Recent advances in causal inference under irregular observation times for the outcome

Janie Coulombe

Department of Mathematics and Statistics, Université de Montréal

EpiBio Seminar Series

Dept. of Epidemiology and Biostatistics, Schulich School of Medicine & Dentistry, Western University

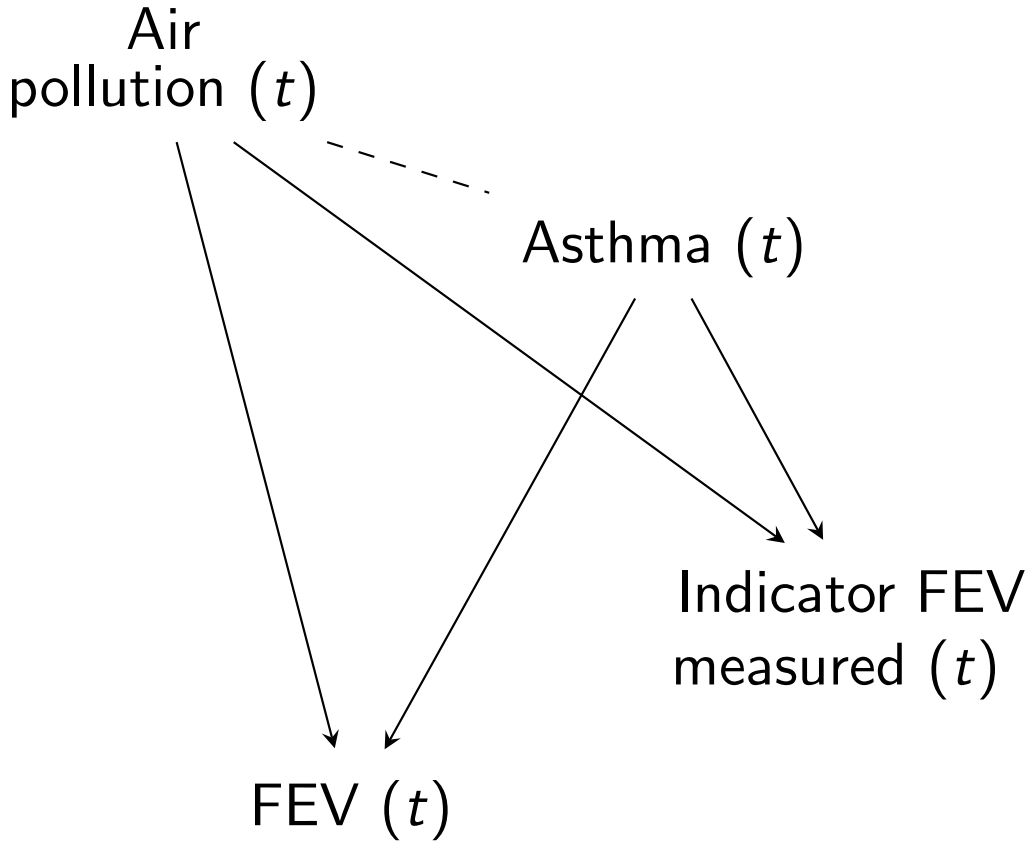
Oct. 6, 2023

Outline

Examples

Ex. 1 - Buzkova and Lumley, 2005

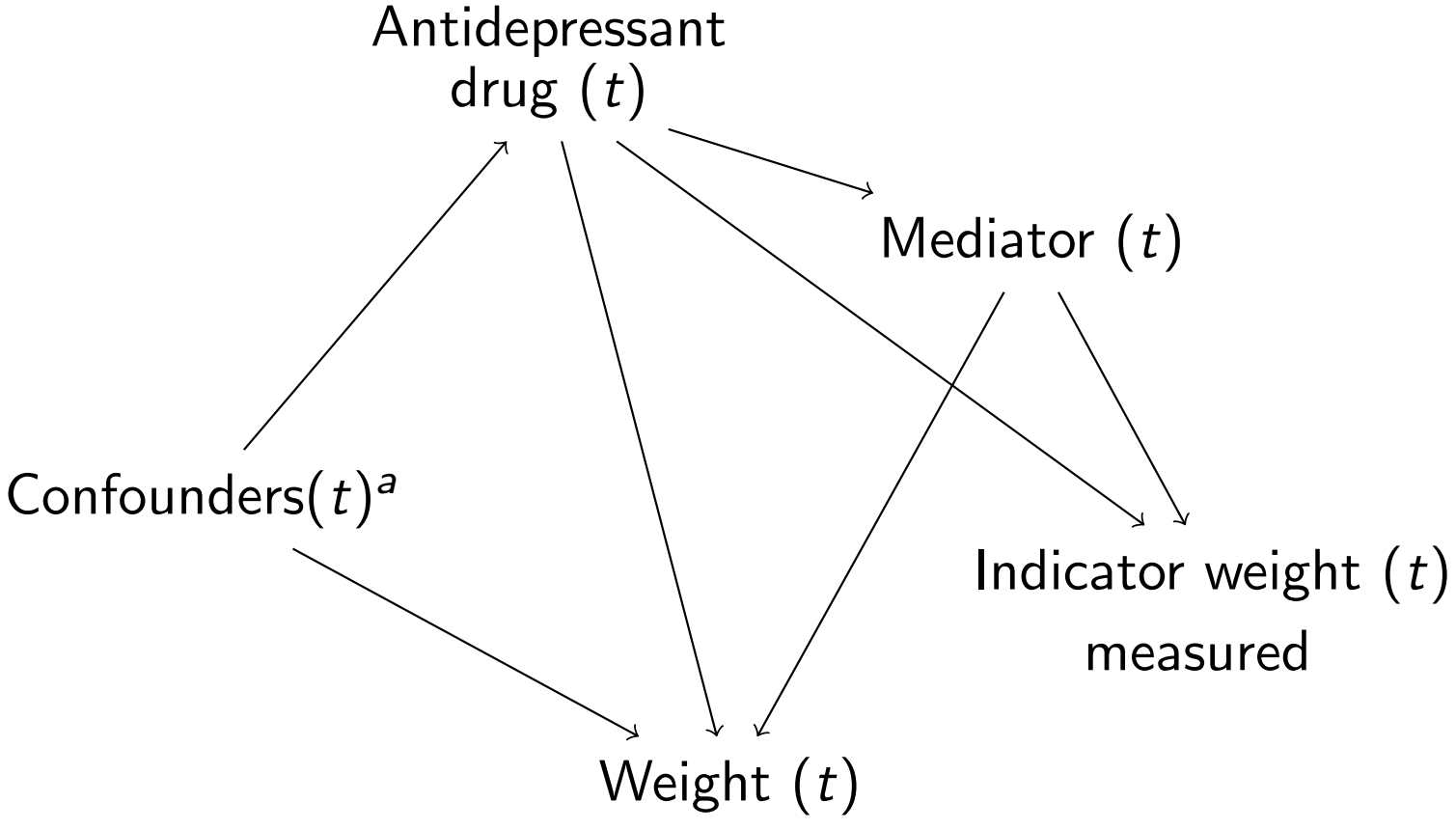
Causal diagram at time t (constructed from their example):



Interest in the marginal association between air pollution and forced expiratory volume (FEV).

Ex. 2 - EHR from the Clinical Practice Research Datalink

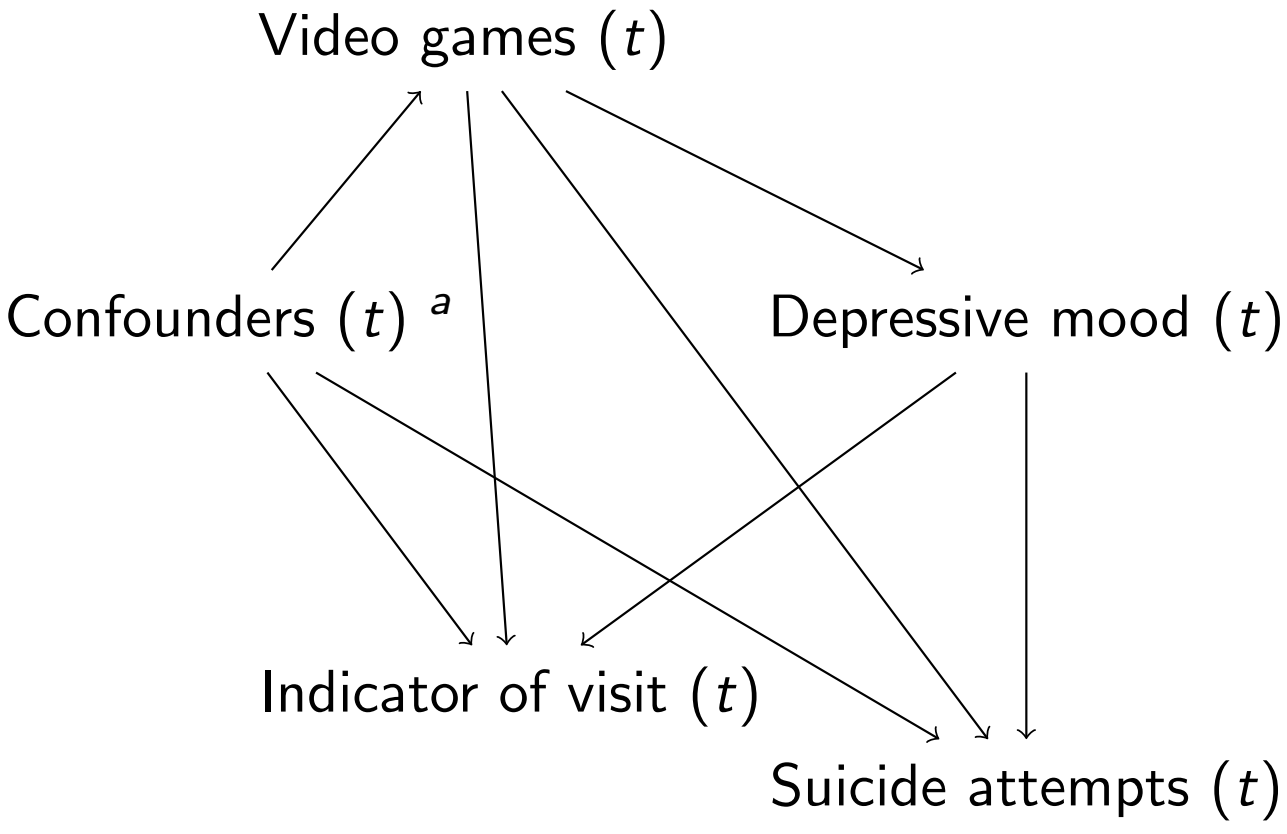
Coulombe, Moodie, Renoux and Platt, 2022 - causal diagram at time t:



[a] Included age, sex, Index of Multiple Deprivation, diabetes, alcohol abuse, anxiety, other psychiatric diseases, number of previous hospitalisations, and others.

