Developing adaptive treatment strategies under irregular measurement times of the outcome: Choosing an optimal antidepressant drug for patients with depression in data from KPW

Janie Coulombe
Université de Montréal
janie.coulombe@umontreal.ca

Joint work with Dr. Erica E.M. Moodie and Dr. Susan M. Shortreed

Centre for Clinical Epidemiology's Seminar Series,
Jewish General Hospital
October, 24, 2023

Outline

Introduction and causal estimands

KPW data and patient health questionnaires

Irregular observation times: What's the problem?

Proposed approach

Diagnostics for missing data and results

Introduction and causal estimands

Adaptive treatment strategies

- Average treatment effect for the choice of an antidepressant drug

Adaptive treatment strategies

- Average treatment effect for the choice of an antidepressant drug
- Question: Can we do more than prescribe the best treatment marginally?

Adaptive treatment strategies

- Average treatment effect for the choice of an antidepressant drug
- Question: Can we do more than prescribe the best treatment marginally?
- Suppose the choice between prescribing citalopram or fluoxetine, two selective serotonin reuptake inhibitors (SSRI)

Adaptive treatment strategies

- Average treatment effect for the choice of an antidepressant drug
- Question: Can we do more than prescribe the best treatment marginally?
- Suppose the choice between prescribing citalopram or fluoxetine, two selective serotonin reuptake inhibitors (SSRI)
- E.g.: Different treatment effets in patients with/without anxiety diagnosis

Research question (2)

- Different names: Tailoring variables, effect modifiers, subgroup effects, precision medicine, adaptive treatment strategies, dynamic treatment regimes...

Research question (2)

- Different names: Tailoring variables, effect modifiers, subgroup effects, precision medicine, adaptive treatment strategies, dynamic treatment regimes...
- I will call the rule an adaptive treatment strategy (ATS)

Research question (2)

- Different names: Tailoring variables, effect modifiers, subgroup effects, precision medicine, adaptive treatment strategies, dynamic treatment regimes...
- I will call the rule an adaptive treatment strategy (ATS)
- Today I will focus on survival outcomes of the type:

  Time to 50 percent reduction in depression symptoms

  but we may be interested in different types of outcomes to optimize (survival, continuous, binary, etc.).

Estimand ψ

Suppose a binary treatment first.

Estimand ψ

Suppose a binary treatment first.

Denote by S^1 and S^0 two potential survival times under treatments $A = 1, 0$ (Neyman-Rubin, 1923 & 1974).

Estimand ψ

Suppose a binary treatment first.

Denote by S^1 and S^0 two potential survival times under treatments $A = 1, 0$ (Neyman-Rubin, 1923 & 1974).

Estimand ψ

Suppose a binary treatment first.

Denote by S^1 and S^0 two potential survival times under treatments $A = 1, 0$ (Neyman-Rubin, 1923 & 1974).

We may be interested in an average treatment effect (ATE) $\psi = \mathbb{E}[S^1 - S^0]$.

Estimand ψ

Suppose a binary treatment first.

Denote by S^1 and S^0 two potential survival times under treatments $A = 1, 0$ (Neyman-Rubin, 1923 & 1974).

We may be interested in an average treatment effect (ATE) $\psi = \mathbb{E}[S^1 - S^0]$.

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

$$\psi = \mathbb{E}[S^1 - S^0 \mid \mathbf{W} = \mathbf{w}]$$

or

$$\psi = \mathbb{E}[\log S^1 - \log S^0 \mid \mathbf{W} = \mathbf{w}]$$

for different values of covariates, $\mathbf{w}$. Under different causal assumptions, this quantity can be estimated.

(is it different from subgroup analyses?)

Previous studies and results

In data from the Clinical Practice Research Datalink (CPRD), we looked for effect modification to:

- Maximize time to a severe depression-related outcome (Coulombe et al., 2021) - the estimand:

$$\psi = \mathbb{E}[\log S^1 - \log S^0 \mid \mathbf{W} = \mathbf{w}]$$

Previous studies and results

In data from the Clinical Practice Research Datalink (CPRD), we looked for effect modification to:

- Maximize time to a severe depression-related outcome (Coulombe et al., 2021) - the estimand:

$$\psi = \mathbb{E}[\log S^1 - \log S^0 \mid \mathbf{W} = \mathbf{w}]$$

- Minimize "detrimental weight changes" (Coulombe et al., 2023) - the estimand:

$$\psi = \mathbb{E}[Y^1(t) - Y^0(t) \mid \mathbf{W} = \mathbf{w}]$$

Previous studies and results

In data from the Clinical Practice Research Datalink (CPRD), we looked for effect modification to:

- Maximize time to a severe depression-related outcome (Coulombe et al., 2021) - the estimand:

$$\psi = \mathbb{E}[\log S^1 - \log S^0 \mid \mathbf{W} = \mathbf{w}]$$

- Minimize "detrimental weight changes" (Coulombe et al., 2023) - the estimand:

$$\psi = \mathbb{E}[Y^1(t) - Y^0(t) \mid \mathbf{W} = \mathbf{w}]$$

We did not find any "good" tailoring variable for the choice of an antidepressant drug, or a class (similar results were obtained by others, see e.g., Iniesta et al. (2016), Green et al. (2017), Taliaz et al. (2021)).

DWOLS and DWSurv to develop ATS

Wallace and Moodie (2015) and Simoneau et al. (2020). Doubly-robust methods based on weighted generalized estimating equations.

DWOLS and DWSurv to develop ATS

Wallace and Moodie (2015) and Simoneau et al. (2020). Doubly-robust methods based on weighted generalized estimating equations.

- S_i the survival time for individual i,
- δ_i the indicator of experiencing the event before the end of study,
- $\mathbf{W}_i$ the potential effect modifiers,
- $\mathbf{X}_i$ some confounders or predictors of the outcome.

DWOLS and DWSurv to develop ATS

Wallace and Moodie (2015) and Simoneau et al. (2020). Doubly-robust methods based on weighted generalized estimating equations.

- S_i the survival time for individual i,
- δ_i the indicator of experiencing the event before the end of study,
- $\mathbf{W}_i$ the potential effect modifiers,
- $\mathbf{X}_i$ some confounders or predictors of the outcome.

Suppose the outcome model:

$$\mathbb{E}[\log S_i | A_i, \mathbf{X}_i, \mathbf{W}_i] = f\{\mathbf{X}_i; \boldsymbol{\beta}\} + A_i \boldsymbol{\psi}' \mathbf{W}_i$$

with the "treatment-free model" $f\{\mathbf{X}_i; \boldsymbol{\beta}\}$ and the "blip" $\boldsymbol{\psi}' \mathbf{W}_i$.

Under that model...

Model:

$$\mathbb{E}[\log S_i | A_i, \mathbf{X}_i, \mathbf{W}_i] = f\{\mathbf{X}_i; \boldsymbol{\beta}\} + A_i \boldsymbol{\psi}' \mathbf{W}_i$$

E.g.: If a patient has

$$\mathbf{W}_i = [\text{age} \;\; \text{sex} \;\; \text{Diabetes} \;\; \text{SES}] = [34 \;\; 1 \;\; 1 \;\; 3]$$

and we estimate with the regression model $\hat{\boldsymbol{\psi}} = [0.3 \;\; 0.2 \;\; 0.03 \;\; 0.001]$ then, clearly, $\hat{\boldsymbol{\psi}}' \mathbf{W}_i > 0$ for that patient.

If we want to maximize $\mathbb{E}[\log S_i | A_i, \mathbf{X}_i, \mathbf{W}_i]$ for that patient, therefore, we need to prescribe them $A_i = 1$ (as opposed to $A_i = 0$).

