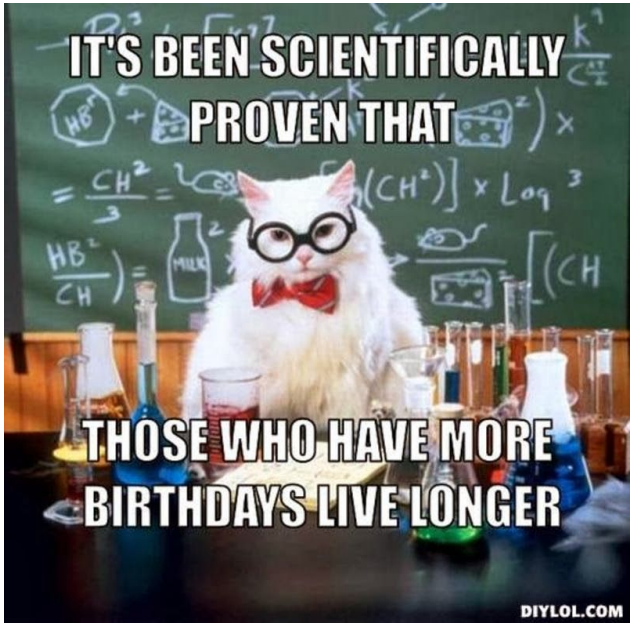
# Causal inference in irregularly observed longitudinal data: Progress and the many unsolved questions

(Bio)Statistics Research Day, McGill University
May 9th 2025

Janie Coulombe
janie.coulombe@umontreal.ca
Assistant Professor

Department of Mathematics and Statistics
Université de Montréal

# Contents



# 0. Why is this even a problem?

# What do we mean by irregular observation?

- Observational data (but could also be in randomized trials)
- Suppose, electronic health records data
- Measurements at the physician office often depend on patient characteristics (such as health condition)

# What do we mean by irregular observation?

- Observational data (but could also be in randomized trials)
- Suppose, electronic health records data
- Measurements at the physician office often depend on patient characteristics (such as health condition)

Questions 1-3

# Causal inference in longitudinal data

Questions 4-5

# Causal inference in longitudinal data

Questions 4-5

- Let exposure be denoted by $A(t) \in \{0, 1\}$
- Denote two potential outcomes[1,2] at time $t$ corresponding to exposure 1 and 0, $Y^1(t)$ and $Y^0(t)$
- If we are interested in a causal parameter, say the ATE

$$E[Y^1(t) - Y^0(t)] = \beta,$$

where the outcome is a longitudinal process indexed by time, then we would need the outcome of everyone at time $t$

1. Neyman, 1923; 2. Rubin, 1974

# Double problem

- Confounding
- Irregular visits: observed data correspond to several individual-specific selections, at different time points

# Double problem

- Confounding
- Irregular visits: observed data correspond to several individual-specific selections, at different time points
- We could study this using causal diagrams like the directed acyclic graph (DAG)

