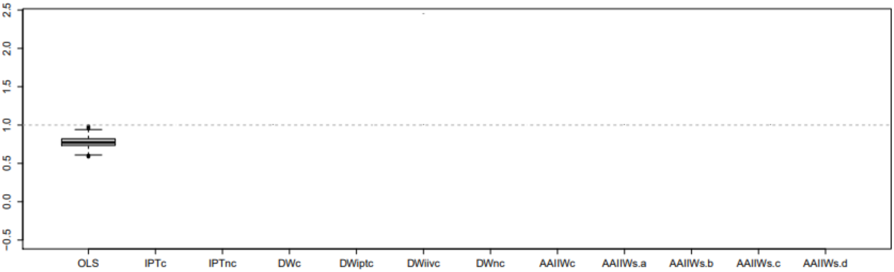
Under all nuisance models correctly specified, the AAIIW attains the semiparametric efficiency bound, and we also find

$$\sigma^2_{AAIIW} - \sigma^2_{DW} = (\beta_0 + \beta_1)^2 E\left[\frac{1-e_1}{e_1} - \frac{2}{e_1}\right] + \beta_0^2 E\left[\frac{1-e_0}{e_0} - \frac{2}{e_0}\right]$$

where $e_0 = P\left\{A(t) = 0 \mid \mathbf{K}(t); \hat{\boldsymbol{\psi}}\right\}, e_1 = P\left\{A(t) = 1 \mid \mathbf{K}(t); \hat{\boldsymbol{\psi}}\right\}$.

Simulation studies, sample size 1000

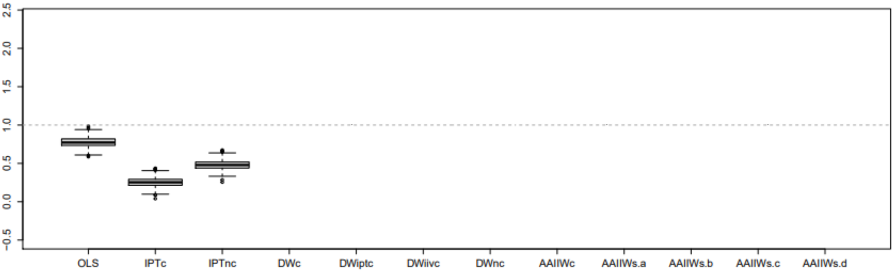
Distribution of 500 estimates:



OLS: Not adjusted for confounding nor irregular observation

Simulation studies, sample size 1000

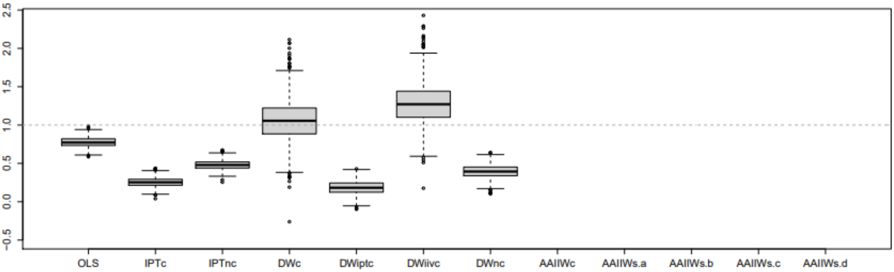
Distribution of 500 estimates:



OLS: Not adjusted for confounding nor irregular observation
IPT: with correctly specified treatment model (c) or not (nc)

Simulation studies, sample size 1000

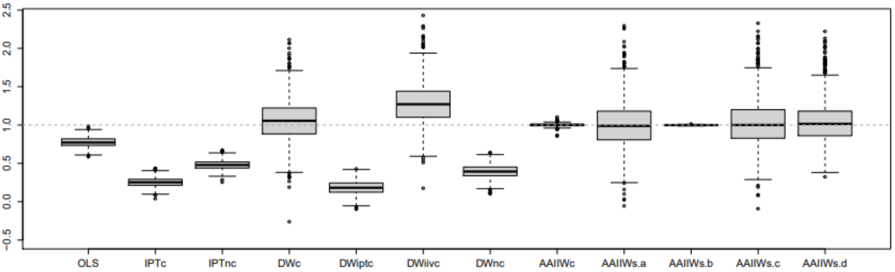
Distribution of 500 estimates:



OLS: Not adjusted for confounding nor irregular observation
IPT: With correctly specified treatment model (c) or not (nc)
DW: With all models (c), only the IPT (iptc), only the IIV (iivc) and none of the model correct (nc)

Simulation studies, sample size 1000

Distribution of 500 estimates:



OLS: Not adjusted for confounding nor irregular observation
IPT: with correctly specified treatment model (c) or not (nc)
DW: With all models (c), only the IPT (iptc), only the IIV (iivc) and none of the model correct (nc)
AAIIWc: All nuisance models correctly specified
AAIIWs.a: Two weight models are correctly specified
AAIIWs.b: Two outcome models are correctly specified
AAIIWs.c: IIV and outcome model conditional on confounders are correct
AAIIWs.d: IPT and outcome model conditional on measurement predictors are correct.

Application the Add Health study (Harris, 2009)

- American data on adolescents (4 waves: 1994-1995, 1996, 2001-2002, and 2008-2009)
- Question: What is the ATE of current therapy counseling on current alcohol consumption?

