Efficient Estimation of Causal Effects Under Two-Phase Sampling with Error-Prone Outcome and Treatment Measurements

ENAR 2025 Spring Meeting

Keith Barnatchez, Kevin Josey, Nima Hejazi, Giovanni Parmigiani, Bryan E. Shepherd, and Rachel Nethery

March 24th, 2025

Roadmap

Background

Methods

Data Application

Discussion

Motivation

Explosion of the use of electronic health record (EHR) data for conducting observational causal inference studies

- Where, at a high level, one is interested in some measure of causal effect of a treatment A on an outcome of interest Y

Motivation

Explosion of the use of electronic health record (EHR) data for conducting observational causal inference studies

- Where, at a high level, one is interested in some measure of causal effect of a treatment A on an outcome of interest Y

And for good reason!

- EHR data is typically cheaper to obtain, fairly representative of patient populations, rich in information on potential confounding factors $\boldsymbol{X}$, and big

Motivation

Explosion of the use of electronic health record (EHR) data for conducting observational causal inference studies

- Where, at a high level, one is interested in some measure of causal effect of a treatment A on an outcome of interest Y

And for good reason!

- EHR data is typically cheaper to obtain, fairly representative of patient populations, rich in information on potential confounding factors $\boldsymbol{X}$, and big

But EHR data tends to present numerous challenges

- Including measurement error in outcomes (Y) and treatments (A) of interest

EHR data: in a perfect world

Suppose we aim to estimate the average treatment effect of a binary treatment A on Y

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

EHR data: in a perfect world

Suppose we aim to estimate the average treatment effect of a binary treatment A on Y

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

Y	A	$\boldsymbol{X}$
Y_1	A_1	$\boldsymbol{X}_1$
Y_2	A_2	$\boldsymbol{X}_2$
Y_3	A_3	$\boldsymbol{X}_3$
Y_4	A_4	$\boldsymbol{X}_4$
Y_5	A_5	$\boldsymbol{X}_5$
Y_6	A_6	$\boldsymbol{X}_6$

In a perfect world, we'd have access to the true outcome + treatment values (+ covariates $\boldsymbol{X}$)

EHR data: in ~~a perfect world~~ reality

Y	Y^*	A	A^*	$\boldsymbol{X}$
NA	Y_1^*	NA	A_1^*	$\boldsymbol{X}_1$
NA	Y_2^*	NA	A_2^*	$\boldsymbol{X}_2$
NA	Y_3^*	NA	A_3^*	$\boldsymbol{X}_3$
NA	Y_4^*	NA	A_4^*	$\boldsymbol{X}_4$
NA	Y_5^*	NA	A_5^*	$\boldsymbol{X}_5$
NA	Y_6^*	NA	A_6^*	$\boldsymbol{X}_6$

EHR data: in ~~a perfect world~~ reality

Y	Y^*	A	A^*	$\boldsymbol{X}$
NA	Y_1^*	NA	A_1^*	$\boldsymbol{X}_1$
NA	Y_2^*	NA	A_2^*	$\boldsymbol{X}_2$
NA	Y_3^*	NA	A_3^*	$\boldsymbol{X}_3$
NA	Y_4^*	NA	A_4^*	$\boldsymbol{X}_4$
NA	Y_5^*	NA	A_5^*	$\boldsymbol{X}_5$
NA	Y_6^*	NA	A_6^*	$\boldsymbol{X}_6$

In practice, we often only have error-prone measurements of Y and A, denoted Y^* and A^*

- Well-documented that using Y^* and A^* in place of Y and A can lead to severely biased causal effect estimates

EHR data: a study design-based workaround

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

In practice, can sometimes spend time + money to obtain gold-standard measurements for a random (typically small) subset of this data

EHR data: a study design-based workaround

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

In practice, selection into this subset can often be controlled, and may depend on the initial error-prone measurements: $R \not\perp\!\!\!\perp A^*, Y^*, \boldsymbol{X}$