Question 6

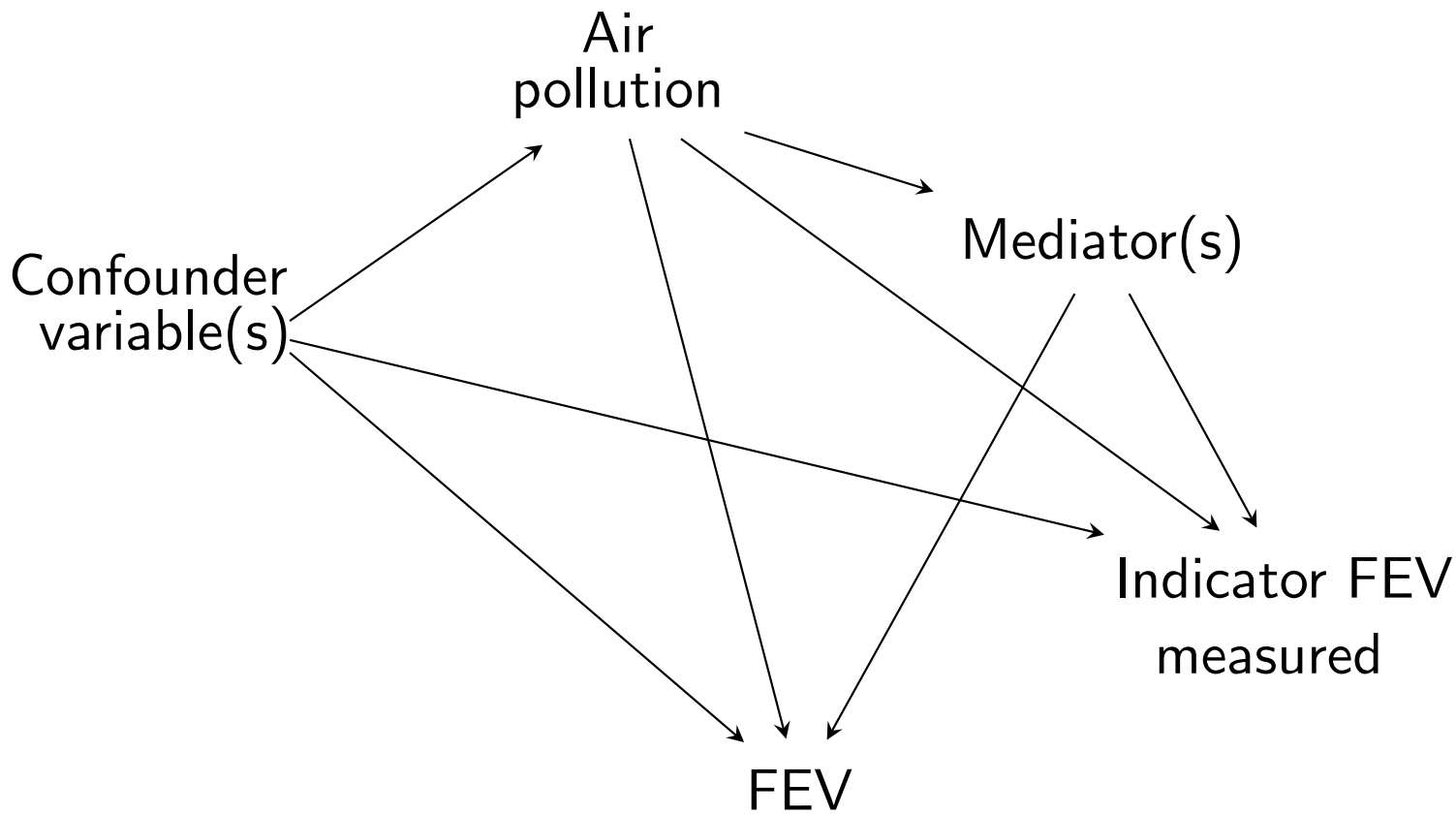
# Double problem

- Confounding
- Irregular visits: observed data correspond to several individual-specific selections, at different time points
- We could study this using causal diagrams like the directed acyclic graph (DAG)

Question 6

# DAG at each time $t$

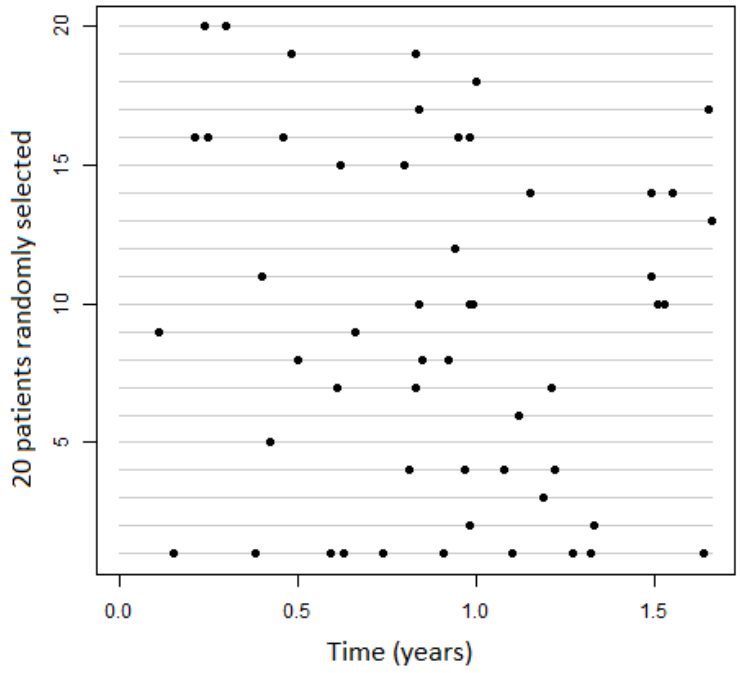
Assume the following causal diagram[1] at time $t$:



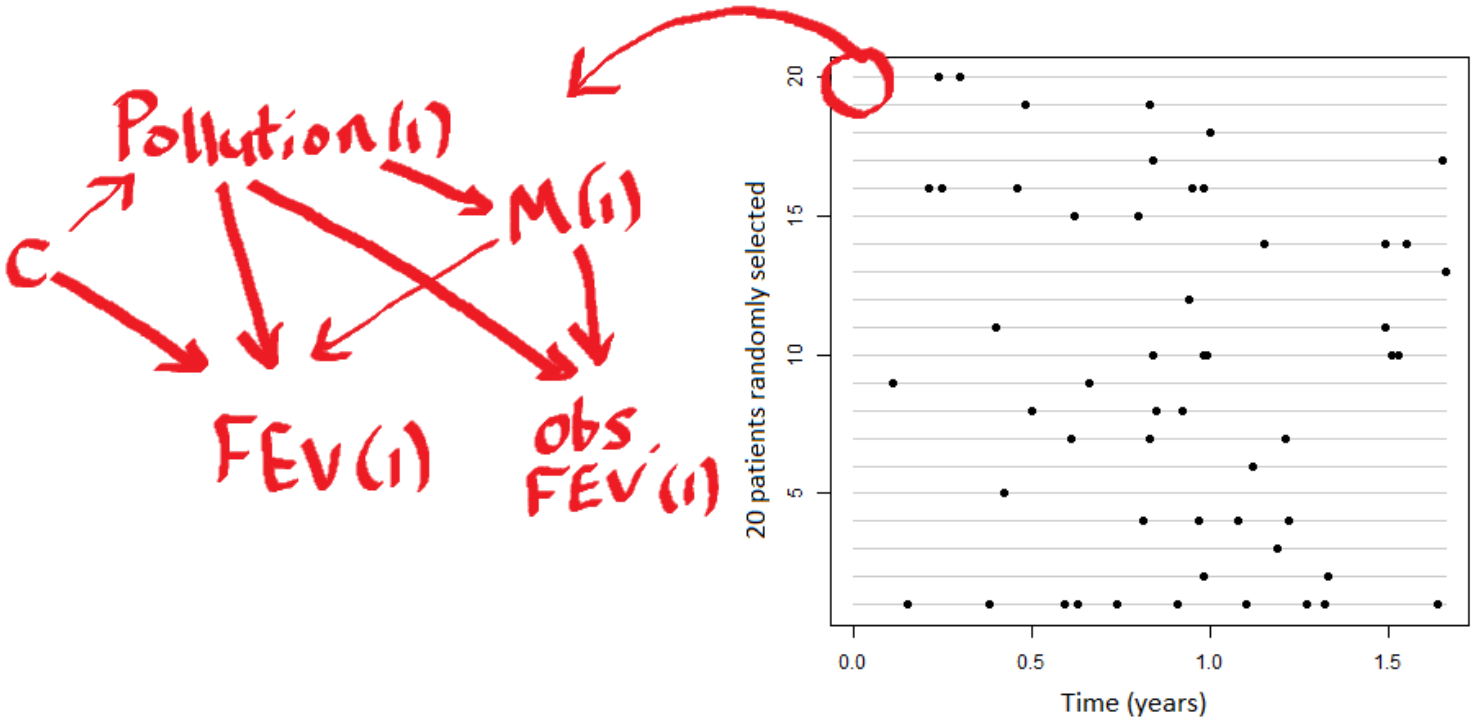

We suppose that all variables except the FEV outcome are measured each day.

1. Coulombe et al., 2021

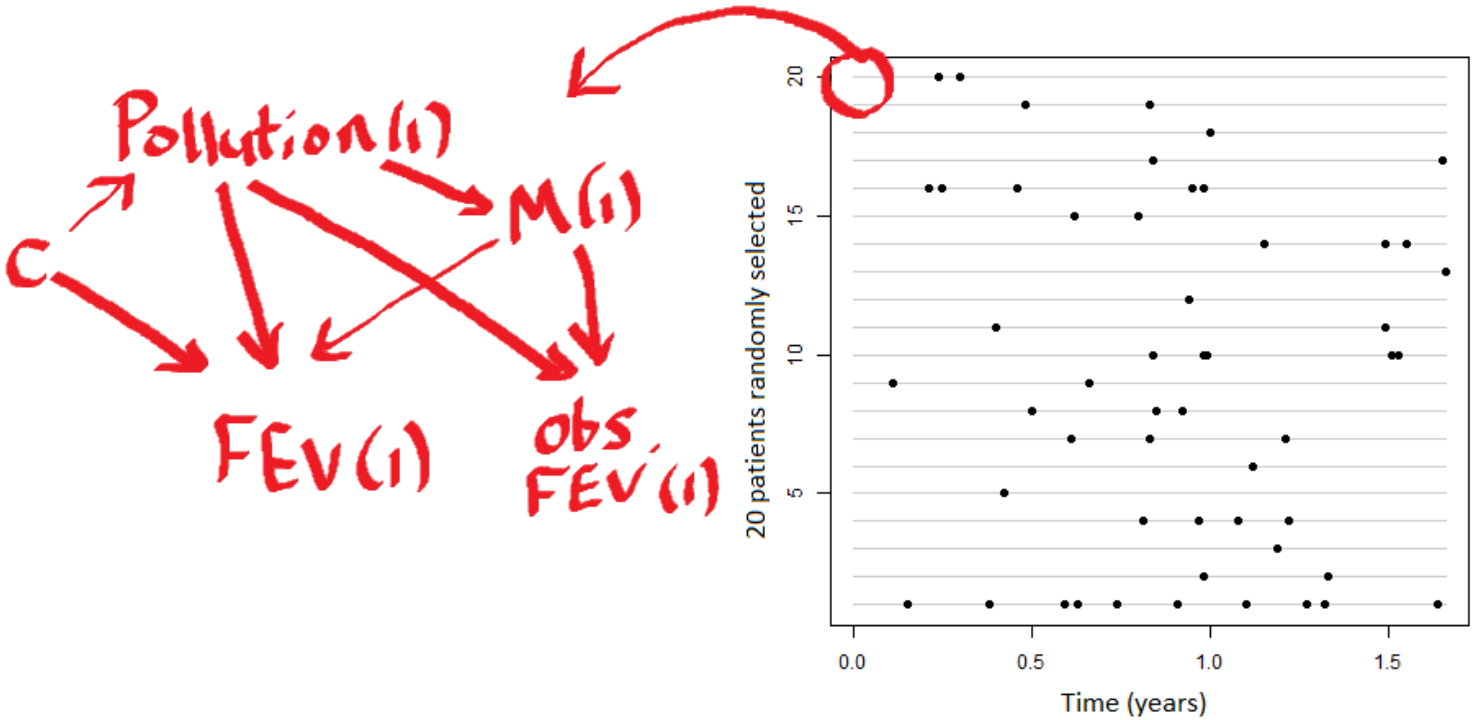
# Measurement times



# Measurement times



# Measurement times




Let $dN(t)$ be an indicator of visiting at time $t$ and $\mathbf{V}(t)$ be the visit predictors. Then, a **proportional rate model** assumes:

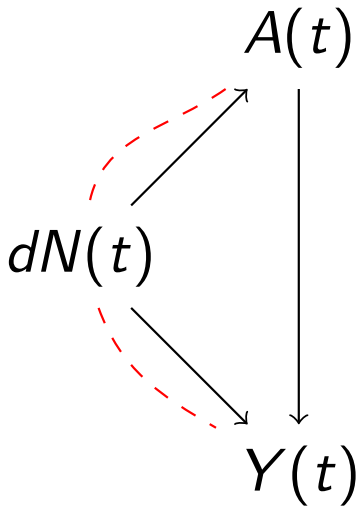
$$E[dN(t) \mid \mathbf{V}(t)] = \xi(t) \exp\{\mathbf{V}(t)\gamma\} \lambda_0(t)$$

# Studying the DAG

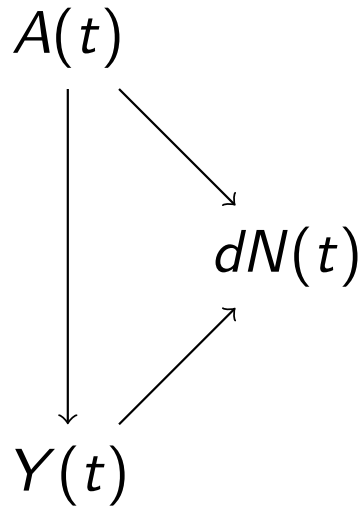
Denote again by $A(t)$, $Y(t)$, $dN(t)$ the exposure, outcome and visit indicator.