In the case of a one-stage (or one-time) treatment rule for a decision at cohort entry, the regression procedure corresponds to solving:

$$U(\boldsymbol{\beta}, \boldsymbol{\psi}) = \sum_{i=1}^{n} \int_0^{\tau} \rho_i \begin{bmatrix} \frac{\partial f\{\mathbf{X}_i;\boldsymbol{\beta}\}}{\partial \boldsymbol{\beta}} \\ A_i(t)\mathbf{W}_i \end{bmatrix} \left[\log(S_i) - f\left\{\mathbf{X}_i;\boldsymbol{\beta}\right\} - A_i(t)\boldsymbol{\psi}'\mathbf{W}_i\right] \delta_i = 0.$$

AIt merely corresponds to a doubly-weighted ordinary least squares estimator with a log-time outcome.

In the case of a one-stage (or one-time) treatment rule for a decision at cohort entry, the regression procedure corresponds to solving:

$$U(\boldsymbol{\beta}, \boldsymbol{\psi}) = \sum_{i=1}^{n} \int_0^{\tau} \rho_i \begin{bmatrix} \frac{\partial f\{\mathbf{X}_i;\boldsymbol{\beta}\}}{\partial \boldsymbol{\beta}} \\ A_i(t)\mathbf{W}_i \end{bmatrix} [\log(S_i) - f\{\mathbf{X}_i;\boldsymbol{\beta}\} - A_i(t)\boldsymbol{\psi}'\mathbf{W}_i]\, \delta_i = 0.$$

AIt merely corresponds to a doubly-weighted ordinary least squares estimator with a log-time outcome.

The weights $\rho_i = \rho_{i1}\rho_{i2}$ address confounding and informative censoring.

Simoneau et al. proposed the use of inverse probability of censoring weights to account for informative censoring multiplied by overlap weights (Li et al., 2018).

KPW data and patient health questionnaires

Kaiser Permanente Washington

- Do we find similar results in American data? (i.e., no effect modification)
- Access to data from **Kaiser Permanente Washington** (KPW)
- Large nonprofit healthcare organization in Washington State that has a virtual data warehouse combining electronic health records data and insurance billing information

Kaiser Permanente Washington

- Do we find similar results in American data? (i.e., no effect modification)
- Access to data from **Kaiser Permanente Washington** (KPW)
- Large nonprofit healthcare organization in Washington State that has a virtual data warehouse combining electronic health records data and insurance billing information
- We gathered a cohort of patients (13 years or older) who received a diagnosis for depression and initiated antidepressant drugs between 2008-2018

Kaiser Permanente Washington

- Do we find similar results in American data? (i.e., no effect modification)
- Access to data from **Kaiser Permanente Washington** (KPW)
- Large nonprofit healthcare organization in Washington State that has a virtual data warehouse combining electronic health records data and insurance billing information
- We gathered a cohort of patients (13 years or older) who received a diagnosis for depression and initiated antidepressant drugs between 2008-2018
- All patients had a one-year retrospective inspection period (and 59% had less than 5 years inspection period)

Kaiser Permanente Washington

- Do we find similar results in American data? (i.e., no effect modification)
- Access to data from **Kaiser Permanente Washington** (KPW)
- Large nonprofit healthcare organization in Washington State that has a virtual data warehouse combining electronic health records data and insurance billing information
- We gathered a cohort of patients (13 years or older) who received a diagnosis for depression and initiated antidepressant drugs between 2008-2018
- All patients had a one-year retrospective inspection period (and 59% had less than 5 years inspection period)
- In KPW data, as opposed to the CPRD, we have access to patient health questionnaires (PHQ)

PHQ-9 questionnaire (a validated tool)

(Kroenke et al., 2001)

- The DSM 5th version (Diagnostic and Statistical Manual) is a classification tool for common mental disorders
- The PHQ-9 scores each of the 9 depression-related items from the DSM from 0 (not at all) to 3 (nearly every day)
- Ex. of question: “Little interest or pleasure in doing things”, “Feeling tired or having little energy”

PHQ-9 questionnaire (a validated tool)

(Kroenke et al., 2001)

- The DSM 5th version (Diagnostic and Statistical Manual) is a classification tool for common mental disorders
- The PHQ-9 scores each of the 9 depression-related items from the DSM from 0 (not at all) to 3 (nearly every day)
- Ex. of question: "Little interest or pleasure in doing things", "Feeling tired or having little energy"
- We can separate:
  - PHQ first 8 items (**PHQ-8**) ranging from 0 to 24, and
  - **PHQ-9i**, the 9th item, which focuses on suicidal ideation or self-harm, ranging from 0 to 3

Challenge in KPW

Time to 50 percent reduction in depression symptoms

becomes

Time to 50 percent reduction in PHQ-9 score (in months)

Challenge in KPW

Time to 50 percent reduction in depression symptoms

becomes

Time to 50 percent reduction in PHQ-9 score (in months)

The PHQ-9 score is supposed to be measured rather regularly (at cohort entry, day 30, 60, 90, 180, 270 and 365 as well as at second medication initiation, initiation of antipsychotics, or initiation of psychotherapy, if there is).

Challenge in KPW

Time to 50 percent reduction in depression symptoms

becomes

Time to 50 percent reduction in PHQ-9 score (in months)

The PHQ-9 score is supposed to be measured rather regularly (at cohort entry, day 30, 60, 90, 180, 270 and 365 as well as at second medication initiation, initiation of antipsychotics, or initiation of psychotherapy, if there is).

In the data, the observation schedule does not reflect that.

Irregular observation times: What's the problem?

The issue with irregular observation times

- In previous work (Coulombe et al, 2020-2023), we demonstrated the bias induced by irregular observation times.

The issue with irregular observation times

- In previous work (Coulombe et al, 2020-2023), we demonstrated the bias induced by irregular observation times. **It is similar to longitudinal selection bias**.
- Patients are not observed as frequently, and not at the same times. Often, these times depend on their health condition and other demographics.
- Patient characteristics could **simultaneously affect** the outcome value and the likelihood of being observed.
- E.g.: When the visit is a collider on a path relating the exposure and the outcome, conditioning on observed data leads to spurious associations (not due to causal effects).

Choice of statistical method

- We previously tackled this issue simultaneously with confounding in settings with an irregularly observed continuous outcome.

Choice of statistical method

- We previously tackled this issue simultaneously with confounding in settings with an irregularly observed continuous outcome.
- We used inverse intensity-of-visit weights (Lin et al., 2004) to account for irregular observation times (these weights are based on a proportional rate model for the visits).

Choice of statistical method

- We previously tackled this issue simultaneously with confounding in settings with an irregularly observed continuous outcome.
- We used inverse intensity-of-visit weights (Lin et al., 2004) to account for irregular observation times (these weights are based on a proportional rate model for the visits).
- In KPW, an **imputation approach** might be preferable to inverse weights.

Choice of statistical method

- We previously tackled this issue simultaneously with confounding in settings with an irregularly observed continuous outcome.
- We used inverse intensity-of-visit weights (Lin et al., 2004) to account for irregular observation times (these weights are based on a proportional rate model for the visits).
- In KPW, an **imputation approach** might be preferable to inverse weights.
- We are interested in a survival outcome and there is no easy way to re-weight observation times, to account for the measurement error in the survival outcome due to irregular observation

  (indeed, inverse-weighting is great for inference on population quantities, but less for "predicting" the outcome at a specific time point)

Proposed approach

Imputation approach

- Inspired by Shortreed et al. (2014), we use a sequential imputation approach for monthly PHQ-8, PHQ-9i and weight outcomes.

Imputation approach

- Inspired by Shortreed et al. (2014), we use a sequential imputation approach for monthly PHQ-8, PHQ-9i and weight outcomes.
- Idea: to borrow information from previous months and baseline to impute future PHQ-8, PHQ-9 and weight values.
- Shortreed et al. call this: "Time-ordered nested conditional imputation approach".

  (what kind of assumptions are we making?)