- Especially common sampling strategy if Y and/or A are rare

EHR data: a study design-based workaround

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

In practice, selection into this subset can often be controlled, and may depend on the initial error-prone measurements: $R \not\perp\!\!\!\perp A^*, Y^*, \boldsymbol{X}$

- **Intuition**: over-sample subjects who contribute more information to the target estimand

Our Contributions

At a high-level, our work addresses the following question: How do we estimate causal effects nonparametrically when

Our Contributions

At a high-level, our work addresses the following question: How do we estimate causal effects nonparametrically when

1. We have error-prone outcome + treatment measurements for all subjects

Our Contributions

At a high-level, our work addresses the following question: How do we estimate causal effects nonparametrically when

1. We have error-prone outcome + treatment measurements for all subjects

2. We have gold-standard treatment + outcome measurements for a *subset* of our EHR data, where...

Our Contributions

At a high-level, our work addresses the following question: How do we estimate causal effects nonparametrically when

1. We have error-prone outcome + treatment measurements for all subjects
2. We have gold-standard treatment + outcome measurements for a *subset* of our EHR data, where...
3. This subset was collected according to a sampling rule that is **dependent** on the initially observed data: X, A^* and Y^* all influence R

Our Contributions

At a high-level, our work addresses the following question: How do we estimate causal effects nonparametrically when

1. We have error-prone outcome + treatment measurements for all subjects

2. We have gold-standard treatment + outcome measurements for a *subset* of our EHR data, where...

3. This subset was collected according to a sampling rule that is **dependent** on the initially observed data: X, A^* and Y^* all influence R

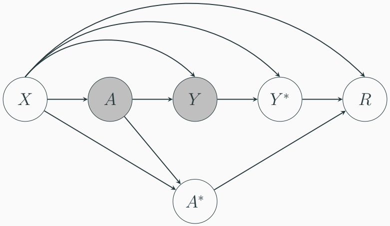
We present two asymptotically equivalent approaches to constructing efficient nonparametric causal effect estimators

Roadmap

Roadmap

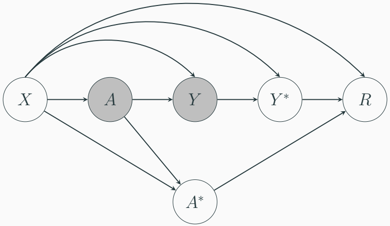
Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



Approach 1: Using the observed data distribution

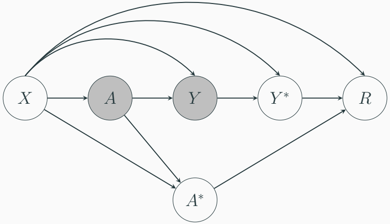
Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



Main goal: Estimate counterfactual means $\mathbb{E}[Y(a)]$ $(a \in \{0,1\})$ *efficiently*, allowing R to depend on $\boldsymbol{X}, Y^*$ and A^*

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

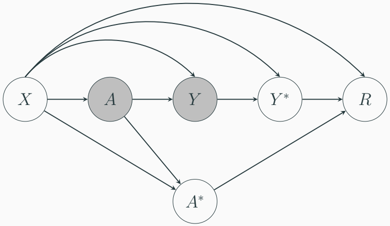


Construction of estimators follows standard semiparametric estimation pipeline:

causal estimand $\mathbb{E}[Y(a)]$ $\underset{\text{Assumptions}}{\longrightarrow}$ stat. estimand $\mathbb{E}[f(\text{data})]$ $\longrightarrow$ plug-in estimator $\hat{\psi}_a^{\text{PI}} = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(\text{data}_i)$ $\underset{\text{(tedious) work}}{\longrightarrow\longrightarrow\longrightarrow}$ debiased est. $\hat{\psi}_a^{\text{OS}}$