# Studying the DAG

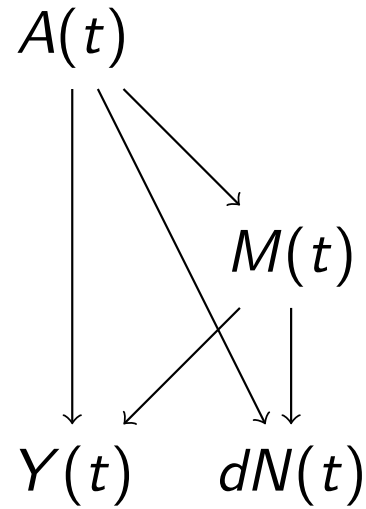
Denote again by $A(t)$, $Y(t)$, $dN(t)$ the exposure, outcome and visit indicator.




Confounder



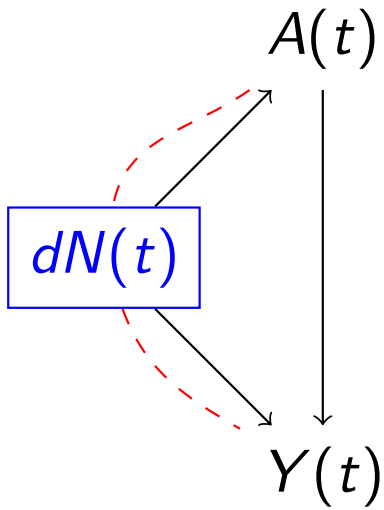

Collider



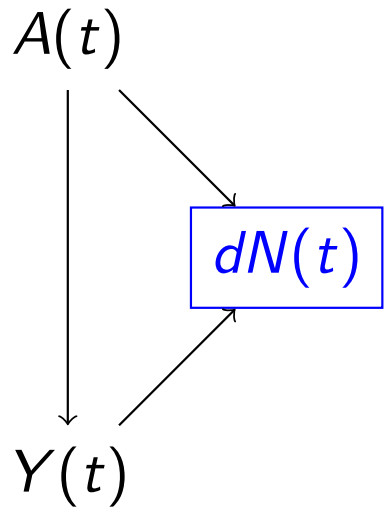

Collider

# Studying the DAG

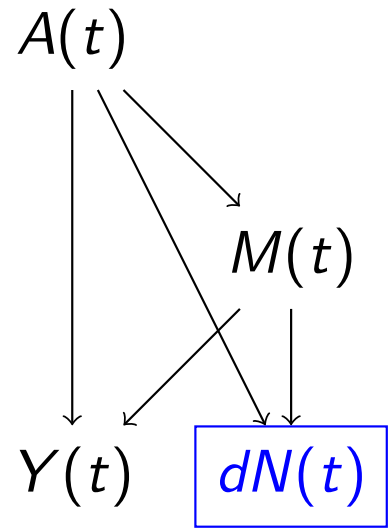
Denote again by $A(t)$, $Y(t)$, $dN(t)$ the exposure, outcome and visit indicator.