Ex. 3 - Application to a continuous exposure

Coulombe, Moodie and Platt, 2021a - diagram at time t:



[a] Included age, sex, SES, ethnicity, different grades in school, trouble relaxing, grooming, seeming bored or impatient, frequency of hanging out with friends, feeling cared about, and no. of cigarettes smoked.

Setting

- In all these examples, we assume **repeated irregular measurements** of the outcome denoted $Y(t)$ at time t
- We are interested in **marginal** associations (or causal effects) of a treatment with the outcome

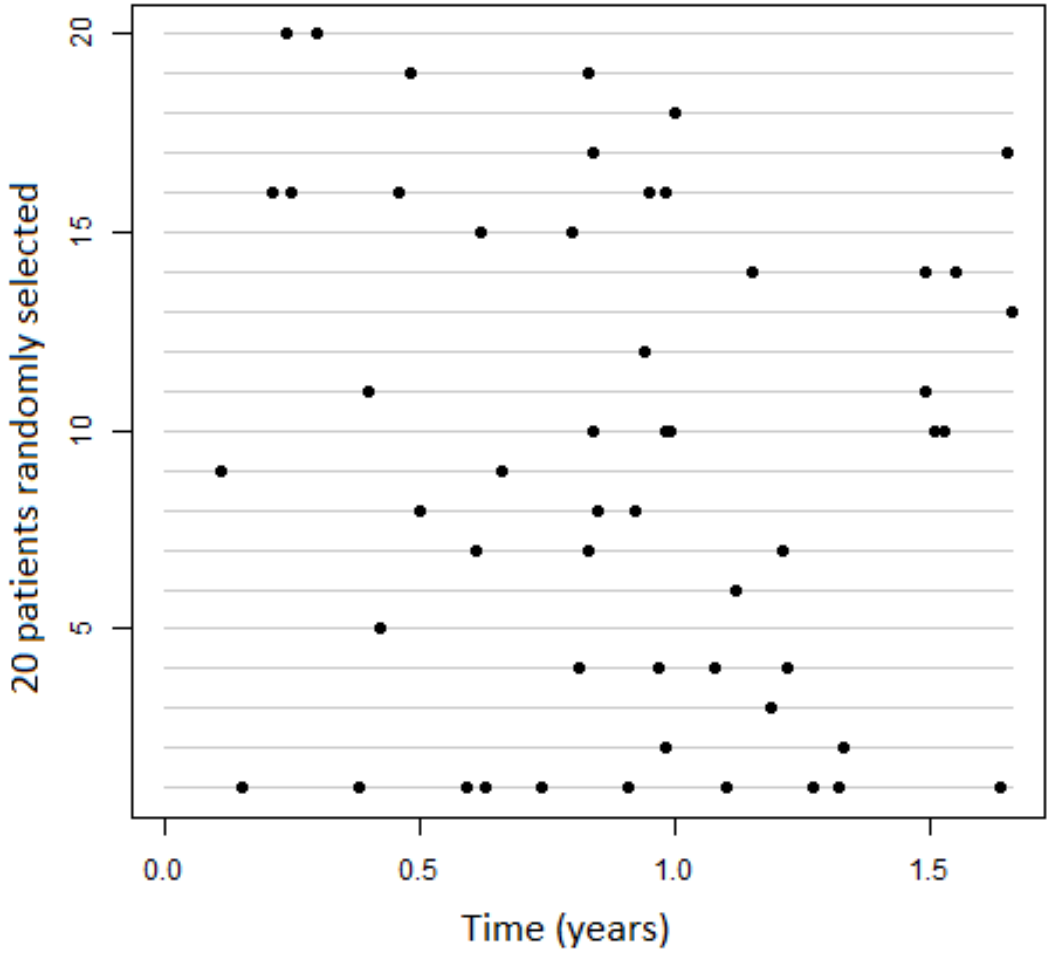
Setting

- In all these examples, we assume **repeated irregular measurements** of the outcome denoted $Y(t)$ at time t
- We are interested in **marginal** associations (or causal effects) of a treatment with the outcome
- We denote the (binary) treatment or exposure $A(t)$.
- We may be interested in binary or continuous treatment.

Setting

- In all these examples, we assume **repeated irregular measurements** of the outcome denoted $Y(t)$ at time t
- We are interested in **marginal** associations (or causal effects) of a treatment with the outcome
- We denote the (binary) treatment or exposure $A(t)$.
- We may be interested in binary or continuous treatment.
- We assume having a random sample from a larger population, corresponding to a cohort of size N followed over τ year(s) of follow-up.

Setting



Note. This plot is called an abacus plot and can be produced with the `IrregLong` package proposed by Pullenayegum (2022).

Estimand

Suppose a binary treatment first.

Estimand

Suppose a binary treatment first. Denote by $Y^1(t)$ and $Y^0(t)$ the two potential outcomes (Neyman-Rubin, 1923 & 1974) corresponding to the different levels of treatment. We assume an acute effect.

Estimand

Suppose a binary treatment first. Denote by $Y^1(t)$ and $Y^0(t)$ the two potential outcomes (Neyman-Rubin, 1923 & 1974) corresponding to the different levels of treatment. We assume an acute effect.

Estimand

Suppose a binary treatment first. Denote by $Y^1(t)$ and $Y^0(t)$ the two potential outcomes (Neyman-Rubin, 1923 & 1974) corresponding to the different levels of treatment. 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

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

Estimand

Suppose a binary treatment first. Denote by $Y^1(t)$ and $Y^0(t)$ the two potential outcomes (Neyman-Rubin, 1923 & 1974) corresponding to the different levels of treatment. 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

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

We may also be interested in developing (one-stage) adaptive treatment strategies, in which case we are more interested in

$$E[Y^1(t) - Y^0(t) \mid X]$$

for different values of X.

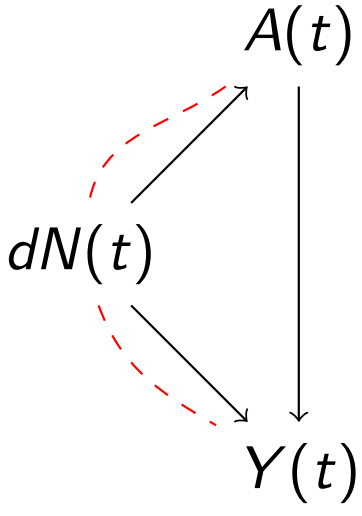
The issue

Why are irregular measurements problematic?

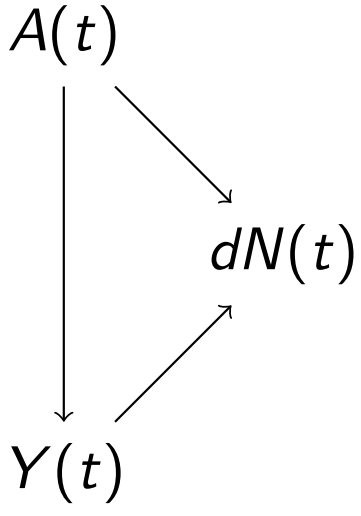
Denote a counting process for the outcome measurements by $N(t)$ and $dN(t)$ and indicator of jump in the process at t.

Why are irregular measurements problematic?

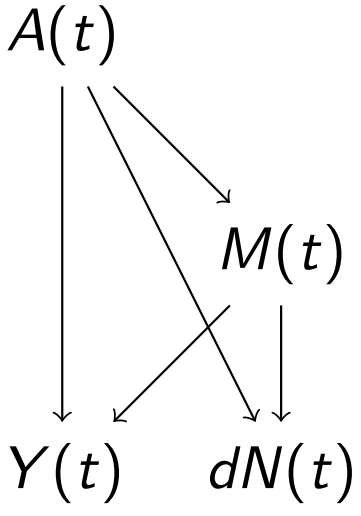
Denote a counting process for the outcome measurements by $N(t)$ and $dN(t)$ and indicator of jump in the process at t.



Confounder



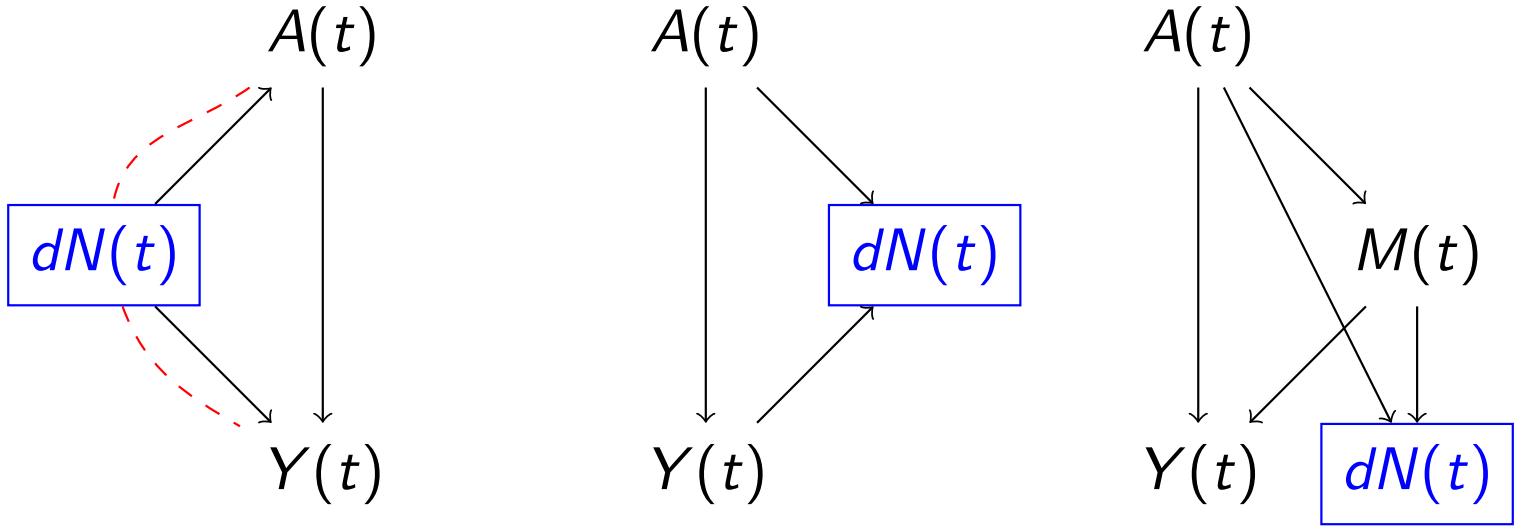
Collider



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.