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

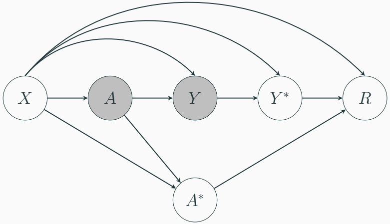


Construction of estimators follows standard semiparametric estimation pipeline:

$$\underset{\mathbb{E}[Y(a)]}{\text{causal estimand}} \xrightarrow[\text{Assumptions}]{} \underset{\mathbb{E}[f(\text{data})]}{\text{stat. estimand}} \longrightarrow \overbrace{\hat{\psi}_a^{\text{PI}} = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(\text{data}_i)}^{\text{plug-in estimator}} \xrightarrow[\text{(tedious) work}]{\longrightarrow\longrightarrow} \overset{\text{debiased est.}}{\hat{\psi}_a^{\text{OS}}}$$

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

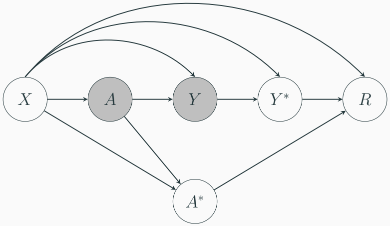


Construction of estimators follows standard semiparametric estimation pipeline:

$$\underset{\text{causal estimand}}{\mathbb{E}[Y(a)]} \xrightarrow[\text{Assumptions}]{} \underset{\text{stat. estimand}}{\mathbb{E}[f(\text{data})]} \longrightarrow \overset{\text{plug-in estimator}}{\hat{\psi}_a^{\text{PI}} = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(\text{data}_i)} \xrightarrow[\text{(tedious) work}]{} \overset{\text{debiased est.}}{\hat{\psi}_a^{\text{OS}}}$$

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

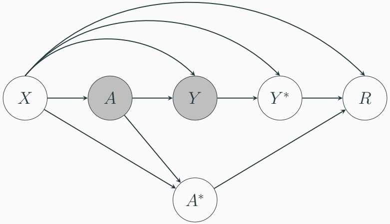


Construction of estimators follows standard semiparametric estimation pipeline:

$$\underset{\text{causal estimand}}{\mathbb{E}[Y(a)]} \underset{\text{Assumptions}}{\longrightarrow} \underset{\text{stat. estimand}}{\mathbb{E}[f(\text{data})]} \longrightarrow \overset{\text{plug-in estimator}}{\hat{\psi}_a^{\text{PI}} = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(\text{data}_i)} \underset{\text{(tedious) work}}{\longrightarrow\longrightarrow\longrightarrow} \overset{\text{debiased est.}}{\hat{\psi}_a^{\text{OS}}}$$

Approach 1: Using the observed data distribution

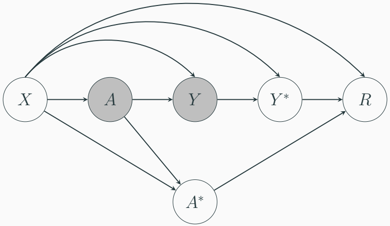
Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



Letting $\boldsymbol{Z} = (\boldsymbol{X}, A^*, Y^*)$, we derive a plug-in estimator

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



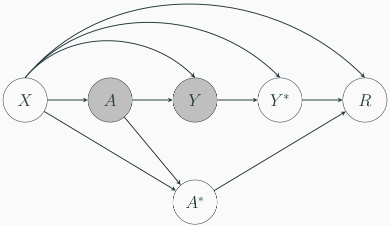
Letting $\boldsymbol{Z} = (\boldsymbol{X}, A^*, Y^*)$, we derive a plug-in estimator

$$\hat{\psi}_a^{\mathsf{PI,1}} = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\mathbb{E}}[\hat{\lambda}_a(\boldsymbol{Z}) \cdot \hat{\mu}_a(\boldsymbol{Z})|\boldsymbol{X}]}{\hat{\mathbb{E}}[\hat{\lambda}_a(\boldsymbol{Z})|\boldsymbol{X}]}$$

where $\hat{\lambda}_a$ and $\hat{\mu}_a$ are imputation functions for A and Y, respectively

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



Letting $\boldsymbol{Z} = (\boldsymbol{X}, A^*, Y^*)$, we derive a plug-in estimator