Imputation approach

- Inspired by Shortreed et al. (2014), we use a sequential imputation approach for monthly PHQ-8, PHQ-9i and weight outcomes.
- Idea: to borrow information from previous months and baseline to impute future PHQ-8, PHQ-9 and weight values.
- Shortreed et al. call this: "Time-ordered nested conditional imputation approach".

  (what kind of assumptions are we making?)
- We do not focus on a particular 2-drug comparison at first.

Imputation approach

- Inspired by Shortreed et al. (2014), we use a sequential imputation approach for monthly PHQ-8, PHQ-9i and weight outcomes.
- Idea: to borrow information from previous months and baseline to impute future PHQ-8, PHQ-9 and weight values.
- Shortreed et al. call this: "Time-ordered nested conditional imputation approach".

  (what kind of assumptions are we making?)
- We do not focus on a particular 2-drug comparison at first.
- We impute values each month, only for the first year of follow-up (12 months).

Data

- Cohort of around 90,000 patients
- Creation of monthly variables: $PHQ8_m$, $PHQ9i_m$, weight_m, $m = 1, ..., 12$

Data

- Cohort of around 90,000 patients
- Creation of monthly variables: $PHQ8_m$, $PHQ9i_m$, weight_m, $m = 1, ..., 12$
- Other variables available:
  - demographics (age, sex, race and ethnicity, insurance type, etc.);
  - medication and treatment (psychotherapy, antipsychotics, etc.); and
  - outcomes (PHQ-8, PHQ-9i, suicide attempt, self-harm, weight, etc.).
- Other longitudinal variables can be transformed into indicators (such as monthly indicators of initiating psychotherapy during month $j, j = 1, ..., 12$)

Time-varying variables we created:

- 1) Initiation and 2) continuation of SGA or FGA
- 1) Initiation and 2) continuation of psychotherapy
- End of the initiating treatment
- Adding a second medication during that month
- Ending the second medication
- PHQ-8 and PHQ-9i measurements closest in time, measured before the month day minus 10 days
- Indicator of self-harm diagnosis, death, death by suicide, hospitalization for depression
- Psychiatric diagnoses: Autism spectrum disorder, anxiety, PTSD, schizophrenia, other psychosis, bipolar disorder, OCD, opioid use disorder, personality disorder, sedative use disorder that occurred anytime before
- Indicator of at least one psychiatric contact on a given month

Imputation

Sequential approach with multiple imputations with chained equations (MICE) using fully conditional specifications:

- Impute baseline variables first
- Impute month 1 data using the baseline information and time-varying variables measured at month 1
- Impute month j data ($j = 2, ..., 12$) using the baseline information + PHQ-8, PHQ-9i and weight imputed at month $j - 1$ and time-varying variables from month j

We created 25 such imputed datasets.

Other important details

- PHQ-8 and PHQ-9i provided different information and were treated separately

Other important details

- PHQ-8 and PHQ-9i provided different information and were treated separately
- Ultimate goal is causal inference: *what does that imply for imputation?*

Other important details

- PHQ-8 and PHQ-9i provided different information and were treated separately
- Ultimate goal is causal inference: *what does that imply for imputation?*
- Treatment of loss to follow-up, post-processing and censoring

Diagnostics for missing data and results

Missing data

- Baseline covariates: Between 0% and 24% missing values
- PHQ-9: 58 % missing at baseline, and between 86% and 99% missing between months 1 and 12
- Naturally, months 4, 5, 7, 8, 10, and 11 not corresponding to the measurement schedule of PHQ-9 contained a lot of missing data.

Diagnostics for causal inference

Suppose we focus on larger groups, which provides more power to detect effect modification.

SSRI: Selective serotonin reuptake inhibitors users (5337 patients, 2910 outcomes).

SNRI: Serotonin and norepinephrine reuptake inhibitors users (68,721 patients, 39,874 outcomes).

We can look at balance in baseline characteristics:

Variable	Before imputation			After imputation		
(% missing)	SNRI	SSRI	SMD[1]	SNRI	SSRI	SMD[1]
Weight (24)	190.58	179.73	0.20	192.06	180.32	0.22
PHQ-8 (58)	14.10	14.57	0.09	14.18	14.56	0.07
PHQ-9i (58)	0.52	0.58	0.07	0.52	0.57	0.06

Large differences here could indicate that data at baseline were missing in certain strata of a third variable (e.g., age) associated to both PHQ and the SSRI-SNRI prescription at baseline.

[1] standardized mean difference

Variable	Before imputation			After imputation		
(% missing)	SNRI	SSRI	SMD[1]	SNRI	SSRI	SMD[1]
Weight (24)	190.58	179.73	0.20	192.06	180.32	0.22
PHQ-8 (58)	14.10	14.57	0.09	14.18	14.56	0.07
PHQ-9i (58)	0.52	0.58	0.07	0.52	0.57	0.06

Large differences here could indicate that data at baseline were missing in certain strata of a third variable (e.g., age) associated to both PHQ and the SSRI-SNRI prescription at baseline.

Looking at missingness in covariates by treatment group could be a form of preliminary diagnostic.

[1] standardized mean difference

Diagnostics for causal inference (2)

Before (left) and after imputation (right):