Application the Add Health study (Harris, 2009)

- American data on adolescents (4 waves: 1994-1995, 1996, 2001-2002, and 2008-2009)
- Question: What is the ATE of current therapy counseling on current alcohol consumption?
- Several confounders of the therapy-alcohol consumption relationship (included age, sex, socioeconomic status (SES), weight, smoking)

Application the Add Health study (Harris, 2009)

- American data on adolescents (4 waves: 1994-1995, 1996, 2001-2002, and 2008-2009)
- Question: What is the ATE of current therapy counseling on current alcohol consumption?
- Several confounders of the therapy-alcohol consumption relationship (included age, sex, socioeconomic status (SES), weight, smoking)
- Emulated missingness in the longitudinal outcome as a function of age, sex, counseling, depressive mood (mediator), SES

Application the Add Health study (Harris, 2009)

- American data on adolescents (4 waves: 1994-1995, 1996, 2001-2002, and 2008-2009)
- Question: What is the ATE of current therapy counseling on current alcohol consumption?
- Several confounders of the therapy-alcohol consumption relationship (included age, sex, socioeconomic status (SES), weight, smoking)
- Emulated missingness in the longitudinal outcome as a function of age, sex, counseling, depressive mood (mediator), SES
- Comparison of OLS, IPT, IIV, DW, and AAIIW with† or without‡ the correct measurement model

Results of the application

Table 1: Estimates (95% bootstrap percentiles confidence intervals) of the marginal effect of counselling on the average number of alcoholic beverages consumed, *Add Health* study, United States, 1996-2008

OLS	IPT	IIV†	IIV‡
0.62 (0.39, 0.75)	0.34 (0.15, 0.48)	0.64 (0.40, 0.77)	0.72 (0.49, 0.92)
DW†	**DW‡**	**AAIIW†**	**AAIIW‡**
0.35 (0.12, 0.50)	0.46 (0.24, 0.67)	0.35 (0.12, 0.51)	0.28 (0.04, 0.54)

†. Uses correctly specified IIV weights. ‡. Uses wrong IIV weights.

- Adjustments bring the estimates towards 0 in general

Results of the application

Table 1: Estimates (95% bootstrap percentiles confidence intervals) of the marginal effect of counselling on the average number of alcoholic beverages consumed, *Add Health* study, United States, 1996-2008

OLS	IPT	IIV†	IIV‡
0.62 (0.39, 0.75)	0.34 (0.15, 0.48)	0.64 (0.40, 0.77)	0.72 (0.49, 0.92)
DW†	DW‡	AAIIW†	AAIIW‡
0.35 (0.12, 0.50)	0.46 (0.24, 0.67)	0.35 (0.12, 0.51)	0.28 (0.04, 0.54)

†. Uses correctly specified IIV weights. ‡. Uses wrong IIV weights.

- Adjustments bring the estimates towards 0 in general
- Wrong IIV weights (Robustness of the AAIIW?):
  - DW‡: 0.46 (0.24, 0.67)
  - AAIIW‡ : 0.28 (0.04, 0.54), i.e., not so far from AAIIW†

Discussion

- Inherent condition on observed data: How does it affect estimation?

Discussion

- Inherent condition on observed data: How does it affect estimation? It requires some deep thoughts about the relationships in the data and the question of interest.

Discussion

- Inherent condition on observed data: How does it affect estimation? It requires some deep thoughts about the relationships in the data and the question of interest.
- Important limitations: Causal and modeling assumptions, acute treatment effect, availability of other longitudinal processes than the outcome.

Discussion

- Inherent condition on observed data: How does it affect estimation? It requires some deep thoughts about the relationships in the data and the question of interest.
- Important limitations: Causal and modeling assumptions, acute treatment effect, availability of other longitudinal processes than the outcome.
- Future avenues of work: Look at the cumulative effect of treatment or the effect of a series of treatments on the irregularly-measured outcome, and develop methods that can address irregularly observed $\mathbf{V}(t)$.

Where to find: The manuscript is under review and available on ArXiv (Coulombe et Yang, 2023).

References I

- Coulombe, Janie and Shu Yang (2023). In: https://arxiv.org/abs/2304.08987.
- Firth, David and KE Series B Bennett (1998). "Robust models in probability sampling". In: Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60.1, pp. 3–21.
- Harris, Kathleen Mullan (2009). "The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007-2009 [machine-readable data file and documentation]". In: Chapel Hill: Carolina Population Center, University of North Carolina at Chapel Hill.
- Horvitz, Daniel G and Donovan J Thompson (1952). "A generalization of sampling without replacement from a finite universe". In: Journal of the American Statistical Association 47.260, pp. 663–685.
- Lin, Haiqun, Daniel O Scharfstein, and Robert A Rosenheck (2004). "Analysis of longitudinal data with irregular, outcome-dependent follow-up". In: Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66.3, pp. 791–813.
- Neyman, Jerzy S (1923). "On the application of probability theory to agricultural experiments. Essay on principles. Section 9 (translation published in 1990)". In: Statistical Science 5.4, pp. 472–480.

References II

Robins, J and A Rotnitzky (1998). "Discussion of "Robust models in probability sampling" by Firth and Bennett". In: Journal of the Royal Statistical Society: Series B (Statistical Methodology), pp. 51–52.

Robins, James M, Miguel Angel Hernan, and Babette Brumback (2000). "Marginal structural models and causal inference in epidemiology". In: Epidemiology 11.5, pp. 550–560.

Robins, James M, Andrea Rotnitzky, and Lue Ping Zhao (1994). "Estimation of regression coefficients when some regressors are not always observed". In: Journal of the American Statistical Association 89.427, pp. 846–866.

Rosenbaum, Paul R and Donald B Rubin (1983). "The central role of the propensity score in observational studies for causal effects". In: Biometrika 70.1, pp. 41–55.

Rubin, Donald B (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies". In: Journal of Educational Psychology 66.5, p. 688.

Tsiatis, Anastasios A (2006). Semiparametric theory and missing data. Springer.