Confounder




Collider

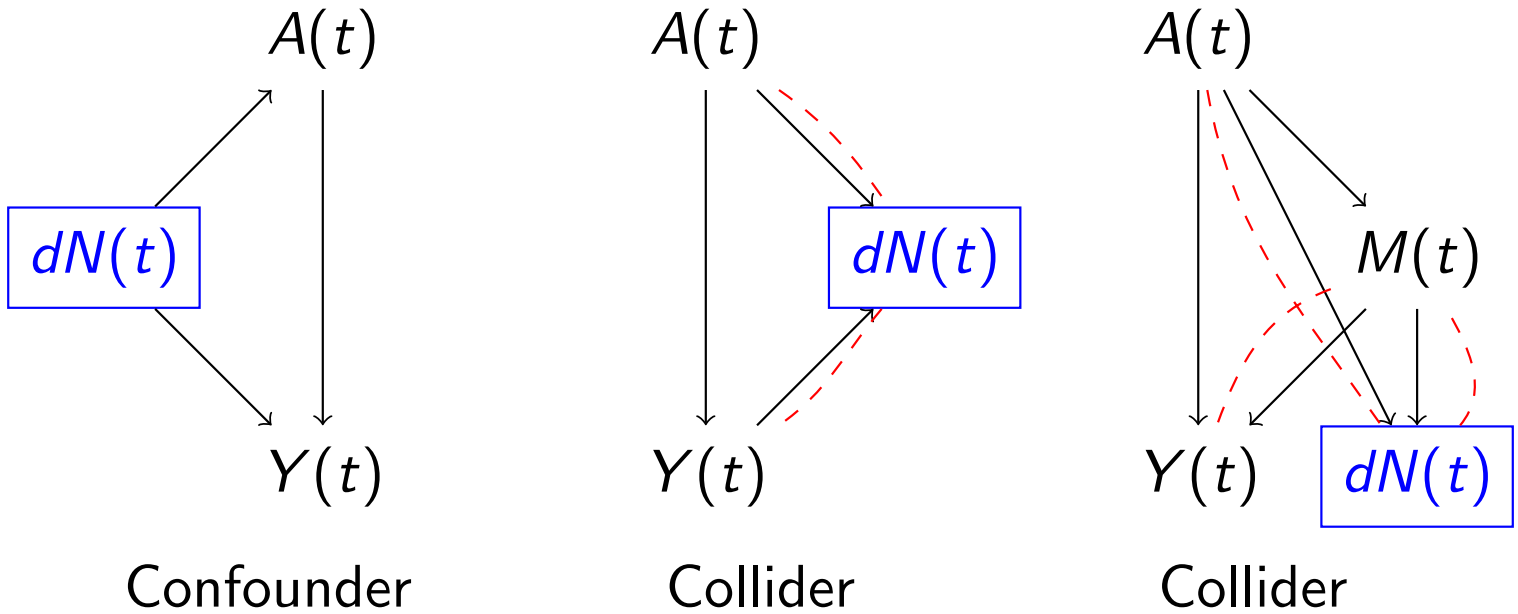



Collider

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

# Studying the DAG

Denote again by $A(t)$, $Y(t)$, $dN(t)$ the exposure, outcome and visit indicator.




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

# Goals of this talk

- Overview of questions, projects, and methods

# Goals of this talk

- Overview of questions, projects, and methods
- Encourage you to think big? And to be curious...

# Goals of this talk

- Overview of questions, projects, and methods
- Encourage you to think big? And to be curious...
- Collaborations and discussions are key! (My own experience)

# Goals of this talk

- Overview of questions, projects, and methods
- Encourage you to think big? And to be curious...
- Collaborations and discussions are key! (My own experience)
- Some students here are thinking about their research topic...

Questions 7-8

# 1. Comparison and the need for new estimators

- Need for causal estimators that are consistent in longitudinal studies with informative visit times

- Need for causal estimators that are consistent in longitudinal studies with informative visit times
- Different estimands might be of interest

- Need for causal estimators that are consistent in longitudinal studies with informative visit times
- Different estimands might be of interest
- Different types of exposure and outcomes could be studied

- Need for causal estimators that are consistent in longitudinal studies with informative visit times
- Different estimands might be of interest
- Different types of exposure and outcomes could be studied
- Also need to study in which settings irregular visit times are problematic (use of causal diagrams!)

# Example of estimands

- Average treatment effect

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

- Conditional treatment effect

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

- Adaptive treatment strategies

- and several others

Might require different assumptions to identify different estimands under irregular visits...

# Assumptions often required

1. Conditional exchangeability

   - Treatments were not randomized to patients, so the outcomes $Y^1(t)$ and $Y^0(t)$ are not measured on similar patients
   - Assume that conditional on patient characteristics/confounders $\mathbf{X}(t)$, we recover randomization:

$$A(t) \perp \{Y^0(t), Y^1(t)\} \mid \mathbf{X}(t)$$

   - We also need an assumption for the visit process:

$$dN(t) \perp Y(t) \mid \mathbf{V}(t)$$

2. Positivity

- Sometimes this assumption is required to avoid dividing by zero
- It also makes sense to assume that, in assessing an average treatment effect, everyone had a chance to receive both treatments and everyone had a chance of visiting at any time

$$0 < P(A(t) = a \mid \mathbf{X}(t)) < 1, a \in \{0, 1\}$$

and

$$0 < E[dN(t) \mid \mathbf{V}(t)]$$

3. Consistency of the outcome

- Suppose that $Y^1(t)$ is the potential outcome at time $t$ if the patient had smoked 30 cigarettes a day, for 7 days
- $Y^0(t)$ is the outcome if they had not smoked
- Suppose that you observe a patient who smoked 25 cigarettes a day, for 7 days...

$$Y(t) \stackrel{?}{=} \mathbb{I}(A(t) = 1)Y^1(t) + \mathbb{I}(A(t) = 0)Y^0(t)$$

# G-formula under irreguler measurements

$$
\begin{aligned}
E[Y^1(t) - Y^0(t)] =& E[Y^1(t)] - E[Y^0(t)] \\
=& E\left[E[Y^1(t) \mid \mathbf{X}(t), \mathbf{V}(t)]\right] - E\left[E[Y^0(t) \mid \mathbf{X}(t), \mathbf{V}(t)]\right] \\
=& E\left[E[Y^1(t) \mid A(t) = 1, \mathbf{X}(t), \mathbf{V}(t)]\right] - \\
& \qquad E\left[E[Y^0(t) \mid A(t) = 0, \mathbf{X}(t), \mathbf{V}(t)]\right] \\
=& E\left[E[Y(t) \mid A(t) = 1, \mathbf{X}(t), \mathbf{V}(t)]\right] - \\
& \qquad E\left[E[Y(t) \mid A(t) = 0, \mathbf{X}(t), \mathbf{V}(t)]\right] \\
=& E\left[E[Y(t) \mid A(t) = 1, dN(t) = 1, \mathbf{X}(t), \mathbf{V}(t)]\right] - \\
& \qquad E\left[E[Y(t) \mid A(t) = 0, dN(t) = 1, \mathbf{X}(t), \mathbf{V}(t)]\right] \\
& \ldots \\
=& E\left[\frac{\mathbb{I}(A(t) = 1, dN(t) = 1) Y(t)}{P(A(t) = 1 \mid \mathbf{X}(t)) E(dN(t) \mid \mathbf{V}(t))}\right] \\
& - E\left[\frac{\mathbb{I}(A(t) = 0, dN(t) = 1) Y(t)}{P(A(t) = 0 \mid \mathbf{X}(t)) E(dN(t) \mid \mathbf{V}(t))}\right].
\end{aligned}
$$

# Inverse intensity of visit (IIV) weights

Once a proportional rate model has been fitted, one can use inverse intensity of visit weights[3] defined as:

$$\text{IIV} = \frac{1}{\exp\left\{\mathbf{V}(t)\hat{\gamma}\right\}}$$

to create a pseudo-population in which covariates $\mathbf{V}(t)$ are not anymore associated with visiting or not at time $t$.