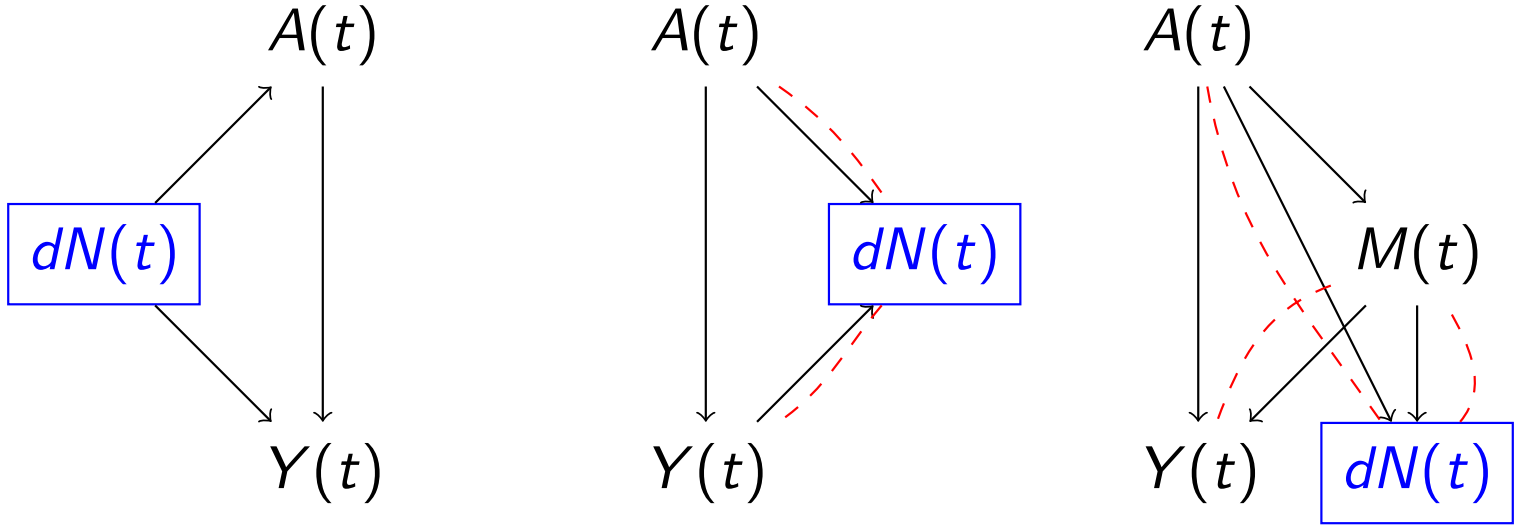


Confounder Collider Collider

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.



Confounder Collider Collider

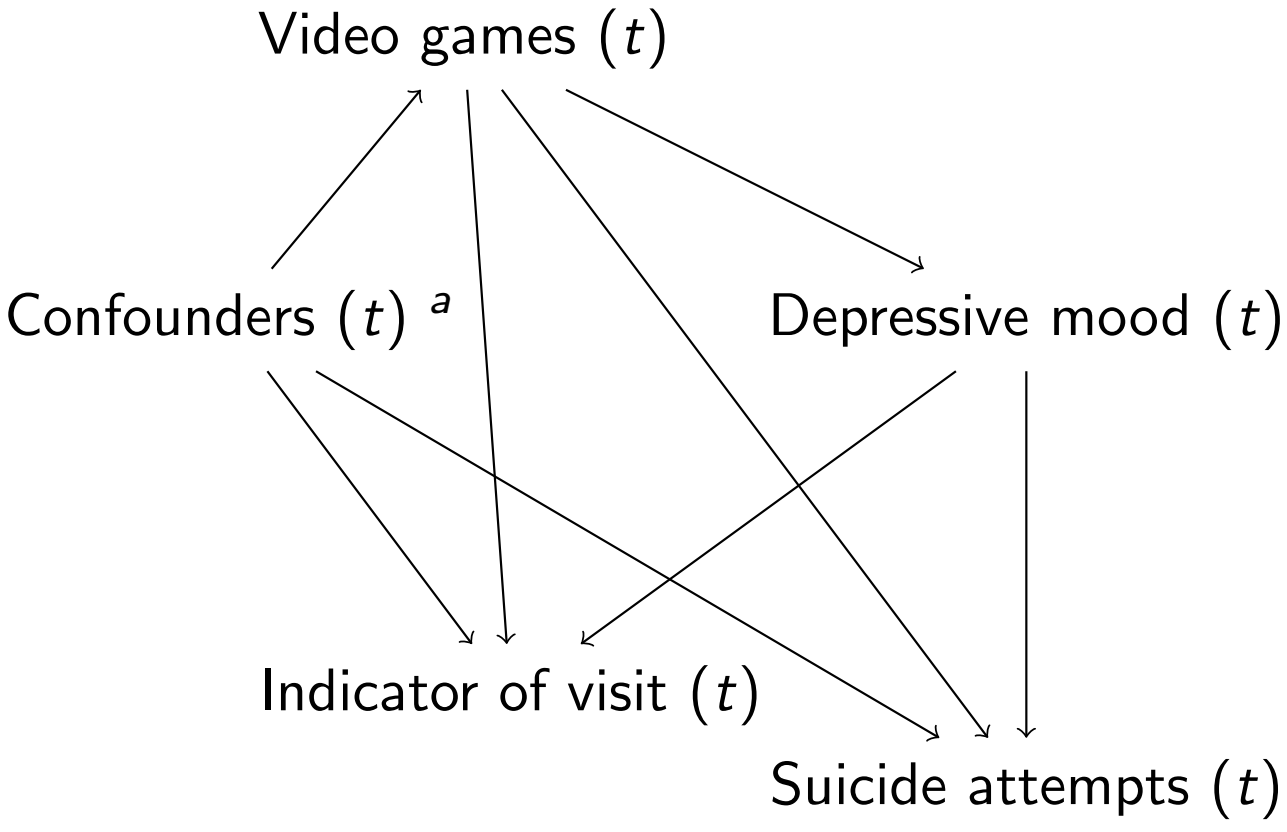
There is almost always a condition on the measurement indicator $dN(t)$ (i.e., a box around it).

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.

Back to an example

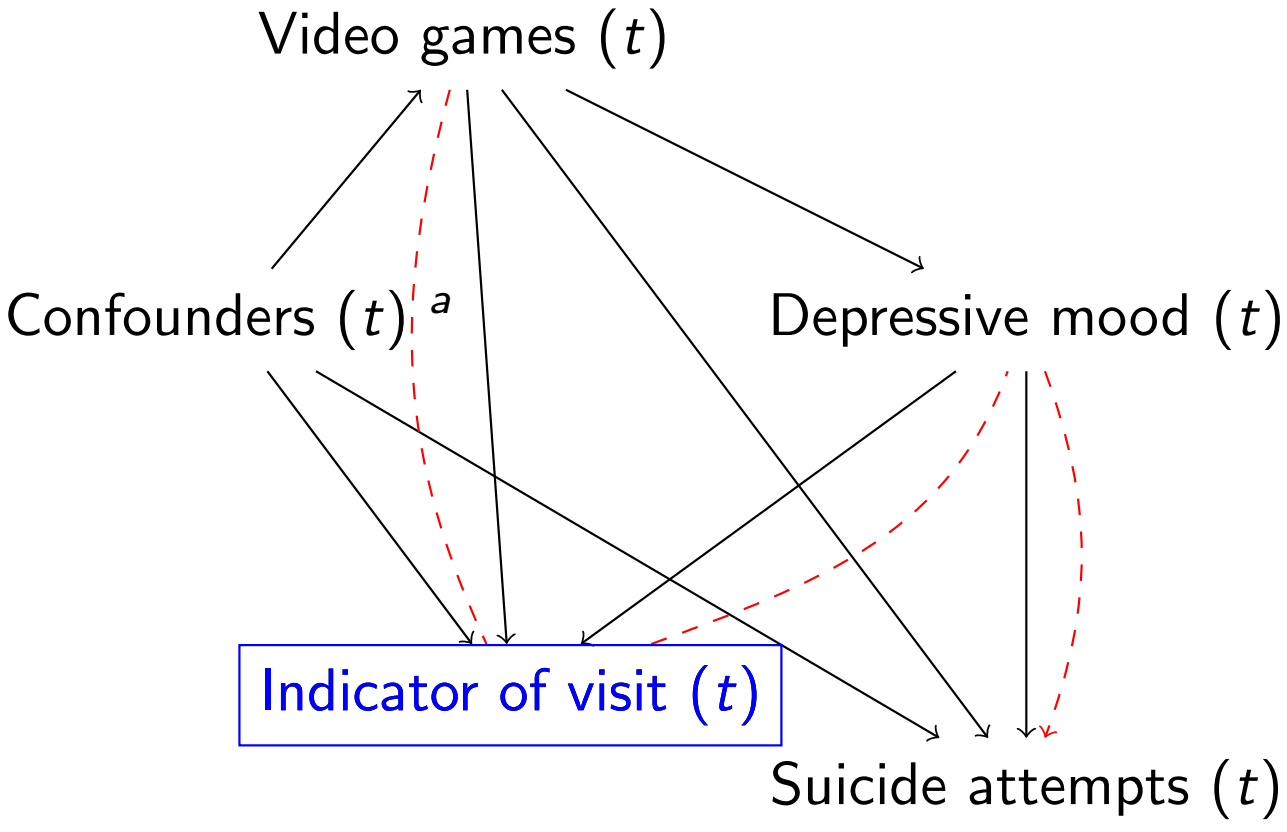
In the study of Coulombe et al. (2021a), we had:



[a] Included age, sex, SES, ethnicity, different grades in school, trouble relaxing, grooming, seeming bored or impatient, frequency of hanging out with friends, feeling cared about, and no. of cigarettes smoked.