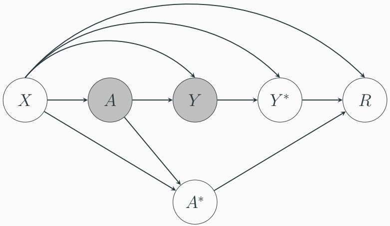
$$\hat{\psi}_a^{\mathsf{PI},1} = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\mathbb{E}}[\hat{\lambda}_a(\boldsymbol{Z}) \cdot \hat{\mu}_a(\boldsymbol{Z}) | \boldsymbol{X}]}{\hat{\mathbb{E}}[\hat{\lambda}_a(\boldsymbol{Z}) | \boldsymbol{X}]}$$

where $\hat{\lambda}_a$ and $\hat{\mu}_a$ are imputation functions for A and Y, respectively

- Interpretation: IPW on imputed values, after marginalizing out post-treatment variables

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



Letting $\boldsymbol{Z} = (\boldsymbol{X}, A^*, Y^*)$, we derive a plug-in estimator

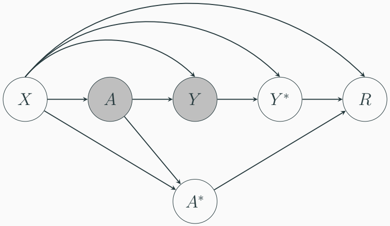
$$\hat{\psi}_a^{\text{PI},1} = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\mathbb{E}}[\hat{\lambda}_a(\boldsymbol{Z}) \cdot \hat{\mu}_a(\boldsymbol{Z})|\boldsymbol{X}]}{\hat{\mathbb{E}}[\hat{\lambda}_a(\boldsymbol{Z})|\boldsymbol{X}]}$$

where $\hat{\lambda}_a$ and $\hat{\mu}_a$ are imputation functions for A and Y, respectively

- Drawback: Inference intractable when nuisance models are fit data-adaptively

Approach 1: Using the observed data distribution

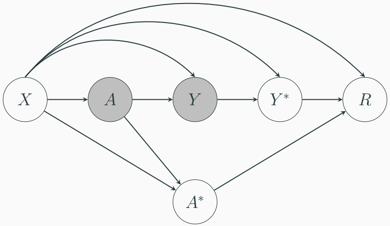
Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



To enable inference, we derive a one-step debiased estimator

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



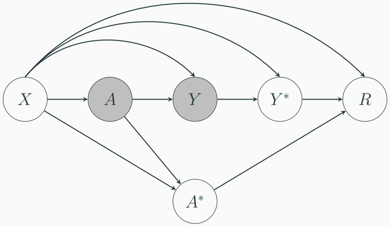
To enable inference, we derive a one-step debiased estimator

$$\hat{\psi}_a^{\mathsf{OS},1} = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\mathbb{E}}[\hat{\lambda}_a(\boldsymbol{Z}) \cdot \hat{\mu}_a(\boldsymbol{Z})|\boldsymbol{X}]}{\hat{\mathbb{E}}[\hat{\lambda}_a(\boldsymbol{Z})|\boldsymbol{X}]} + \widehat{\mathsf{BC}}$$

where $\widehat{\mathsf{BC}}$ is a *bias correction* term based on the efficient influence function for ψ_a

Approach 1: Using the observed data distribution

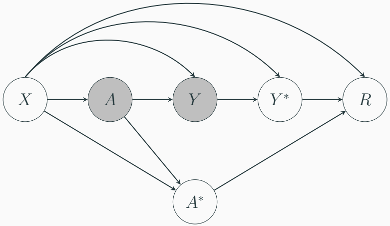
Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



We document multiple properties of $\hat{\psi}_a^{\mathsf{OS},1}$

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

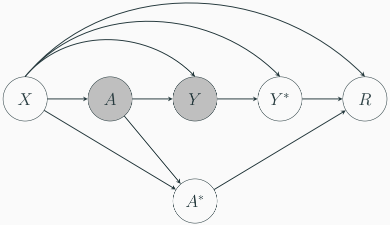


We document multiple properties of $\hat{\psi}_a^{\mathsf{OS},1}$

1. Bias correction enables valid inference when nuisance models are fit with flexible ML methods that converge at $n^{1/4}$ rates, however...