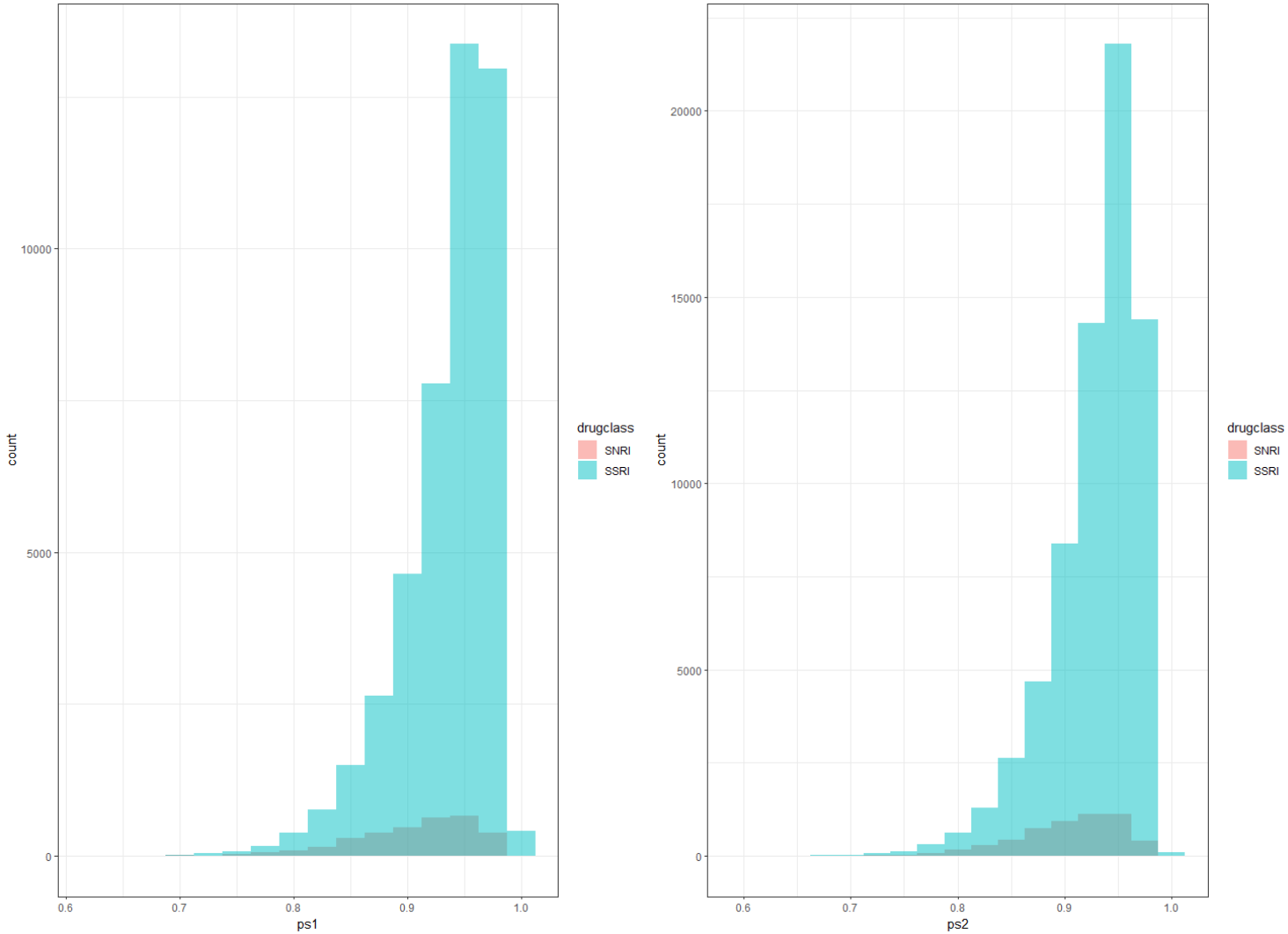


Figure 1: Overlap in the propensity score distributions. Blue: SSRI, Pink: SNRI

Tailoring variables (and confounders) for SSRI-SNRI comparison

- Age, sex, race and ethnicity, weight at cohort entry, tobacco use, Charlson comorbidity index[2]
- Psychotherapy (previous year)
- Anxiety or generalized anxiety disorder (GAD)
- Indicator of other psychiatric diagnosis at cohort entry (autism spectrum, obsessive-compulsive, bipolar, personality, sedative use, or alcohol use disorders, schizophrenia, PTSD[3])
- Number of hospit. for mental health diagnosis or suicide attempt or self-harm (previous 6 months)
- Number of antidepressant drugs in previous 5 years
- Had a baseline PHQ score
- PHQ-8 and PHQ-9i at baseline

[2] categorizes comorbidities based on the risk of mortality

[3] post-traumatic stress disorder

Time to 50% PHQ reduction - SSRI vs SNRI

We applied the DWSurv approach to each of the 25 imputed datasets.

Time to 50% PHQ reduction - SSRI vs SNRI

We applied the DWSurv approach to each of the 25 imputed datasets.

Out of 25 imputed datasets (number of times the effect modifier is significant):

- Sex (3)
- Anxiety or GAD (2)
- Diagnosis for a psychiatric diagnosis other than anxiety, GAD (2)
- No. of mental health inpatient stays in previous 6 months (2)
- No. mental health visits in previous 5 years (2)
- Had baseline PHQ (2)
- Psychotherapy (1)
- No. prior AD in previous 5 years (1)

Time to 50% PHQ reduction - SSRI vs SNRI

We applied the DWSurv approach to each of the 25 imputed datasets.

Out of 25 imputed datasets (number of times the effect modifier is significant):

- Sex (3)
- Anxiety or GAD (2)
- Diagnosis for a psychiatric diagnosis other than anxiety, GAD (2)
- No. of mental health inpatient stays in previous 6 months (2)
- No. mental health visits in previous 5 years (2)
- Had baseline PHQ (2)
- Psychotherapy (1)
- No. prior AD in previous 5 years (1)

The most "important" effect modifier is only significant 3 times out of 25...

Discussion

- Similar results as with CPRD (we also looked at more specific comparisons, like citalopram-fluoxetine)

Discussion

- Similar results as with CPRD (we also looked at more specific comparisons, like citalopram-fluoxetine)
- Longitudinal imputation: Lots of missing values in time (although high correlation between repeated measurements)

Discussion

- Similar results as with CPRD (we also looked at more specific comparisons, like citalopram-fluoxetine)
- Longitudinal imputation: Lots of missing values in time (although high correlation between repeated measurements)
- Results depend on the time discretization we used for imputation

Discussion

- Similar results as with CPRD (we also looked at more specific comparisons, like citalopram-fluoxetine)
- Longitudinal imputation: Lots of missing values in time (although high correlation between repeated measurements)
- Results depend on the time discretization we used for imputation
- Could have used a "fancier" imputation approach, forcing smoothness in mean and interactions in the models (principle of congeniality) or random effects (see Debray et al., 2023)

Discussion

- Similar results as with CPRD (we also looked at more specific comparisons, like citalopram-fluoxetine)
- Longitudinal imputation: Lots of missing values in time (although high correlation between repeated measurements)
- Results depend on the time discretization we used for imputation
- Could have used a "fancier" imputation approach, forcing smoothness in mean and interactions in the models (principle of congeniality) or random effects (see Debray et al., 2023)
- What is preferable between inverse-weighting methods and imputation here?

Discussion

- Similar results as with CPRD (we also looked at more specific comparisons, like citalopram-fluoxetine)
- Longitudinal imputation: Lots of missing values in time (although high correlation between repeated measurements)
- Results depend on the time discretization we used for imputation
- Could have used a "fancier" imputation approach, forcing smoothness in mean and interactions in the models (principle of congeniality) or random effects (see Debray et al., 2023)
- What is preferable between inverse-weighting methods and imputation here? Depends on our assumptions! (+ causal assumptions)

Acknowledgements

Drs. Erica E. M. Moodie
and Susan M. Shortreed

and Dr. Christel Renoux for the
invitation.

Funding and support:

NSERC CRSNG

compute canada | calcul canada

NIH
National Institute of Mental Health

Contact:
janie.coulombe@umontreal.ca

References I

Coulombe, J., Moodie, E. E. M., et Platt, R. W. (2020) Weighted regression analysis to correct for informative monitoring times and confounders in longitudinal studies. *Biometrics*, 77(1), pp. 162-174.

Coulombe, J., Moodie, E. E. M., Shortreed, et al. (2021) Can the risk of severe depression-related outcomes be reduced by tailoring the antidepressant therapy to patient characteristics?. *American Journal of Epidemiology*, 190(7), pp. 1210-1219.

Coulombe, J., Moodie, E. E. M., et Platt, R. W. (2021) Estimating the marginal effect of a continuous exposure on an ordinal outcome using data subject to covariate-driven treatment and visit processes. *Statistics in Medicine*, 40(26), pp. 5746-5764.

Coulombe, J., Moodie, E. E. M., Renoux, C., et al. (2023) Estimating individualized treatment rules in longitudinal studies with covariate-driven observation times. *Statistical Methods in Medical Research*, 32(5), pp. 868-884.

Debray, T. P., Simoneau, G., Copetti, M., Platt, R. W., Shen, C., Pellegrini, F., et de Moor, C. (2023). Methods for comparative effectiveness based on time to confirmed disability progression with irregular observations in multiple sclerosis. *Statistical Methods in Medical Research*, 32(7), pp. 1284-1299.

Green E., Goldstein-Piekarski, A., Schatzberg, A. F., et al. (2017) Personalizing antidepressant choice by sex, body mass index, and symptom profile: An iSPOT-D report. *Personalized medicine in psychiatry*, 1, pp. 65-73.

Iniesta, R., Malki, K., Maier, W., et al. (2016) Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of psychiatric research*, 78, pp. 94-102.

Kroenke K., Spitzer, R. L., and Williams, J. BW. (2001) The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9), pp. 606-613.

Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018) Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521), pp. 390-400.

References II

Lin, H., Scharfstein, D. O., et Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3), pp. 791-813.

Shortreed, S. M., Laber, E., Scott Stroup, T., et al. (2014) A multiple imputation strategy for sequential multiple assignment randomized trials. *Statistics in medicine*, 33(24), pp. 4202-4214.

Simoneau, G., Moodie, E. E. M., Nijjar, J. S., et al. (2020) Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association*, 115(531), pp. 1531-1539.

Taliaz, D., Spinrad, A., Barzilay, R., et al. (2021) Optimizing prediction of response to antidepressant medications using machine learning and integrated genetic, clinical, and demographic data. *Translational psychiatry*, 11(1), pp. 381.