Back to an example

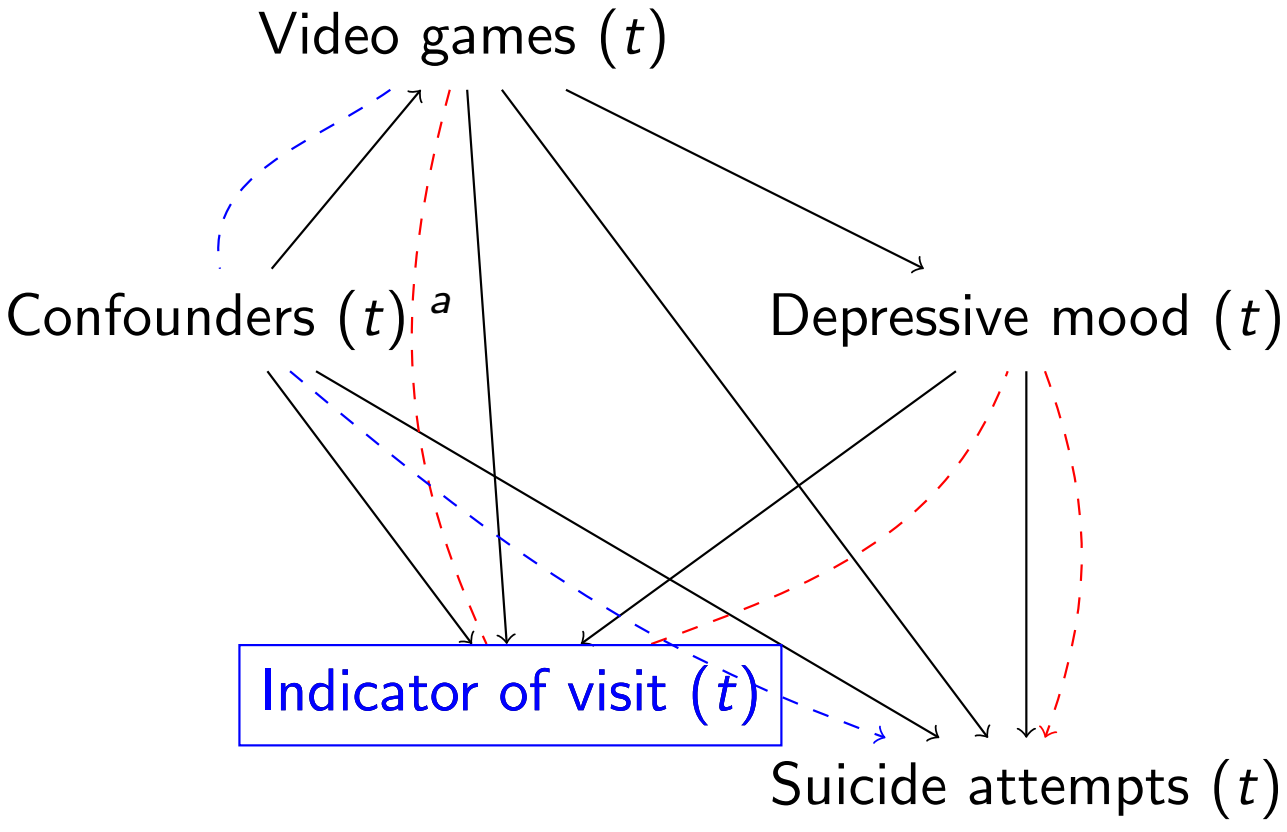
In the study of Coulombe et al. (2021a), we had:



[a] Included age, sex, SES, ethnicity, different grades in school, trouble relaxing, grooming, seeming bored or impatient, frequency of hanging out with friends, feeling cared about, and no. of cigarettes smoked.

Back to an example

In the study of Coulombe et al. (2021a), we had:



[a] Included age, sex, SES, ethnicity, different grades in school, trouble relaxing, grooming, seeming bored or impatient, frequency of hanging out with friends, feeling cared about, and no. of cigarettes smoked.

Methods

Notation

Let:

- $A_i(t)$ and $Y_i(t)$ the binary treatment and the continuous outcome

Notation

Let:

- $A_i(t)$ and $Y_i(t)$ the binary treatment and the continuous outcome
- $\mathbf{K_i(t)}$ be the confounding variables

Notation

Let:

- $A_i(t)$ and $Y_i(t)$ the binary treatment and the continuous outcome
- $\mathbf{K_i(t)}$ be the confounding variables
- $dN_i(t)$ the visit indicator ($=$ observation of $Y_i(t)$)
- C_i the follow-up time of patient i
- $\xi_i(t) = \mathbb{I}(C_i \geq t)$ the indicator for still being at risk at time t

Notation

Let:

- $A_i(t)$ and $Y_i(t)$ the binary treatment and the continuous outcome
- $\mathbf{K_i(t)}$ be the confounding variables
- $dN_i(t)$ the visit indicator ($=$ observation of $Y_i(t)$)
- C_i the follow-up time of patient i
- $\xi_i(t) = \mathbb{I}(C_i \geq t)$ the indicator for still being at risk at time t
- We assume $dN_i(t) \perp Y_i(t) | \mathbf{V_i(t)}$ for a set of variables $\mathbf{V_i(t)}$

Causal assumptions

General assumptions we make with the methods I present today:

- 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): $\left\{Y^0(t), Y^1(t)\right\} \perp A(t) \mid \mathbf{K}(t)$
- Conditional exchangeability:
  $$\text{NUC} + \{Y^0(t), Y^1(t)\} \perp dN(t) \mid \mathbf{V}(t)$$

First estimator

Coulombe, Moodie and Platt (2021b).

The first estimator, which is doubly-weighted (I will denote it by DW), solves the following equations for β_0 and β_1 (the ATE):

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

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

Model for the visit indicator

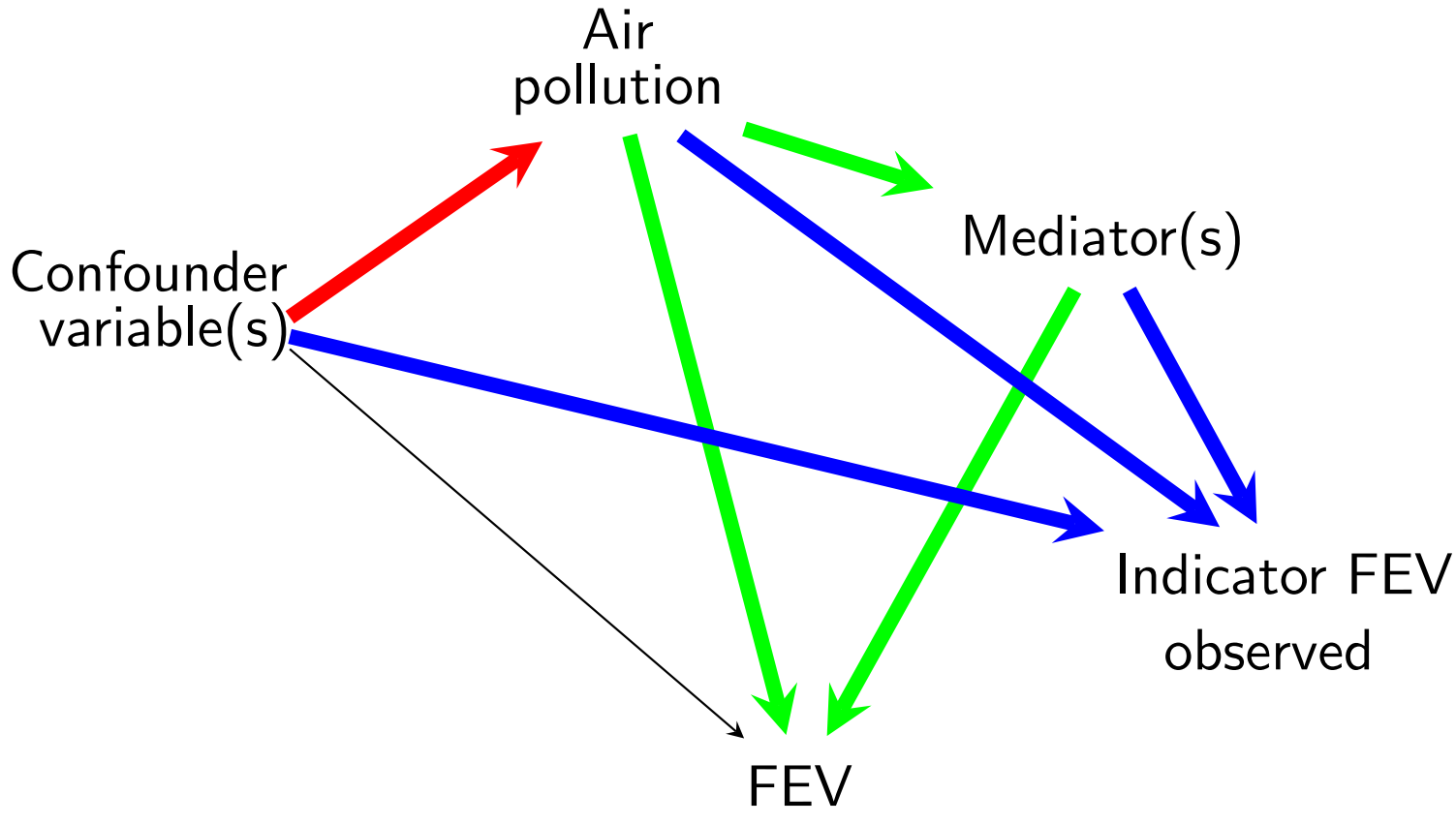
In general, a proportional rate model is used for $E[dN(t) \mid \mathbf{V}(t); \hat{\gamma}]$ - i.e., we estimate the parameters for the model

$$E[dN(t) \mid \mathbf{V}(t); \gamma] = \lambda_0(t) \exp\left\{\gamma^T \mathbf{V}(\mathbf{t})\right\} dt$$

which assumes a proportional effect of covariates $\mathbf{V}(\mathbf{t})$ on visit (or observation) rates.