3. Lin, Scharfstein and Rosenheck, 2004

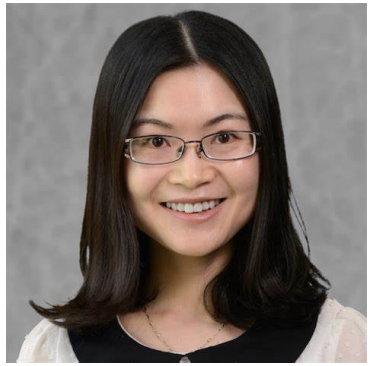
Doctoral work with prof. Erica E. M. Moodie and Robert W. Platt from McGill University.

Led to three different projects[4–6] using inverse weighting:

- Continuous outcome and binary treatment
- Ordinal outcome and continuous treatment
- Endogeneity (feedback between the visits and the covariates)

4. Coulombe, Moodie, and Platt, 2021a; 5. Coulombe, Moodie, and Platt, 2021b; 6. Coulombe, Moodie, Platt, and Renoux, 2022

# Making it more efficient...



Limit of the previous approaches?

Inverse-weighted estimating equations are not the most efficient.

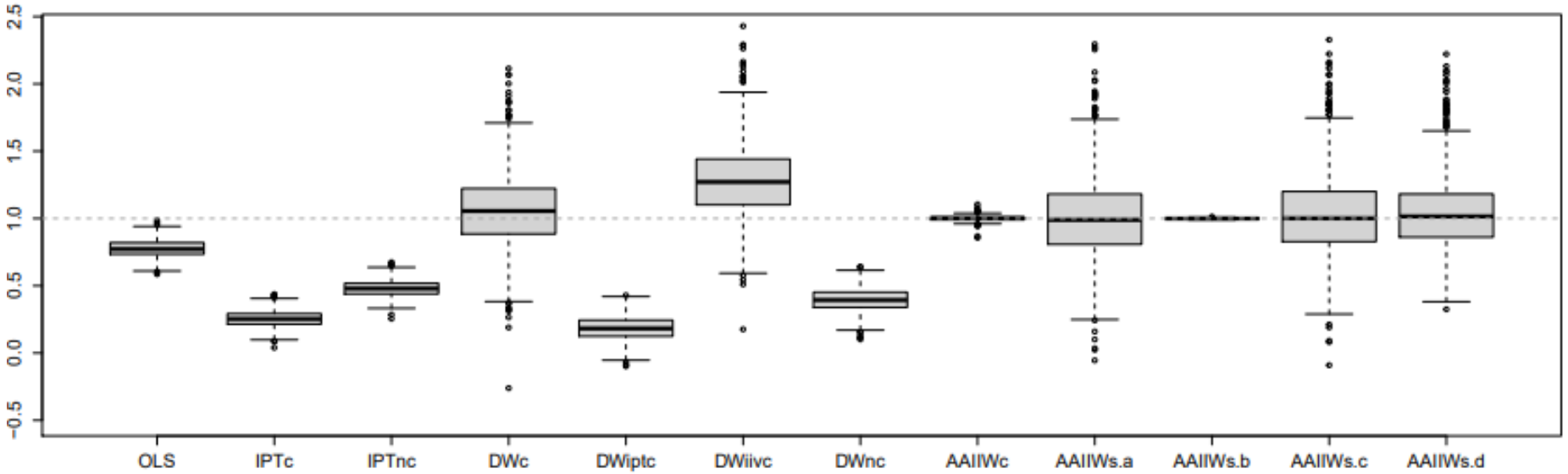
They also rely on the correct specification of the treatment and the visit models.

We further extended this work to construct a more robust and efficient estimator[7], with prof. Shu Yang from North Carolina State University.