Approach 1: Using the observed data distribution

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



We document multiple properties of $\hat{\psi}_a^{\mathsf{OS},1}$

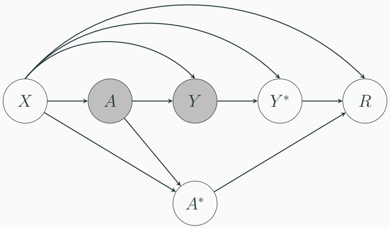
1. Bias correction enables valid inference when nuisance models are fit with flexible ML methods that converge at $n^{1/4}$ rates, however...

2. This bias correction term introduces numerous unstable weighting terms that can harm finite sample performance
   - Particularly concerning, as validation samples tend to be small in practice

Roadmap

Approach 2: Complete-data projection

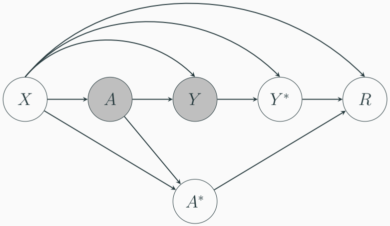
Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



Approach 2 is based on a well-developed, but relatively underutilized, framework for constructing debiased estimators under missing data (van der Laan and Robins 2003)

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

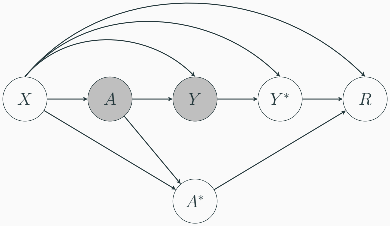


Approach 2 is based on a well-developed, but relatively underutilized, framework for constructing debiased estimators under missing data (van der Laan and Robins 2003)

- Links (i) the estimator we'd ideally construct under complete data to (ii) the observed data structure, where Y and A are partially missing

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



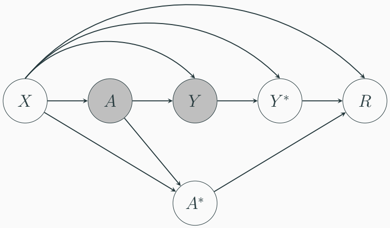
Idea: If we had complete data, could construct a plug-in estimator $\hat{\psi}_a^{\text{PI},2} = \frac{1}{n}\sum_{i=1}^{n} \hat{m}_a(\boldsymbol{X}_i)$,

$$\hat{\psi}_a^{\text{OS},1} = \hat{\psi}_a^{\text{PI},2} + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{m}_a(\boldsymbol{X}_i) + \frac{I(A_i = a)}{g_a(\boldsymbol{X}_i)}\{Y_i - \hat{m}_a(\boldsymbol{X}_i)\} - \hat{\psi}_a^{\text{PI},2}\right)$$

where $m_a(\boldsymbol{X}) = \mathbb{E}(Y|A = a, \boldsymbol{X})$ and $g_a(\boldsymbol{X}) = \mathbb{P}(A = a|\boldsymbol{X})$

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



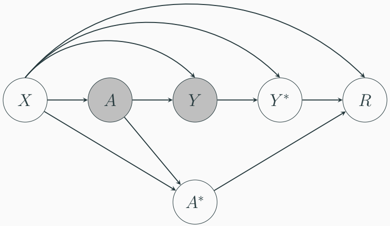
Idea: If we had complete data, could construct a plug-in estimator $\hat{\psi}_a^{\text{PI,2}} = \frac{1}{n}\sum_{i=1}^n \hat{m}_a(\boldsymbol{X}_i)$, as well as an AIPW estimator

$$\hat{\psi}_a^{\text{OS,2}} = \hat{\psi}_a^{\text{PI,2}} + \frac{1}{n}\sum_{i=1}^n \left(\hat{m}_a(\boldsymbol{X}_i) + \frac{I(A_i = a)}{\hat{g}_a(\boldsymbol{X}_i)}\{Y_i - \hat{m}_a(\boldsymbol{X}_i)\} - \hat{\psi}_a^{\text{PI,2}} \right)$$

where $m_a(\boldsymbol{X}) = \mathbb{E}(Y|A = a, \boldsymbol{X})$ and $g_a(\boldsymbol{X}) = \mathbb{P}(A = a|\boldsymbol{X})$