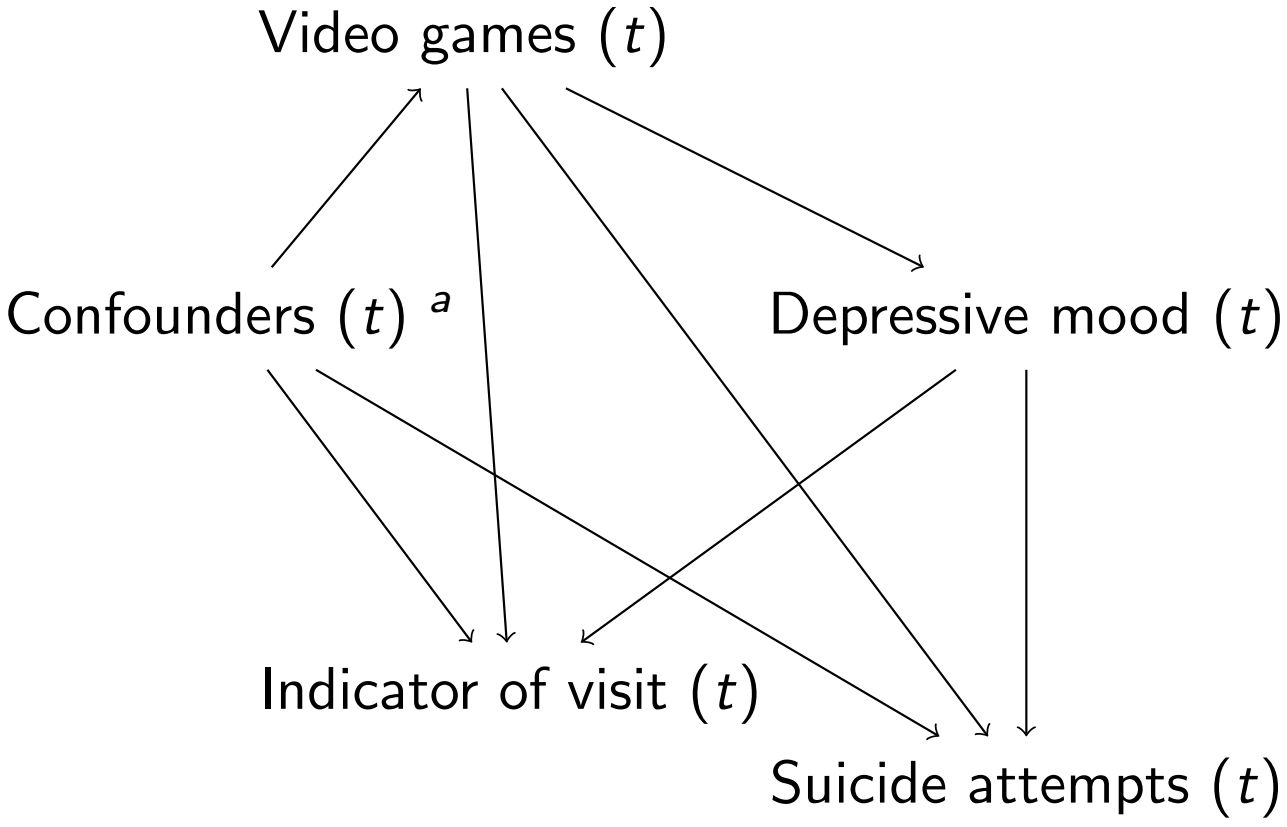
The Andersen and Gill model (1982) can be used to estimate the γ parameters.

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



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

Now, suppose the following



[a] Included age, sex, SES, ethnicity, different grades in school, trouble relaxing, grooming, seeming bored or impatient, frequency of hanging out with friends, feeling cared about, and no. of cigarettes smoked.

Additional challenge: The exposure is continuous (somewhat) and the outcome is ordinal.

Extension with the proportional odds model

Generalized IPT weights can be used to model the exposure as a function of confounders (Robins et al., 2000).

The proportional odds model (POM) (McCullagh, 1980) can account for the ordinal nature of the outcome:

$$\boldsymbol{\zeta}_i(\mathbf{t}) = \begin{bmatrix} \mathbb{P}(Y_i(t) \leq 1 | D_i(t)) \\ \mathbb{P}(Y_i(t) \leq 2 | D_i(t)) \\ \dots \\ \mathbb{P}(Y_i(t) \leq J | D_i(t)) \end{bmatrix}, \text{ and } \mathbf{Q_i}(\mathbf{t}) = \begin{bmatrix} \mathbb{I}(Y_i(t) \leq 1) \\ \mathbb{I}(Y_i(t) \leq 2) \\ \dots \\ \mathbb{I}(Y_i(t) \leq J) \end{bmatrix}.$$

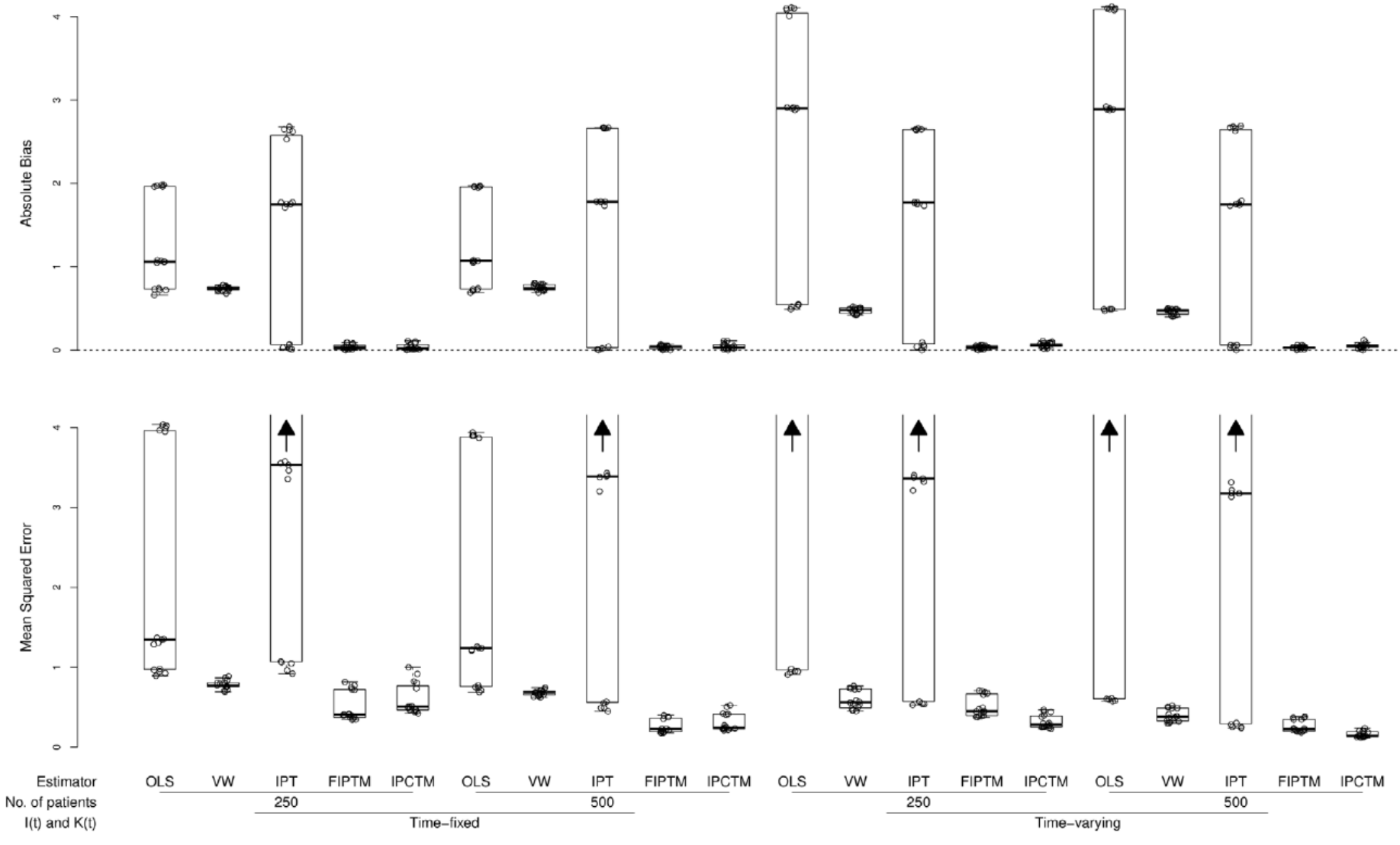
The new estimating equation is given by:

$$\mathbb{E}\left[\int_0^\tau \frac{\mathbf{e}(\mathbf{t};\boldsymbol{\psi})(\mathbf{Q}(\mathbf{t}) - \boldsymbol{\zeta}(\mathbf{t}))}{\boldsymbol{\varphi}(\mathbf{t};\boldsymbol{\gamma}_\mathbf{V})} \mathbf{dN}(\mathbf{t})\right] = \mathbf{0},$$

where $\zeta_{i,j}(t) = \text{expit}(\alpha_j - \beta_D D_i(t))$.

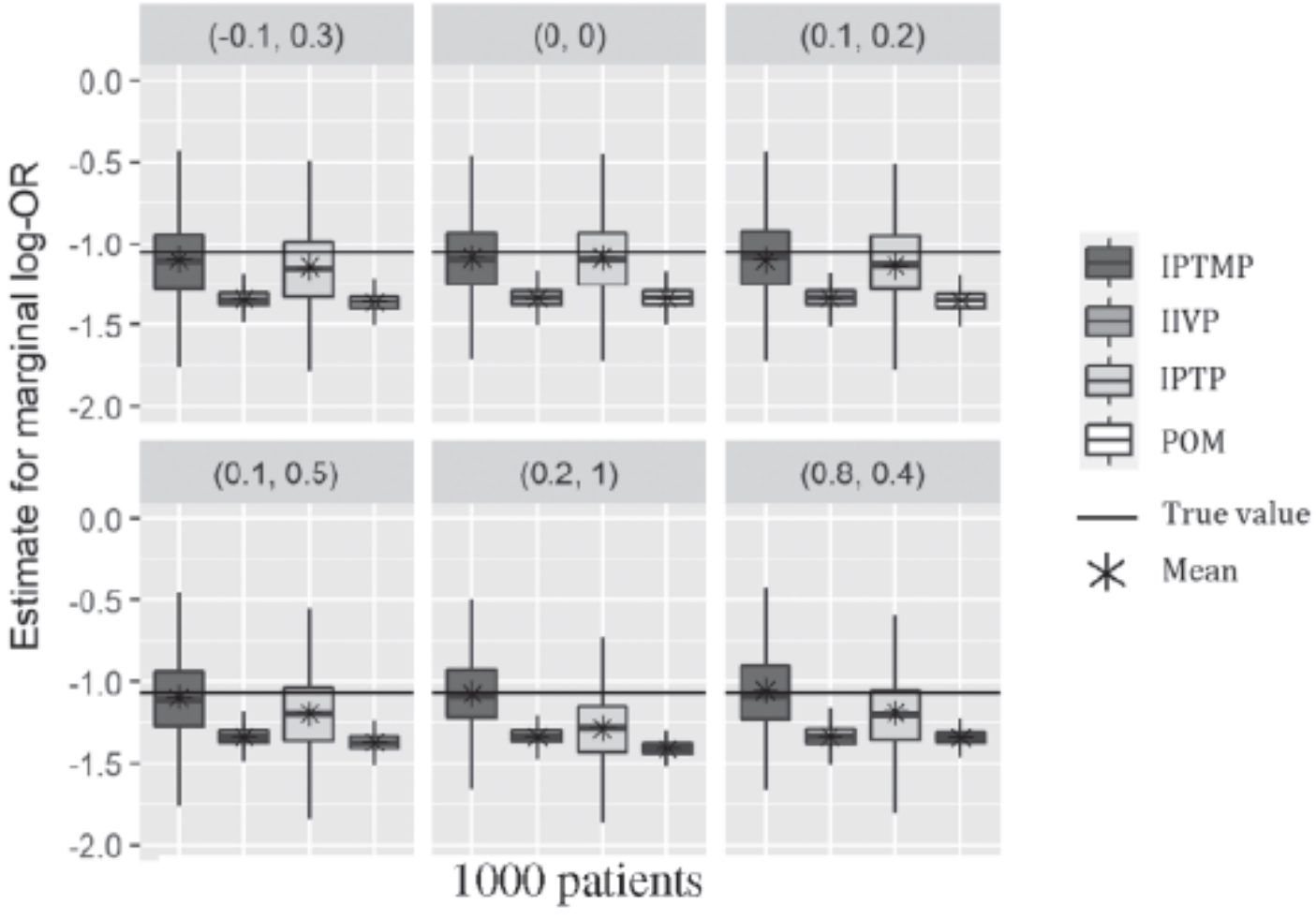
Performance

1. Method for continuous outcomes (DW):



Performance

2. Method for ordinal outcomes (POM-DW):



In applications...