Wallace, M. P. and Moodie, E. E. M. (2015) Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71(3), pp. 636-644.

Appendix

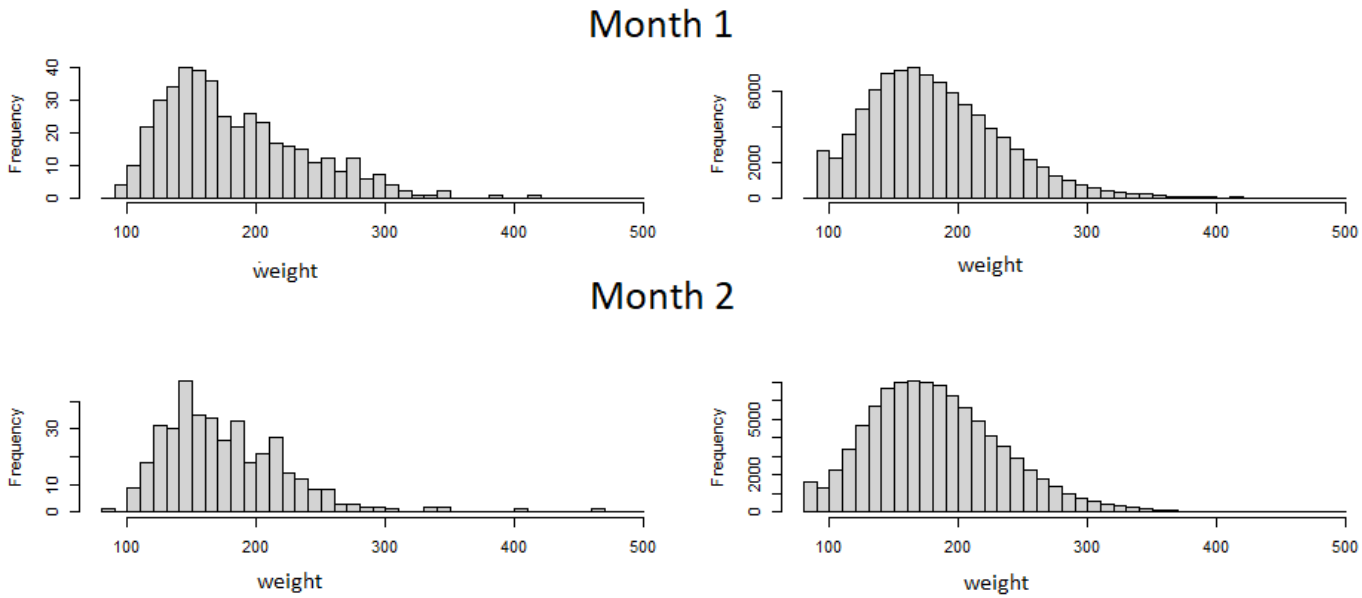
Keeping complete cases until month 3

Correlation matrix for PHQ-8 scores:

	baseline	month 1	month 2	month 3
baseline	1.00	0.50	0.45	0.38
month 1	0.50	1.00	0.68	0.60
month 2	0.45	0.68	1.00	0.74
month 3	0.38	0.60	0.74	1.00

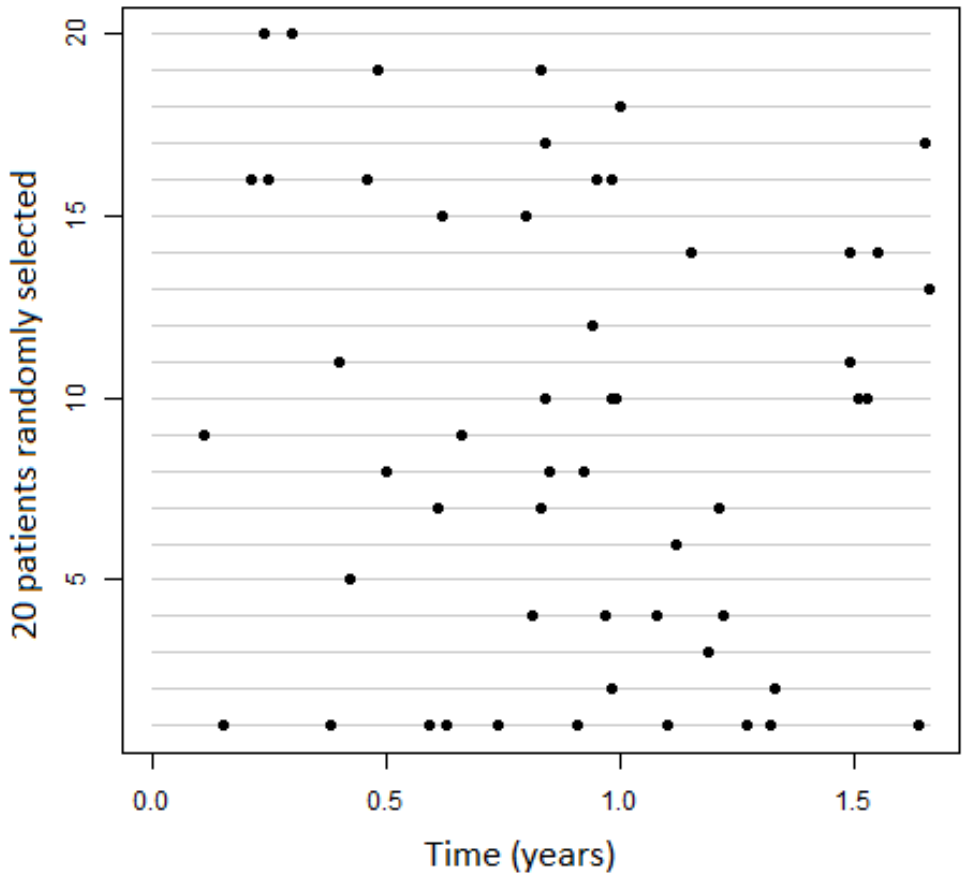
Diagnostics for the imputation

Example with weight (left, before; right, after imputation)



Setting (simulation study)

Proportional rate model for the observation times:



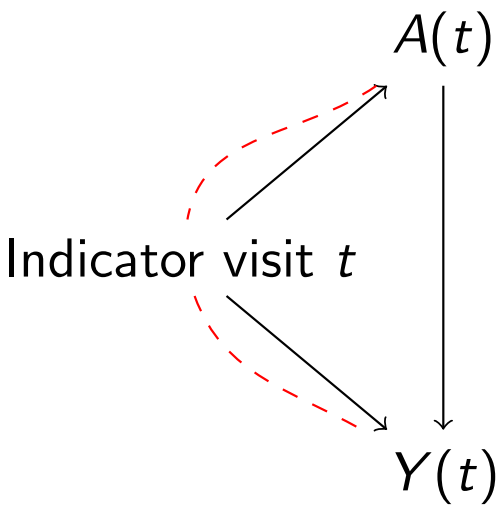
Note. This plot is called an abacus plot and can be produced with the `IrregLong` package proposed by Pullenayegum (2022).

Why are irregular measurements problematic?

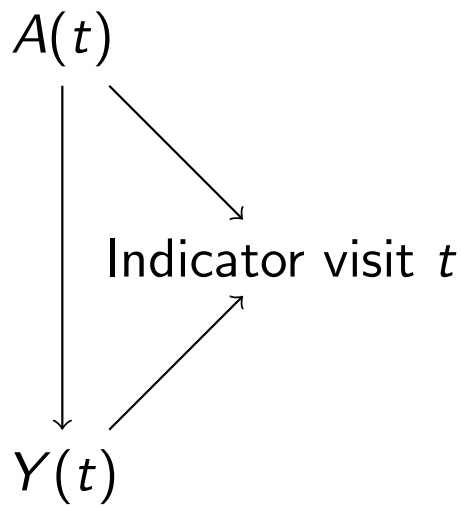
Suppose an indicator of visit at time t - it can be related to other variables in many different ways:

Why are irregular measurements problematic?