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



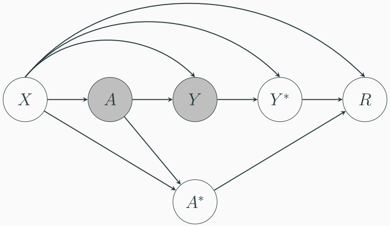
Idea: If we had complete data, could construct a plug-in estimator $\hat{\psi}_a^{\text{PI},2} = \frac{1}{n}\sum_{i=1}^n \hat{m}_a(\boldsymbol{X}_i)$, as well as an AIPW estimator

$$\hat{\psi}_a^{\text{OS},2} = \hat{\psi}_a^{\text{PI},2} + \frac{1}{n}\sum_{i=1}^n \left(\hat{m}_a(\boldsymbol{X}_i) + \frac{I(A_i = a)}{\hat{g}_a(\boldsymbol{X}_i)}\{Y_i - \hat{m}_a(\boldsymbol{X}_i)\} - \hat{\psi}_a^{\text{PI},2} \right)$$

Above, $m_a(\boldsymbol{X})$ and $\hat{g}_a(\boldsymbol{X})$ can be estimated with weighted regressions that add weights $R/\mathbb{P}(R = 1|\boldsymbol{Z})$ to the underlying loss function

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



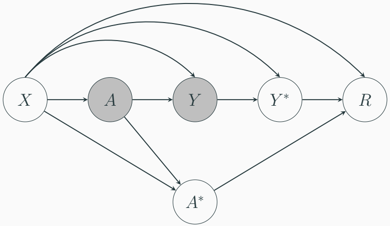
$$\hat{\psi}_a^{\mathsf{OS},2} = \hat{\psi}_a^{\mathsf{PI},2} + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{m}_a(\boldsymbol{X}_i) + \frac{I(A_i = a)}{\hat{g}_a(\boldsymbol{X}_i)}\{Y_i - \hat{m}_a(\boldsymbol{X}_i)\} - \hat{\psi}_a^{\mathsf{PI},2}\right)$$

Issue: The bias correction terms are <u>only observed</u> when $R_i = 1$

- Above estimator is infeasible

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



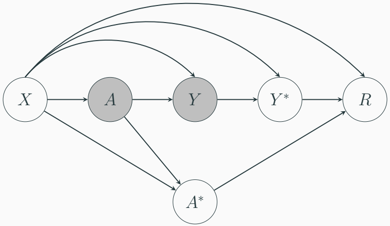
Key idea: Treat the bias correction terms as pseudo-outcomes:

$$\hat{\psi}_a^{\mathsf{OS,2}} = \hat{\psi}_a^{\mathsf{PI,2}} + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{m}_a(\boldsymbol{X}_i) + \frac{I(A_i = a)}{\hat{g}_a(\boldsymbol{X}_i)}\{Y_i - \hat{m}_a(\boldsymbol{X}_i)\} - \hat{\psi}_a^{\mathsf{PI,2}}\right)$$

$$= \hat{\psi}_a^{\mathsf{PI,2}} + \frac{1}{n}\sum_{i=1}^{n}\underbrace{\chi_a(\boldsymbol{O}_i; \hat{m}_a, \hat{g}_a)}_{\text{pseudo outcome}}$$

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

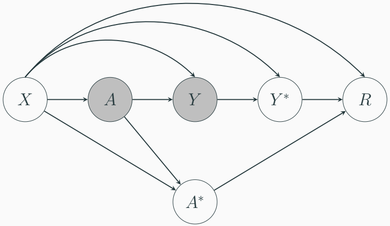


Key idea: Perform AIPW on the pseudo-outcome:

$$\hat{\psi}_a^{\mathsf{OS},2} = \hat{\psi}_a^{\mathsf{PI},2} + \frac{1}{n}\sum_{i=1}^{n} \chi_a(\boldsymbol{O}_i; \hat{m}_a, \hat{g}_a)$$