It was interesting to evaluate how much covariates affect the observation of the outcome...

Example 1: Effect of **depressive mood** on the outcome **weight**

TABLE 3 Average rate ratios and 95% confidence intervals for variables in the proportional intensity model for monitoring times

Variable	Rate ratio	95 % CI
Depressive mood	0.93	0.84; 1.02
Smoking	1.08	1.03; 1.13
Age	0.94	0.93; 0.94
Sex = female	1.04	1.01; 1.07
SES	1.00	0.99; 1.01

In applications...

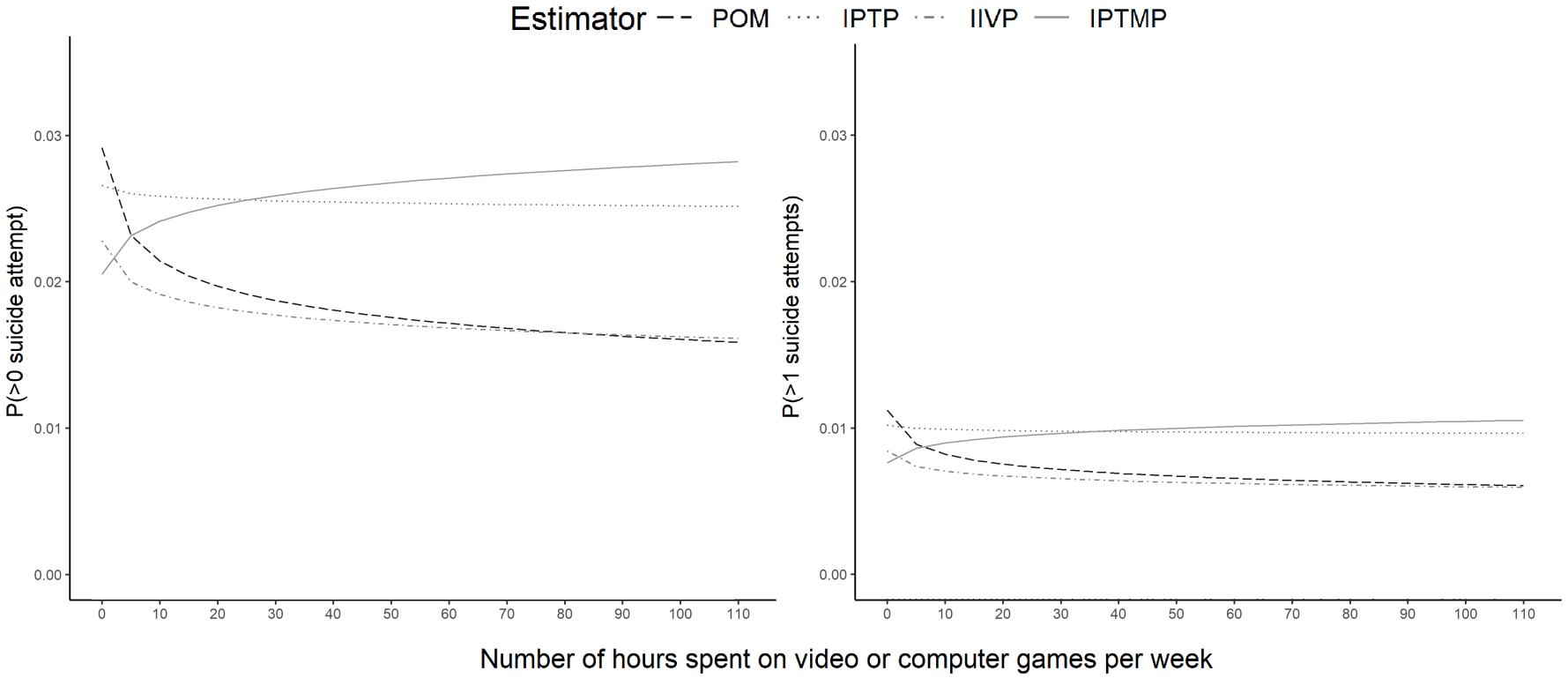
Example 2: Effect of **video games** on the outcome **suicide attempts**

TABLE 3 Estimated rate ratios (95% CI) for the monitoring model, *Add Health* study, United States, 1994 to 2008, $n = 6504$ individuals

Variable	Rate ratio (bootstrap 95% CI)
Number of hours spent on video or computer games	1.00 (1.00, 1.00)
Frequency of feeling depressed (Ref. = Never or rarely)	
Sometimes	1.00 (0.97, 1.02)
A lof of the time	0.99 (0.94, 1.03)
Most of the time or all the time	1.01 (0.94, 1.08)
Age	0.93 (0.93, 0.94)
Sex (female)	1.11 (1.09, 1.13)
SES	1.01 (1.01, 1.02)
Race (Ref. = White)	
Black/African American	0.93 (0.91, 0.95)
American Indian/Alaskan Native	0.96 (0.87, 1.04)
Asian/Pacific Islander	0.92 (0.88, 0.97)
Other	0.90 (0.86, 0.94)

In applications...

Example 2: Effect of **video games** on the outcome **suicide attempts**



In applications...

Example 3: Effect of **antidepressants** on **weight variations**

TABLE 2
Rate ratios from the visit intensity model, Clinical Practice Research Datalink, U.K., 1998–2017

Variable	Rate Ratio	95% CI
Citalopram (Ref.: Fluoxetine)	0.95	0.93, 0.96*
Age at baseline	1.00	0.99, 1.00
Sex (Ref.: Female)	0.77	0.76, 0.78*
IMD at baseline	1.06	1.05, 1.06*
Smoking (Ref.: Never)		
Ever	0.92	0.91, 0.94*
Missing	0.23	0.23, 0.23*
Diabetes	2.07	2.02, 2.12*
Alcohol abuse	1.14	1.07, 1.22*
Anxiety or GAD	1.00	0.98, 1.03
Psychiatric diagnosis	1.03	0.93, 1.13
Number of hospitalisations in prior month	0.98	0.96, 1.01
Antipsychotic drugs	1.12	1.07, 1.18*
Benzodiazepine drugs	1.20	1.16, 1.23*
Lipid lowering drugs	1.21	1.18, 1.25*

Abbreviations: IMD, Index of multiple deprivation; GAD, Generalized anxiety disorder. *Confidence interval does not contain 1.

Adaptive treatment strategies

Irregular observation times can also occur in data used to develop adaptive treatment strategies (ATS).

ATS are functions of causal effects so they require consistent estimation of causal effects.

We were interested in developing a one-stage treatment rule to pick an optimal antidepressant drug to reduce detrimental weight changes.

Adaptive treatment strategies

Let:

- $\mathbf{Q}(\mathbf{t})$ be a matrix of "pure predictors" that could be effect modifiers for the treatment's effect on the outcome

Adaptive treatment strategies

Let:

- $\mathbf{Q}(\mathbf{t})$ be a matrix of "pure predictors" that could be effect modifiers for the treatment's effect on the outcome
- The set $\mathbf{X}^{\beta}(\mathbf{t}) = [\mathbf{1} \;\; \mathbf{K}(\mathbf{t}) \;\; \mathbf{Q}(\mathbf{t})]$ that contains the confounders

Adaptive treatment strategies

Let:

- $\mathbf{Q}(\mathbf{t})$ be a matrix of "pure predictors" that could be effect modifiers for the treatment's effect on the outcome
- The set $\mathbf{X}^{\beta}(\mathbf{t}) = [\mathbf{1} \;\; \mathbf{K}(\mathbf{t}) \;\; \mathbf{Q}(\mathbf{t})]$ that contains the confounders
- A matrix containing the effect modifiers, denoted by $\mathbf{X}^{\psi}(\mathbf{t})$

Adaptive treatment strategies

Let:

- $\mathbf{Q}(\mathbf{t})$ be a matrix of "pure predictors" that could be effect modifiers for the treatment's effect on the outcome
- The set $\mathbf{X}^{\beta}(\mathbf{t}) = [\mathbf{1} \;\; \mathbf{K}(\mathbf{t}) \;\; \mathbf{Q}(\mathbf{t})]$ that contains the confounders
- A matrix containing the effect modifiers, denoted by $\mathbf{X}^{\psi}(\mathbf{t})$
- $\mathbf{X}(\mathbf{t}) = \left[\mathbf{X}^{\beta}(\mathbf{t}) \;\; \mathbf{X}^{\psi}(\mathbf{t})\right]$ the design matrix for the estimating equations to come

Adaptive treatment strategies

Let:

- $\mathbf{Q}(\mathbf{t})$ be a matrix of "pure predictors" that could be effect modifiers for the treatment's effect on the outcome
- The set $\mathbf{X}^{\beta}(\mathbf{t}) = [\mathbf{1} \;\; \mathbf{K}(\mathbf{t}) \;\; \mathbf{Q}(\mathbf{t})]$ that contains the confounders
- A matrix containing the effect modifiers, denoted by $\mathbf{X}^{\psi}(\mathbf{t})$
- $\mathbf{X}(\mathbf{t}) = \left[\mathbf{X}^{\beta}(\mathbf{t}) \;\; \mathbf{X}^{\psi}(\mathbf{t})\right]$ the design matrix for the estimating equations to come

(all the rest stays the same!)

Adaptive treatment strategies

We extended in Coulombe et al. (2023) an approach proposed by Wallace and Moodie (2015).