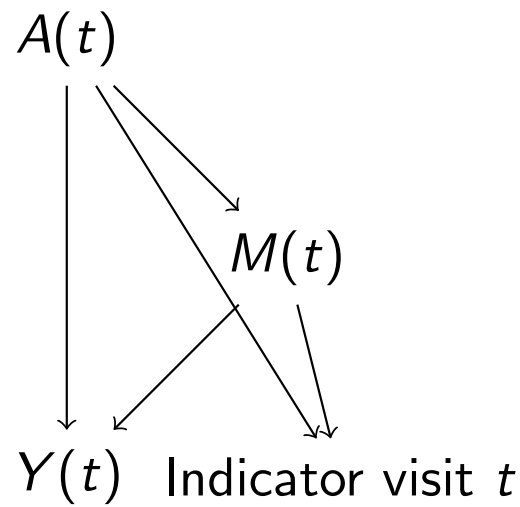
Suppose an indicator of visit at time t - it can be related to other variables in many different ways:



Confounder



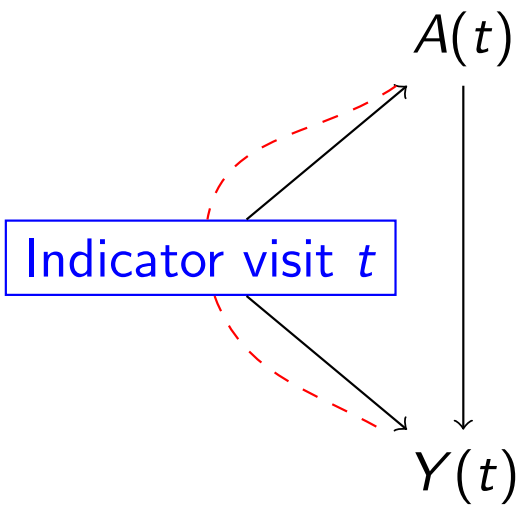
Collider



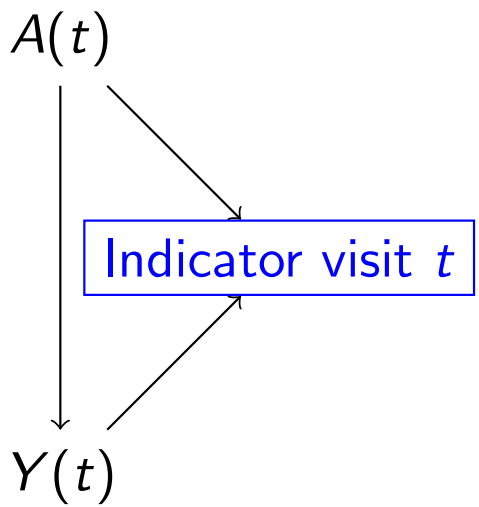
Collider

Why are irregular measurements problematic?

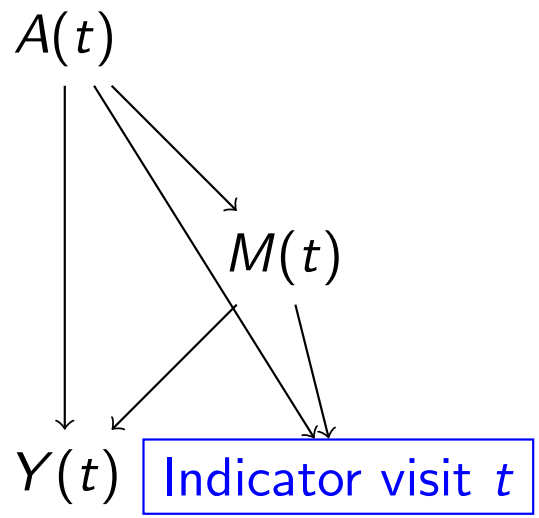
Suppose an indicator of visit at time t - it can be related to other variables in many different ways:



Confounder



Collider

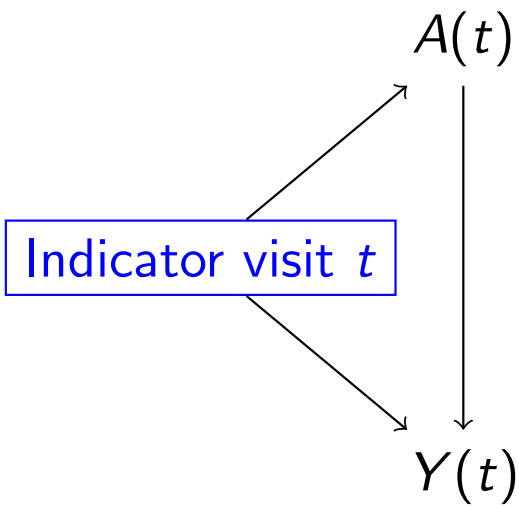


Collider

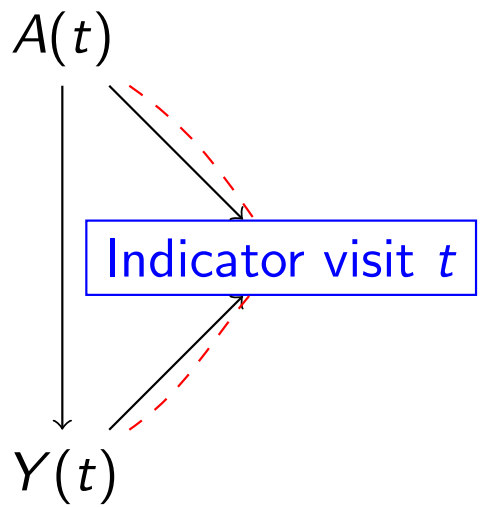
In general, in a standard analysis, there is a condition on the measured data, i.e., on the indicator of visit at time t (i.e., a box around it).

Why are irregular measurements problematic?

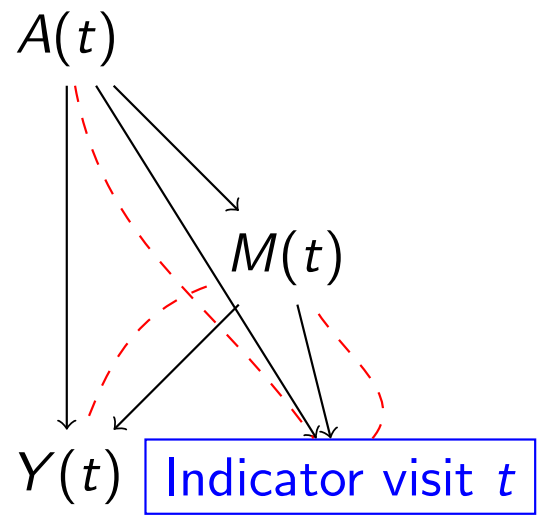
Suppose an indicator of visit at time t - it can be related to other variables in many different ways:



Confounder



Collider



Collider

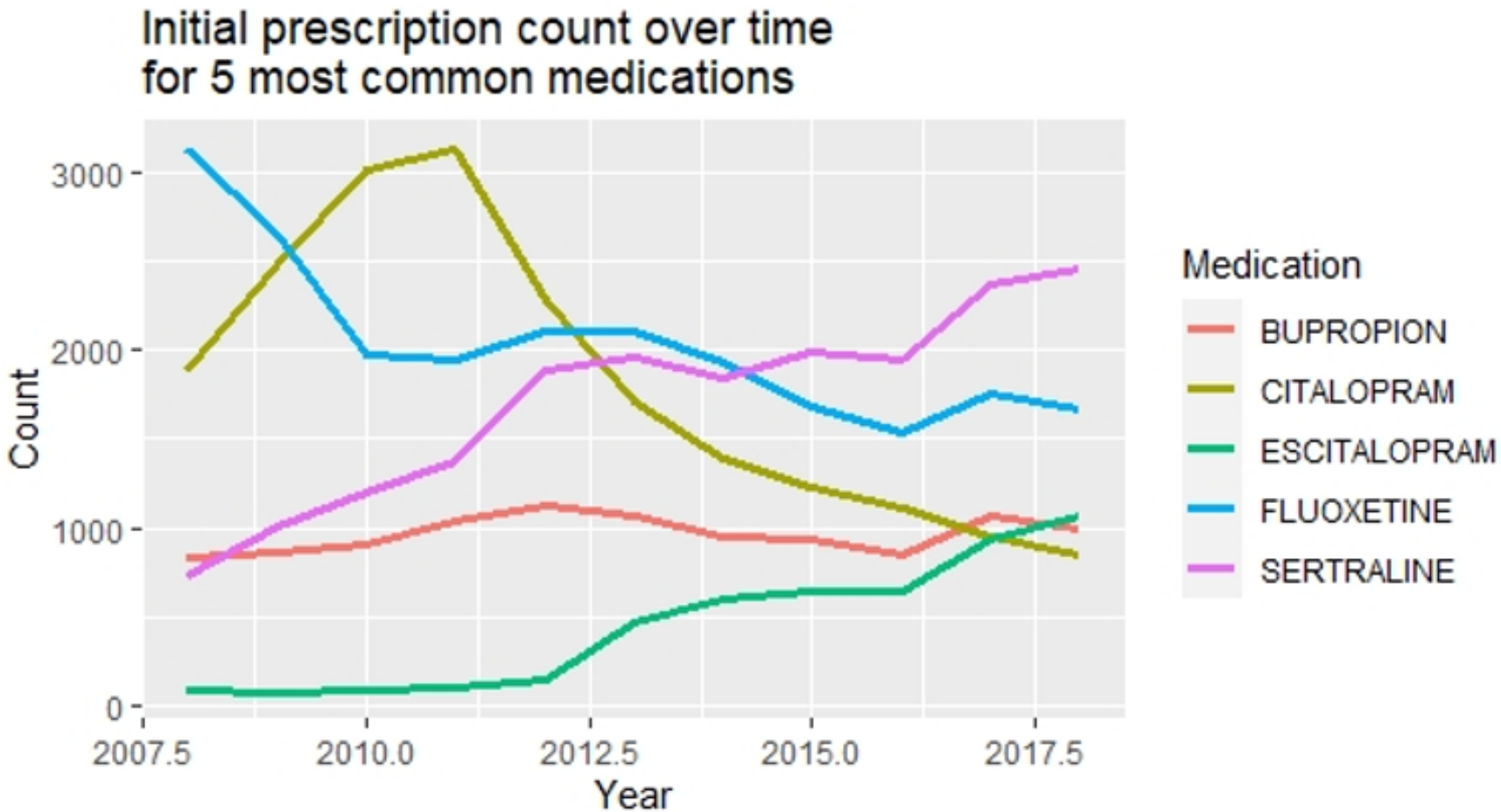
In general, in a standard analysis, there is a condition on the measured data, i.e., on the indicator of visit at time t (i.e., a box around it).

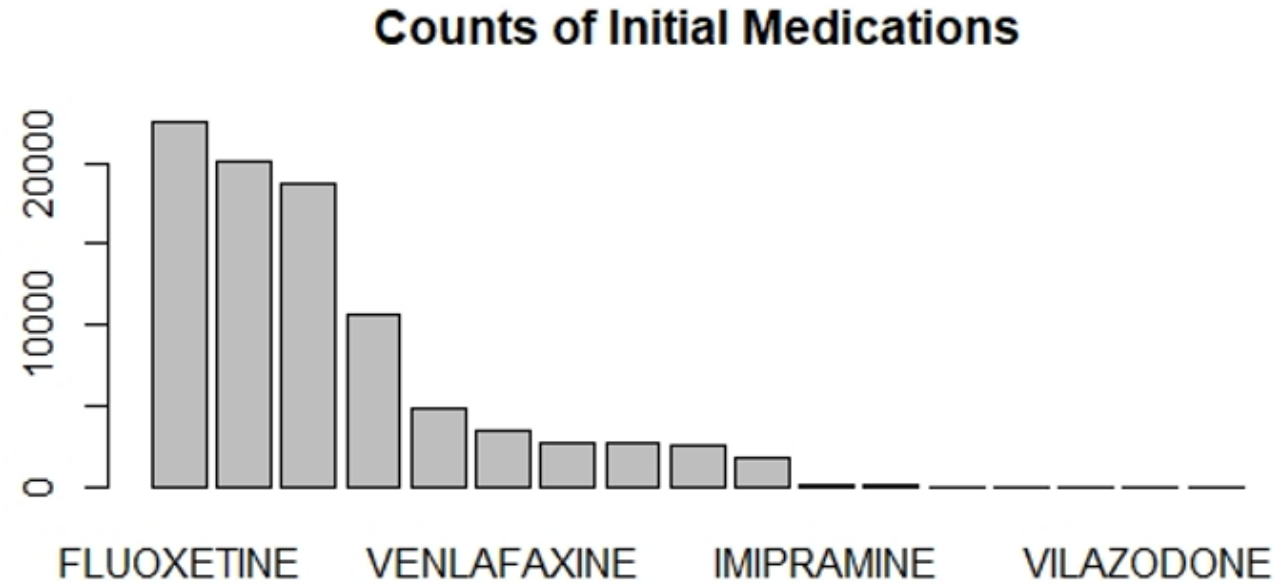
Double missingness mechanism

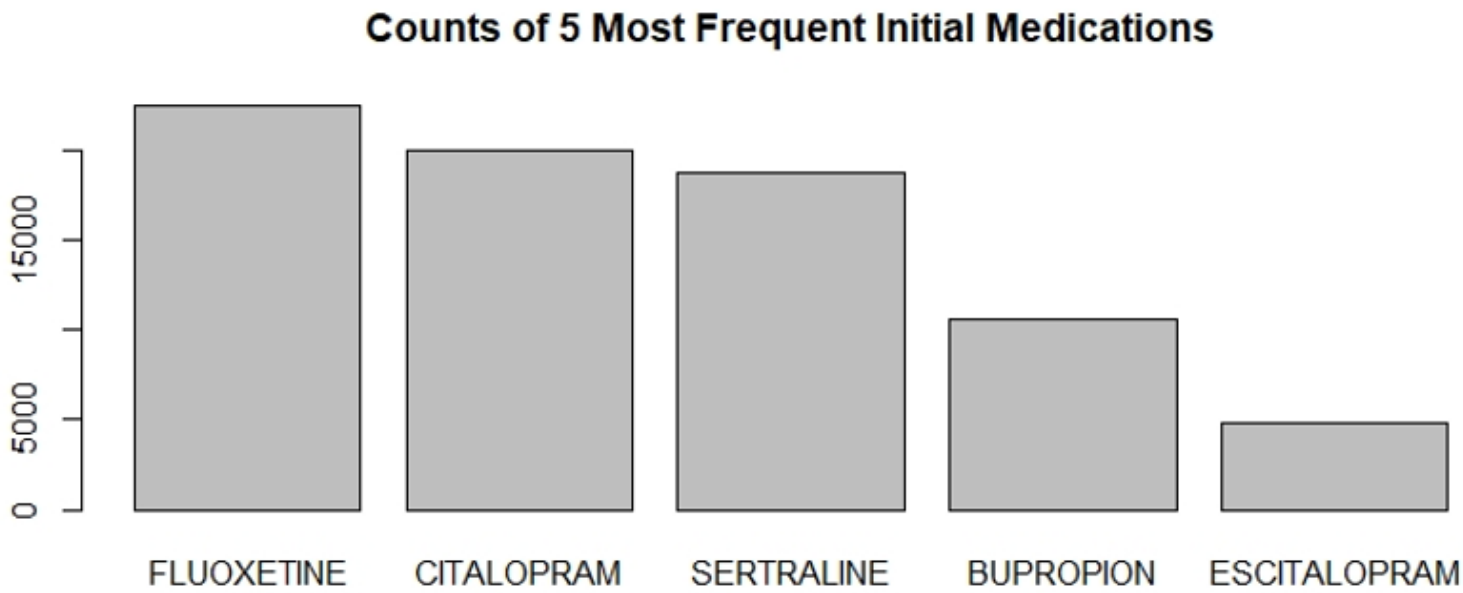
- We observe one of the two potential outcomes informatively, according to the treatment mechanism.
- We do not observe $Y^1(t)$ and $Y^0(t)$ at all times t but only at covariate-dependent measurement times.
- We previously proposed doubly-weighted estimators that tackle these missingness problems (Coulombe et al., 2020-2023).