7. Coulombe and Yang, 2024

# Simulation studies, sample size 1000

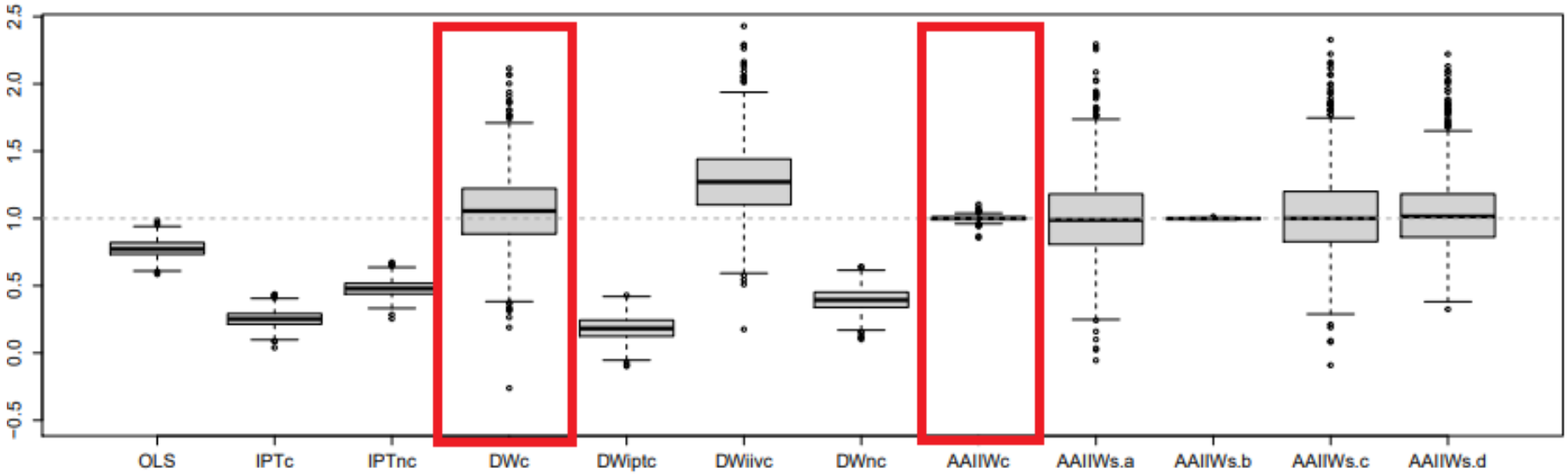
Distribution of 500 estimates:




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.

# 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.

# Joint models

Joint models can be useful if the visit and the outcome processes are still dependent in time, even after adjusting for patient covariates $\mathbf{V}(t)$.

They make different assumptions than our conditional exchangeability assumption. For instance, in Liang et al.[8], they use:

$$\begin{aligned} Y(t) =& \alpha_0(t) + \beta_0 X(t) + Z_1 W(t) + \epsilon(t), \\ E[dN(t) \mid Z_2, V] =& I(C \geq t) Z_2 \exp\left(\gamma_0' V\right) \lambda_0(t), \\ E[Z_1 \mid Z_2] =& \theta_0 (Z_2 - 1) \end{aligned}$$

for $Z_1, Z_2$ subject-specific random effects.

Recent works on irregular visit times have used this type of approach[8–10].

8. Liang, Lu, and Ying, 2009; 9. Pullenayegum, Birken, Maguire et al, 2023; 10. Sun, Song and Zhou, 2012

# Current work comparing inverse weighting and imputation

You may also wonder: Can't the data be imputed at times when there is no visit?

How does that approach compare to an inverse-weighting approach?

Recently, we investigated these questions with Dr. Erica E. M. Moodie and Dr. Susan M. Shortreed (Kaiser Permanente Washington).

# Some of our findings...

- Imputation requires coarsening of the data (e.g., weekly, monthly)
- With coarsening, the adjustment for confounding may be poorer
- Inverse weighting approaches are more straightforward but less convenient when other processes are observed irregularly
- Under similar assumptions or constraints, both imputation and inverse weighting approaches performed similarly overall

(This is work under review)

# 2. Developing guidelines

# Issue

- Even if literature on irregular measurement times is expanding...

# Issue

- Even if literature on irregular measurement times is expanding... a lot of research using longitudinal data does not consider irregular measurement times

# Issue

- Even if literature on irregular measurement times is expanding... a lot of research using longitudinal data does not consider irregular measurement times
- Different assumptions are made to avoid it, or they are not considered at all

# Issue

- Even if literature on irregular measurement times is expanding... a lot of research using longitudinal data does not consider irregular measurement times
- Different assumptions are made to avoid it, or they are not considered at all
- Researchers from other fields, analysts, and other users of these approaches are not sure when irregular measurement times are an issue

# Proposed solution

Project led by Prof. Eleanor Pullenayegum from the Dalla Lana School of Public Health at UToronto and the Hospital for Sick Children.

We are developing guidelines for researchers working with observational, longitudinal data at risk of irregular measurement times. These include considerations for:

- The estimand,
- Data collection,
- Analytic plan,
- Reporting

# 3. Something (new) to learn from missing data

# What if missing data provide information?

- We mostly don't like missing data! (who's with me?)

# What if missing data provide information?

- We mostly don't like missing data! (who's with me?)
- What if missing data could tell us something about some causal effects?

# What if missing data provide information?

- We mostly don't like missing data! (who's with me?)
- What if missing data could tell us something about some causal effects?
- Idea: Contrasting the outcome under different visit patterns (after adjusting for confounding, visit predictors, etc.) to study the causal effect of visits
- Do it dynamically? Construction of dynamic monitoring regimes...

# Extension of DTR methods...

In this recent work[11], we extended dynamic weighted ordinary least squares (dWOLS), an approach developed by Wallace and Moodie[12].

These authors assumed the following outcome model:

$$E[Y(t) \mid A(t), \mathbf{X}(t), \mathbf{Q}(t)] = f(\mathbf{X}(t); \psi_0) + \psi_1 A(t)\mathbf{Q}(t),$$

where $\mathbf{X}(t)$ are potential confounders and $\mathbf{Q}(t)$ tailoring variables.

Each month, should a patient visit their physician based on their personal characteristics?

If they should visit, should they also receive an add-on treatment?

We found...

11. Coulombe, El-Riachi, Du et al, 2025 (preprint); 12. Wallace and Moodie, 2015

# 4. Unsolved questions

# Cumulative effect of treatment

- The estimators we developed in previous work look at the causal "acute" treatment effect at time $t$.
- Mathilde Dicaire-Cartier, a master's student working with me, is looking at a cumulative exposure (duration) with irregularly observed outcomes.
- Her work assumes normality of the exposure distribution.