The outcome model includes a treatment-free model and a blip:

$$\mathbb{E}[Y_i(t)|A_i(t), \mathbf{X}_i(t)] = f\left\{\mathbf{X}_i^{\beta}(t); \boldsymbol{\beta}\right\} + A_i(t)\psi^T\mathbf{X}_i^{\psi}(t)$$

The following rule:

"**Treat with citalopram if the outcome expectation is optimized under citalopram, and with fluoxetine otherwise**"

becomes

"**Treat with citalopram if** $\hat{\psi}^T\mathbf{X}_i^{\psi}(t) \geq 0$**, and with fluoxetine otherwise.**"

Adaptive treatment strategies

The (doubly-robust) estimating equation to solve:

$$U(\boldsymbol{\beta}, \boldsymbol{\psi}; \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\omega}}) = \sum_{i=1}^{n} \int_0^{\tau} \varphi_i(\hat{\boldsymbol{\gamma}}, \mathbf{V_i(t)}) e_i(\hat{\boldsymbol{\omega}}, \mathbf{K_i(t)})$$

$$\times \begin{bmatrix} \frac{\partial f\left\{\mathbf{X}_i^{\beta}(t); \boldsymbol{\beta}\right\}}{\partial \boldsymbol{\beta}} \\ A_i(t)\mathbf{X}_i^{\psi}(t) \end{bmatrix} \left[Y_i(t) - f\left\{\mathbf{X}_i^{\beta}(t); \boldsymbol{\beta}\right\} - A_i(t)\boldsymbol{\psi}^T \mathbf{X}_i^{\psi}(t) \right] dN_i(t) = \mathbf{0}.$$

We can compute the asymptotic variance using theory on two-step estimators (Newey et McFadden, 1994).

Adaptive treatment strategies

The (doubly-robust) estimating equation to solve:

$$U(\boldsymbol{\beta}, \boldsymbol{\psi}; \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\omega}}) = \sum_{i=1}^{n} \int_0^{\tau} \varphi_i(\hat{\boldsymbol{\gamma}}, \mathbf{V_i(t)}) e_i(\hat{\boldsymbol{\omega}}, \mathbf{K_i(t)})$$

$$\times \begin{bmatrix} \frac{\partial f\left\{\mathbf{X}_i^{\beta}(t); \boldsymbol{\beta}\right\}}{\partial \boldsymbol{\beta}} \\ A_i(t)\mathbf{X}_i^{\psi}(t) \end{bmatrix} \left[Y_i(t) - f\left\{\mathbf{X}_i^{\beta}(t); \boldsymbol{\beta}\right\} - A_i(t)\boldsymbol{\psi}^T \mathbf{X}_i^{\psi}(t) \right] dN_i(t) = \mathbf{0}.$$

We can compute the asymptotic variance using theory on two-step estimators (Newey et McFadden, 1994).

Note that it requires an acute treatment effect assumption, no carryover effect of subsequent treatments (Dong, 2020), and it is doubly robust if the weights satisfy the balancing property and the blip is correctly specified (Wallace and Moodie, 2015).

Adaptive treatment strategies

The (doubly-robust) estimating equation to solve:

$$U(\boldsymbol{\beta}, \boldsymbol{\psi}; \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\omega}}) = \sum_{i=1}^{n} \int_0^{\tau} \varphi_i(\hat{\boldsymbol{\gamma}}, \mathbf{V_i(t)}) e_i(\hat{\boldsymbol{\omega}}, \mathbf{K_i(t)})$$

$$\times \begin{bmatrix} \frac{\partial f\left\{\mathbf{X}_i^{\beta}(t); \boldsymbol{\beta}\right\}}{\partial \boldsymbol{\beta}} \\ A_i(t)\mathbf{X}_i^{\psi}(t) \end{bmatrix} \left[Y_i(t) - f\left\{\mathbf{X}_i^{\beta}(t); \boldsymbol{\beta}\right\} - A_i(t)\boldsymbol{\psi}^T \mathbf{X}_i^{\psi}(t) \right] dN_i(t) = \mathbf{0}.$$

We can compute the asymptotic variance using theory on two-step estimators (Newey et McFadden, 1994).

Note that it requires an acute treatment effect assumption, no carryover effect of subsequent treatments (Dong, 2020), and it is doubly robust if the weights satisfy the balancing property and the blip is correctly specified (Wallace and Moodie, 2015). **The observation model must be correctly specified** for consistency.

In simulation studies...

Table 1: Results of the simulation study ($M = 1000$ simulations) - error rate for the optimal treatment decision

Sample size	Parameters γ^{υ}	No. obs. times mean (IQR)	Error rate $\hat{\psi}_{DW1}$	$\hat{\psi}_{DW2}$	$\hat{\psi}_{DW3}$	$\hat{\psi}_{DW4}$	$\hat{\psi}_{OLS}$	$\hat{\psi}_{IPT}$
250	1	3 (1-3)	0.02	0.01	0.01	0.04	0.03	0.04
	2	3 (2-5)	0.05	0.06	0.05	0.16	0.15	0.16
	3	6 (3-9)	0.06	0.03	0.03	0.26	0.25	0.26
	4	10 (8-12)	0.01	0.01	0.00	0.01	0.00	0.01
500	1	3 (1-3)	0.01	0.01	0.01	0.03	0.03	0.03
	2	3 (1-5)	0.02	0.03	0.02	0.14	0.13	0.14
	3	6 (3-9)	0.04	0.02	0.02	0.25	0.25	0.25
	4	10 (8-12)	0.00	0.00	0.00	0.00	0.00	0.00

υ.1. (-2, -0.3, 0.2, -1.2); 2. (0.3, -0.6, -0.4, -0.3); 3. (0.4, -0.8, 1, 0.6); 4. (0, 0, 0, 0). Abbrev.: IQR, interquartile range.

Similar results for the comparison of the blip coefficients and the blip evaluated at patients' characteristics.

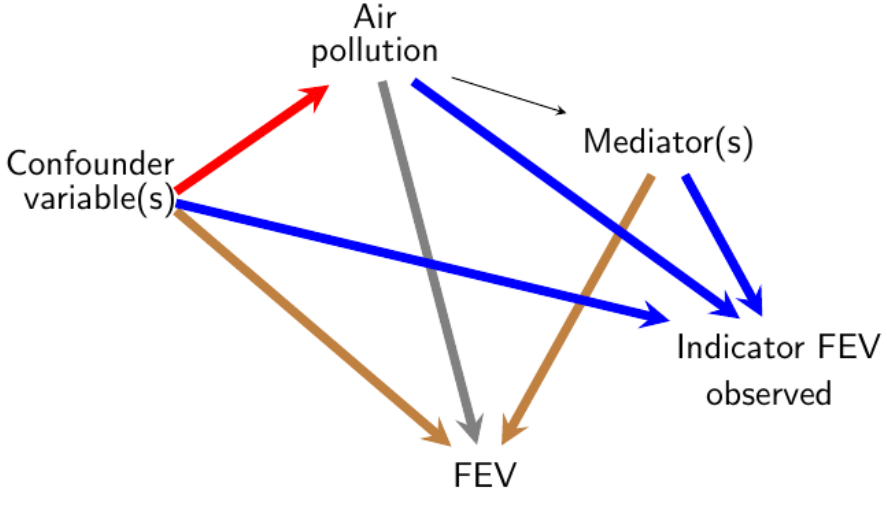
Two weaknesses of all the estimators previously discussed

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

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:

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



Novel estimator

It solves the following estimating equations:

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

where

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

(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 $\mathbf{V}(t)$)

Relative asymptotic efficiency

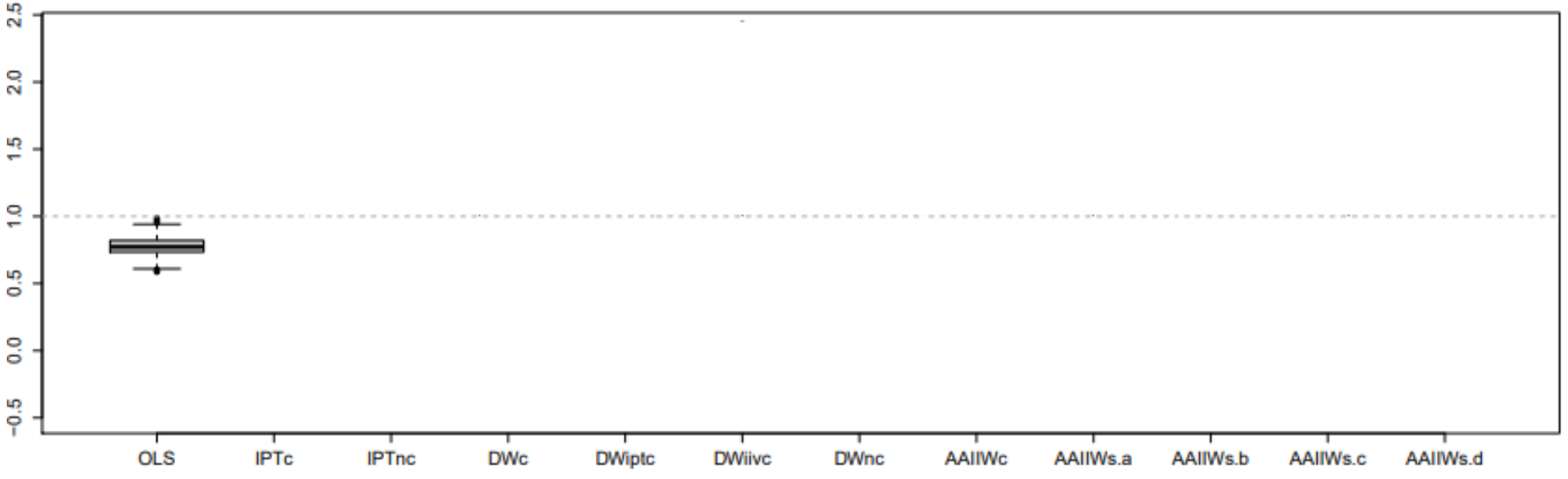
Under all nuisance models correctly specified, the AAIIW attains the semiparametric efficiency bound, and we also found

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

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

Simulation studies, sample size 1000

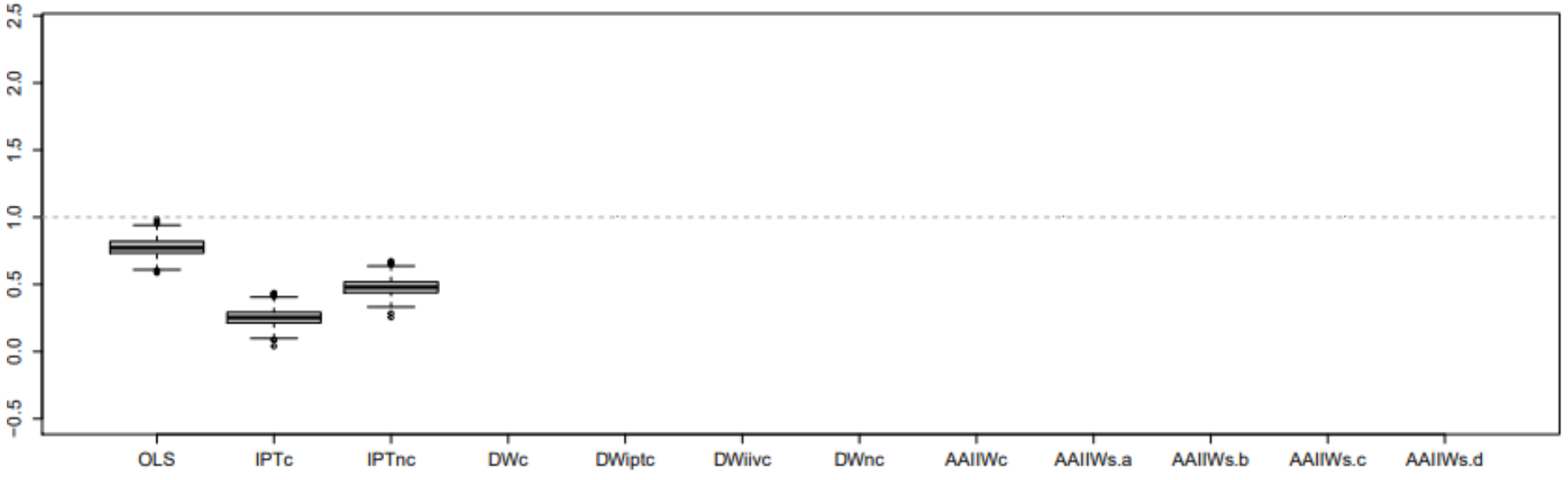
Distribution of 500 estimates:



OLS: Not adjusted for confounding nor irregular observation

Simulation studies, sample size 1000

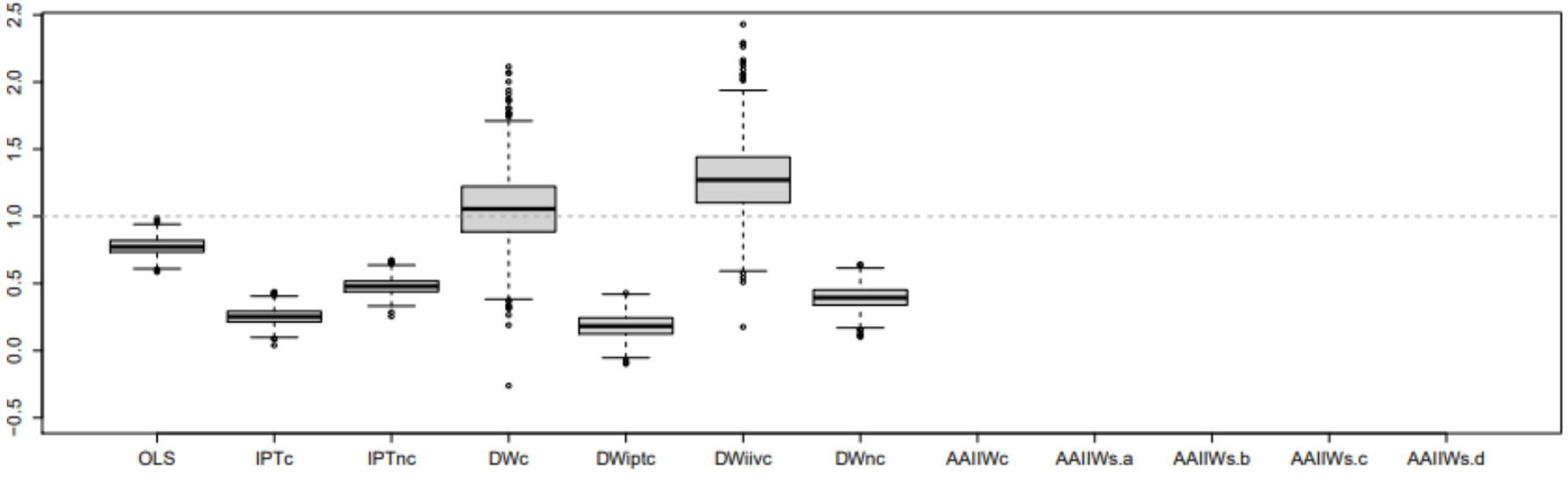
Distribution of 500 estimates:



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

Simulation studies, sample size 1000

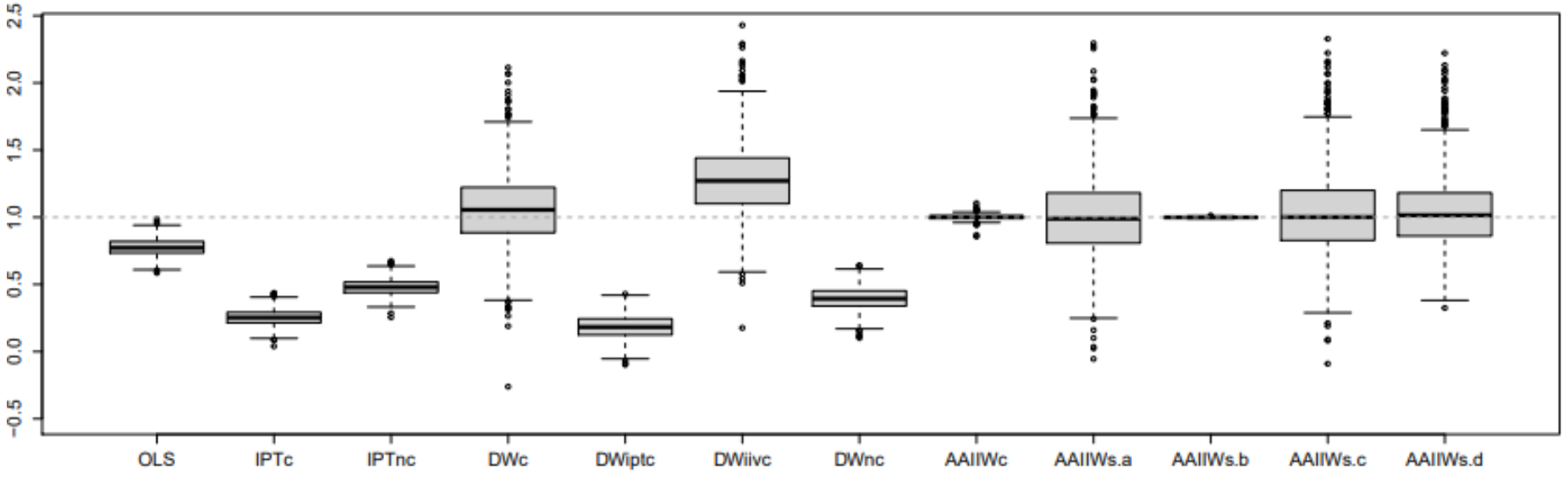
Distribution of 500 estimates:



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†

We could compare to the estimates using complete data (around 0.31)

MNAR: Method of Pullenayegum et al.

Recently, Pullenayegum et al. (2023) published a new approach for causal inference under irregular observation times for the settings in which visits occur not at random (data MNAR).

The approach relies on random effects to jointly model the outcome $Y(t)$ and the observation $dN(t)$ processes (conditional effects rather than marginal).

Their estimator performs better than our doubly-weighted estimator when visits are MNAR.

See their paper in *Statistics in Medicine* for more details.

Clusters: Method of Debray et al.

Multiple imputations using random effects in the outcome model can be used to account for hierarchies in the data.

Their method assumes hospital and patient hierarchies in repeated measurements.

See their paper in *Statistical Methods in Medical Research* (2023) for more details.

Shortreed et al. (2014) also previously proposed an imputation approach for longitudinal outcomes, for SMART studies (sequential multiple assignment randomized trials), which also depend on causal effects.

Some details to further study:

- Joint models: Random effects in causal inference
- Imputation:
  - How many time points when we impute the longitudinal outcome?
  - Imputation: Does it answer the same question as IIV-weighting?
  - Imputation: Depending on whether we know more about the outcome (biological) mechanism or its observation mechanism, we should choose one of the two types of approaches...

Discussion and future avenues

Assumptions we've made

- Causal and modeling assumptions
- Acute treatment effect
- Availability of other longitudinal processes than the outcome

Challenges left

- Cumulative effect of treatment or non-acute effects
- Dynamic treatment regimes with varying gap times
- New techniques for when predictors of the observation times are not always available

Why is it important?

Selection bias is well known, but it is typically considered in cross-sectional studies only (or only at the level of a baseline covariate or a one-time measured covariate on which there is selection).

There is often an inherent condition on the observed data, including in longitudinal studies.

How does it affect estimation?

It requires some deep thoughts about the relationships in the data and the question of interest.

References I

Andersen, Per Kragh and Richard D Gill (1982). "Cox's regression model for counting processes: a large sample study". In: The Annals of Statistics, pp. 1100–1120.

Buzkova, Petra and Thomas Lumley (2005). "Semiparametric loglinear regression for longitudinal measurements subject to irregular, biased follow-up". In.

Coulombe, Janie, Erica EM Moodie, and Robert W Platt (2021a). "Estimating the marginal effect of a continuous exposure on an ordinal outcome using data subject to covariate-driven treatment and visit processes". In: Statistics in Medicine 40.26, pp. 5746–5764.

— (2021b). "Weighted regression analysis to correct for informative monitoring times and confounders in longitudinal studies". In: Biometrics 77.1, pp. 162–174.

Coulombe, Janie and Shu Yang (2023). In: https://arxiv.org/abs/2304.08987.

Coulombe, Janie et al. (2022). "Estimation of the marginal effect of antidepressants on body mass index under confounding and endogenous covariate-driven monitoring times". In: Annals of Applied Statistics 16, pp. 1868–1890.

Coulombe, Janie et al. (2023). "Estimating individualized treatment rules in longitudinal studies with covariate-driven observation times". In: Statistical Methods in Medical Research 32.5, pp. 868–884.

References II

Debray, Thomas PA et al. (2023). "Methods for comparative effectiveness based on time to confirmed disability progression with irregular observations in multiple sclerosis". In: Statistical Methods in Medical Research, p. 09622802231172032.

Dong, Larry (2020). Evaluating the Use of Generalized Dynamic Weighted Ordinary Least Squares for Individualize McGill University (Canada).

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.

McCullagh, Peter (1980). "Regression models for ordinal data". In: Journal of the Royal Statistical Society: Series B (Methodological) 42.2, pp. 109–127.

References III

Newey, KW and D McFadden (1994). "Large sample estimation and hypothesis". In: Handbook of Econometrics 4, pp. 2112–2245.

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.

Pullenayegum, Eleanor and Maintainer Eleanor Pullenayegum (2022). "Package 'IrregLong'". In.

Pullenayegum, Eleanor M et al. (2023). "Causal inference with longitudinal data subject to irregular assessment times". In: Statistics in Medicine.

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.

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.

Shortreed, Susan M et al. (2014). "A multiple imputation strategy for sequential multiple assignment randomized trials". In: Statistics in medicine 33.24, pp. 4202–4214.

References IV

Wallace, Michael P and Erica EM Moodie (2015). "Doubly-robust dynamic treatment regimen estimation via weighted least squares". In: Biometrics 71.3, pp. 636–644.