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.
- We previously proposed doubly-weighted estimators that tackle these missingness problems (Coulombe et al., 2020-2023).
- However, in the KPW analysis, we are interested in outcomes of the type **Time-to 50% reduction in depression symptoms**.
- Recovering regularity in the observation of the PHQ outcome might be preferable to inverse-weighting strategies (connection to the causal estimand here...)


Initial prescription count over time
for 5 most common medications
Count
3000
2000
1000
0
2007.5
2010.0
2012.5
2015.0
2017.5
Year
Medication
BUPROPION
CITALOPRAM
ESCITALOPRAM
FLUOXETINE
SERTRALINE


Counts of Initial Medications
0
10000
20000
FLUOXETINE
VENLAFAXINE
IMIPRAMINE
VILAZODONE


Counts of 5 Most Frequent Initial Medications
0
5000
15000
FLUOXETINE
CITALOPRAM
SERTRALINE
BUPROPION
ESCITALOPRAM

Variable	KPWA dataset		CPRD dataset	
	Citalopram	Fluoxetine	Citalopram	Fluoxetine
All patients, No.	17,505	17,737	137,791	108,712
Age, mean (SD), y	47.6 (18.6)	46.6 (19.0)	43.4 (18.2)	40.7 (16.3)
Age group, No. (%), y				
18-24	2257 (12.9)	2952 (16.6)	20,665 (15.0)	17,950 (16.5)
25-34	2661 (15.2)	2549 (14.4)	31,382 (22.8)	27,248 (25.1)
35-44	3147 (18.0)	2884 (16.3)	29,927 (21.7)	25,564 (23.5)
45-54	3261 (18.6)	3145 (17.7)	22,309 (16.2)	17,523 (16.1)
≥55	6179 (35.3)	6207 (35.0)	33,508 (24.3)	20, 427 (18.8)
Female, No. (%)	11,796 (67.4)	11,912 (67.2)	87,561 (63.6)	71,826 (66.1)
Year of cohort entry, No. (%)				
1998-2005	0	0	40,077 (29.1)	56,502 (52.0)
2006-2011	9103 (52.0)	7873 (44.4)	65,297 (47.4)	38,780 (35.7)
2012-2017	8402 (48.0)	9864 (55.6)	32,417 (23.5)	13,430 (12.4)

Baseline variables

- Initiating treatment, the number of days of supply of the initiating treatment, the number of days enrolled prior to index date,
- an urban-rural indicator (integer from 1-most urban to 6-most rural), the number of days between the depression diagnostic and index date, sex at birth, age, type of insurance, ethnicity, calendar year at index date, an indicator of low income in the neighborhood, an indicator of low college graduation in the neighborhood, an indicator of neighborhood poverty based on the Federal poverty level,
- the Charlson comorbidity index at index date, height in inches at index date, weight in pounds at index date, the number of days between weight was measured and the index date, an indicator of alcohol use disorder in the year prior index date, of GAD, ASD, PTSD, OCD, OUD, personality disorder, SUD, of use of tobacco, all in the year prior index date,
- the no. of inpatient stays with mental health diagnostic in the 6 months prior index date, the no. of suicide attempts in the 6 months prior index date, an indicator of an inpatient stay with mental health diagnostic in the 5 years prior index date, the no. of different antidepressants taken in the 5 years prior index date, the number of mental health specialty visits in the 5 years prior index date,

- an indicator of antidepressant use in the year prior (that include only the drugs trazodone, doxepin, or amitriptyline, as patients who have used other types of antidepressant in the year prior have been excluded),
- an indicator of any psychotherapy visit in the year prior index date,
- an indicator that the baseline PHQ exists and is measured within -15 to +3 days from the index date of that episode,
- the PHQ8 score at index date, the 9th item of the PHQ score, the number of days between the time when the PHQ scores were last measured and the index date, the count of prior PHQ 9 scores that were measured in the year prior to index date and including the next 15 days after index date, the average PHQ 8 score (first 8 items) in the year prior index date, the average of the PHQ 9th item in the year prior index date, the minimum of the PHQ9 9th item in the year prior index date, the maximum of the PHQ9 9th item in the year prior index date, the episode number (as some individuals have more than one).

Missing data

- Least missing PHQ data at the time of psychotherapy initiation (39% missing)
- Weight often missing, but there is an exception in FGA initiators, only about 14% missing.
- Race, whether Hispanic, and height were all missing for about 22%-24% of episodes.
- Weight and PHQ attached to multiple events that could occur far in time; sometimes same values as baseline values. Weight: true for 40% for the end of the initial medication to 15% for starting an initial FGA. For PHQ, true for a much smaller proportion of events (less than 7 percent)
- Large decrease in missingness from 2008 to 2009 and then proportion with baseline PHQs increases until 2013 when it begins to drop off slightly.
- For many of the other variables, it not possible to determine missingness as there is no way to distinguish unrecorded from no diagnosis from no disease.

Missingness (2)

- In terms of censoring, all observations included had at least 365 days enrolled before the index date, which reduced the censoring of initial information about diagnoses and episodes.
- However, three variables looking at five years of past information (about antidepressants, mental health hospitalizations, and self-harm incidents) were included, and 59% of episodes did not have 5 years of enrollment before the index date.
- Finally, the death days, suicide flag, self-harm days, and hospital days variables could be considered censored for most episodes as for most episodes (84% - 98% depending) the event never occurred before the end of study or end of enrollment.

Below is a summary of the days till the initial medication ended and the proportion who were prescribed a follow-up medication in the multiple episode group by episode number.

Episode Number	Days before Initial Medication 1 End				Count	Proportion prescribed		
	Min	Mean	Median	Max		Second AD	Initial FGA	Initial SGA
1	3	136.39	67	2505	9443	0.296	0.017	0.011
2	3	198.93	90	3102	9443	0.339	0.024	0.022
3	12	152.54	84	2105	1215	0.314	0.026	0.023
4	4	130.99	90	1001	146	0.295	0.021	0.014
5	30	100.17	30	408	12	0.667	0.167	0.083

Episodes with and without Baseline PHQ

Looking at baseline characteristics, those without a PHQ are much more likely to have unknown race and to have missing weight and/or height information. The average age is about five years older in the group without PHQs. Those without a baseline PHQ are somewhat more likely to have lower neighborhood income and education.

Looking at the prior diagnosis history, for about half the diagnosis proportion or past event mean is higher for those with a baseline PHQ, as might be expected. One exception is the opioid use diagnosis, though the difference is small. Other exceptions that seem (informally) more potentially meaningful are the mean prior mental health inpatient episodes and mean prior AD prescriptions. Also, the proportion with PT in the prior year is higher for those with baseline PHQs, as expected.

Looking at the episode year, as expected the mean episode year is about a year greater for those episodes with baseline PHQs due to the time trend in missingness of PHQs.

Looking at differences with regards to treatment patterns, a higher proportion of the episodes with baseline PHQs engage in PT throughout the episode, a lower proportion of episodes with a baseline PHQ involve AP prescriptions, and the mean length of time until a second AD is prescribed is longer for those without baseline PHQs, while the mean days on the initial treatment are very close.