# ...and dynamic treatment regimes

- Dynamic treatment regimes (DTRs) often focus on a fixed, common number of stages (time points).
- What should we do with irregular observation times ? Is it even a problem? Under certain assumptions, not necessarily.
- In a 2023 paper[13], we assessed the development of adaptive treatment strategies (one-stage DTRs) under irregular visits
- Recently, Dong et al.[14] proposed a Bayesian approach for the estimation of optimal DTRs using irregularly observed data.

13. Coulombe, Moodie, Shortreed et al, 2023; 14. Dong, Pullenayegum, Thiébaut et al, 2025

# Machine learning and other methods for nuisance models

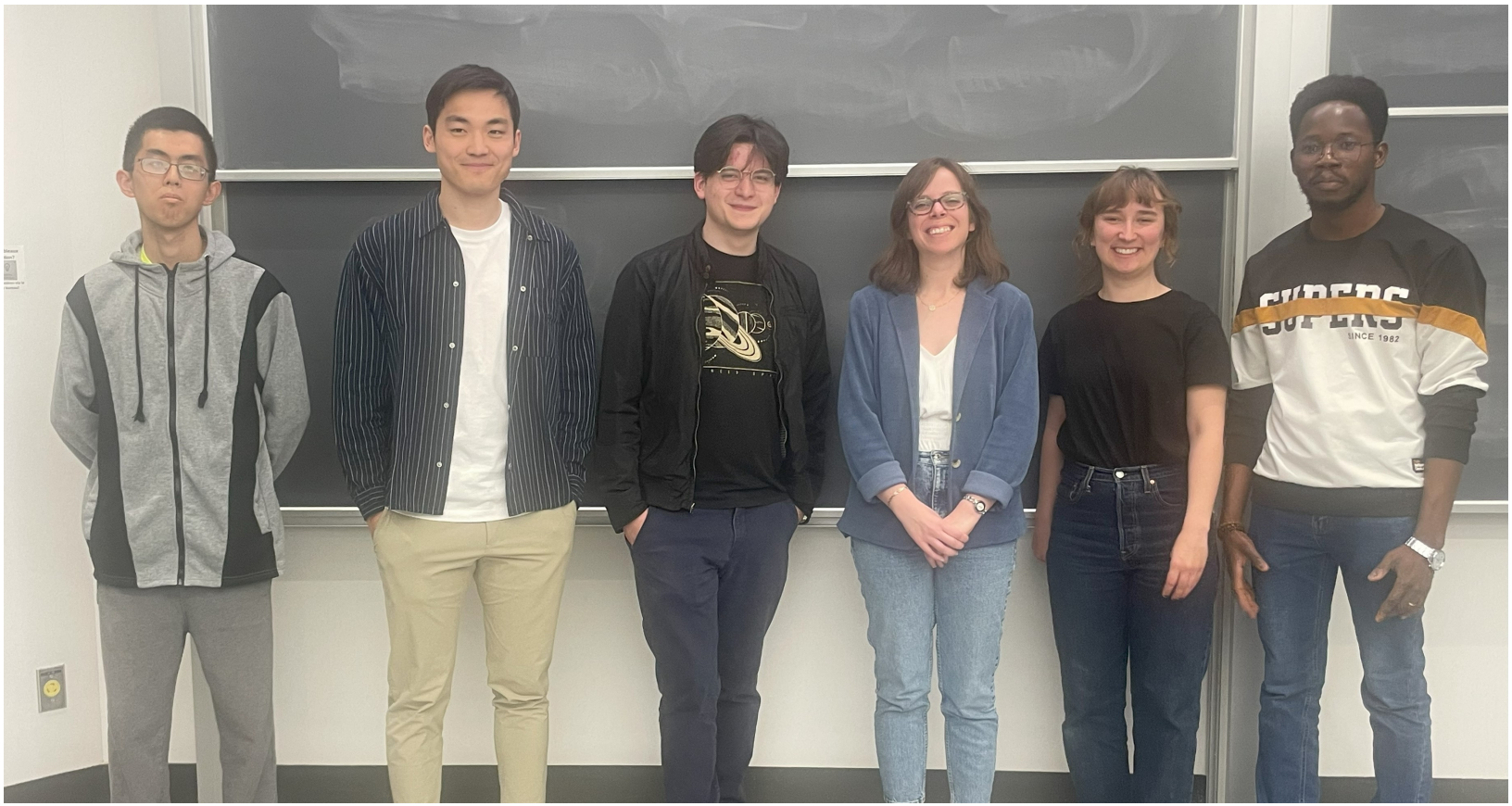
- So far, machine learning tools were not used in the visit nuisance model.
- In our work with Dr. Shu Yang, we developed the most efficient estimator for the ATE from its nonparametric class.
- We could incorporate machine learning in that estimator.
- Semiparametric models, parametric models, etc.

# 5. Conclusion

# In summary...

- There's been progress in this area
- Lot of ongoing work by different groups
- We hope that our set of guidelines on irregular visit times (work to come) will help users of these methods (Pullenayegum et al.)
- Collaboration is key - Contacts are important!
- One seemingly "straightforward" topic can lead to so many questions...
- Discussions with collaborators and students can generate new ideas!

# Lab



From left to right: Zong Yang Yu, Si Ming Xu, Emiliano Aviles Astorga, Janie, Mathilde Dicaire-Cartier, and Paguidame Sambiani.

# References

This work is supposed by an NSERC discovery grant, a FRQS chercheur-boursier J1, and a Banting-CANSSI discovery award.

Numbered references and slides can be found on my Github page, at https://janiecoulombestat.github.io/ (under the tab *Talks*).

**Many thanks to the organizers for the invitation and thanks for your attention!**