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1

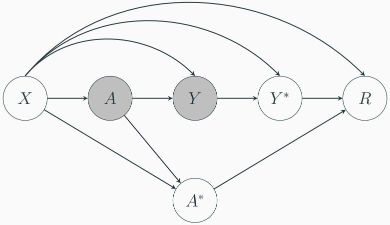


Key idea: Perform AIPW on the pseudo-outcome:

$$\hat{\psi}_a^{\mathsf{OS,2}} = \hat{\psi}_a^{\mathsf{PI,2}} + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varphi}_a(\boldsymbol{Z}_i) + \frac{R_i}{\mathbb{P}(R_i = 1|\boldsymbol{Z}_i)}\{\chi_a(\boldsymbol{O}_i;\hat{m}_a,\hat{g}_a) - \hat{\varphi}_a(\boldsymbol{Z}_i)\}\right)$$

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



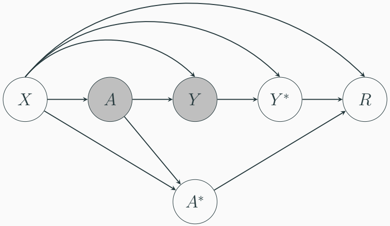
Key idea: Perform AIPW on the pseudo-outcome:

$$\hat{\psi}_a^{\text{OS},2} = \hat{\psi}_a^{\text{PI},2} + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varphi}_a(\boldsymbol{Z}_i) + \frac{R_i}{\mathbb{P}(R_i = 1|\boldsymbol{Z}_i)}\{\chi_a(\boldsymbol{O}_i; \hat{m}_a, \hat{g}_a) - \hat{\varphi}_a(\boldsymbol{Z}_i)\}\right)$$

where $\hat{\varphi}_a(\boldsymbol{Z}) = \hat{\mathbb{E}}[\chi_a(\boldsymbol{O}; \hat{m}_a, \hat{g}_a)|\boldsymbol{Z}, R = 1]$

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



Key idea: Perform AIPW on the pseudo-outcome:

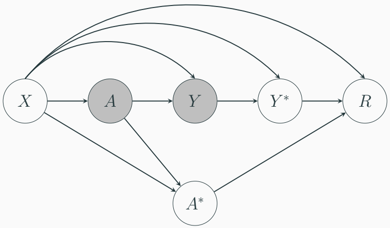
$$\hat{\psi}_a^{\mathsf{OS},2} = \hat{\psi}_a^{\mathsf{PI},2} + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varphi}_a(\boldsymbol{Z}_i) + \frac{R_i}{\mathbb{P}(R_i = 1|\boldsymbol{Z}_i)}\{\chi_a(\boldsymbol{O}_i; \hat{m}_a, \hat{g}_a) - \hat{\varphi}_a(\boldsymbol{Z}_i)\}\right)$$

where $\hat{\varphi}_a(\boldsymbol{Z}) = \mathbb{E}[\chi_a(\boldsymbol{O}_i; \hat{m}_a, \hat{g}_a)|\boldsymbol{Z}, R = 1]$

- Feasible estimator!
- When sampling probabilities are known, $\hat{\psi}_a^{\mathsf{OS},2}$ is doubly-robust in the traditional sense

Approach 2: Complete-data projection

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	0
Y_3	Y_3^*	A_3	A_3^*	X_3	0
Y_4	Y_4^*	A_4	A_4^*	X_4	1
Y_5	Y_5^*	A_5	A_5^*	X_5	0
Y_6	Y_6^*	A_6	A_6^*	X_6	1



Connections between $\hat{\psi}_a^{\mathsf{OS},1}$ and $\hat{\psi}_a^{\mathsf{OS},2}$:

- Can be shown that asymptotically, Approach 1 and Approach 2 estimators are *equivalent*
  - Approach 2 can be viewed as a re-parameterization of Approach 1
- Both estimators have unique sources of finite sample instability

Roadmap

Background

Methods

Data Application

Discussion

The Vanderbilt Comprehensive Care Clinic (VCCC)

EHR database with information on $\approx$ 1900 patients living with HIV between 1998-2011 that began receiving care at the VCCC

- Numerous variables recorded with error
    - Date of antiretroviral therapy (ART) initiation
    - Occurrence of AIDS-defining events (ADEs)

The Vanderbilt Comprehensive Care Clinic (VCCC)

EHR database with information on $\approx$ 1900 patients living with HIV between 1998-2011 that began receiving care at the VCCC

- Numerous variables recorded with error
  - Date of antiretroviral therapy (ART) initiation
  - Occurrence of AIDS-defining events (ADEs)

- Reliable measurements on demographic information, risk factors and baseline lab measurements (baseline covariates, $\boldsymbol{X}$)

The Vanderbilt Comprehensive Care Clinic (VCCC)

EHR database with information on $\approx$ 1900 patients living with HIV between 1998-2011 that began receiving care at the VCCC

- Numerous variables recorded with error
  - Date of antiretroviral therapy (ART) initiation
  - Occurrence of AIDS-defining events (ADEs)

- Reliable measurements on demographic information, risk factors and baseline lab measurements (baseline covariates, $\boldsymbol{X}$)

- Team at Vanderbilt University Medical Center manually validated *every* patient's observation in the database
  - Effectively yielding a fully validated dataset alongside an error-prone dataset

The Vanderbilt Comprehensive Care Clinic (VCCC)

EHR database with information on $\approx$ 1900 patients living with HIV between 1998-2011 that began receiving care at the VCCC

- Numerous variables recorded with error
  - Date of antiretroviral therapy (ART) initiation
  - Occurrence of AIDS-defining events (ADEs)
- Reliable measurements on demographic information, risk factors and baseline lab measurements (baseline covariates, $\boldsymbol{X}$)
- Team at Vanderbilt University Medical Center manually validated *every* patient's observation in the database
  - Effectively yielding a fully validated dataset alongside an error-prone dataset

Causal estimand: Average causal effect of starting ART within 1 month of first visit (A) on 3-year (post initial visit) risk of suffering an ADE (Y)

VCCC application

"Reveal" increasingly larger shares of the validated data $\mathbb{P}(R = 1)$, over-sampling those with $A^* = 1$ and $Y^* = 1$

- E.g. 5%, 10%, $15\%, \ldots$
- Implement proposed estimators at each share

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
	Y_2^*		A_2^*	X_2	0
	Y_3^*		A_3^*	X_3	0
	Y_4^*		A_4^*	X_4	0
	Y_5^*		A_5^*	X_5	0
	Y_6^*		A_6^*	X_6	0
	Y_7^*		A_7^*	X_7	0
	Y_8^*		A_8^*	X_8	0
	Y_9^*		A_9^*	X_9	0
	Y_{10}^*		A_{10}^*	X_{10}	0

VCCC application

"Reveal" increasingly larger shares of the validated data $\mathbb{P}(R = 1)$, over-sampling those with $A^* = 1$ and $Y^* = 1$

- E.g. 5%, 10%, $15\%, \ldots$
- Implement proposed estimators at each share

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	1
	Y_3^*		A_3^*	X_3	0
	Y_4^*		A_4^*	X_4	0
	Y_5^*		A_5^*	X_5	0
	Y_6^*		A_6^*	X_6	0
	Y_7^*		A_7^*	X_7	0
	Y_8^*		A_8^*	X_8	0
	Y_9^*		A_9^*	X_9	0
	Y_{10}^*		A_{10}^*	X_{10}	0

VCCC application

"Reveal" increasingly larger shares of the validated data $\mathbb{P}(R=1)$, over-sampling those with $A^*=1$ and $Y^*=1$

- E.g. 5%, 10%, $15\%, \ldots$
- Implement proposed estimators at each share

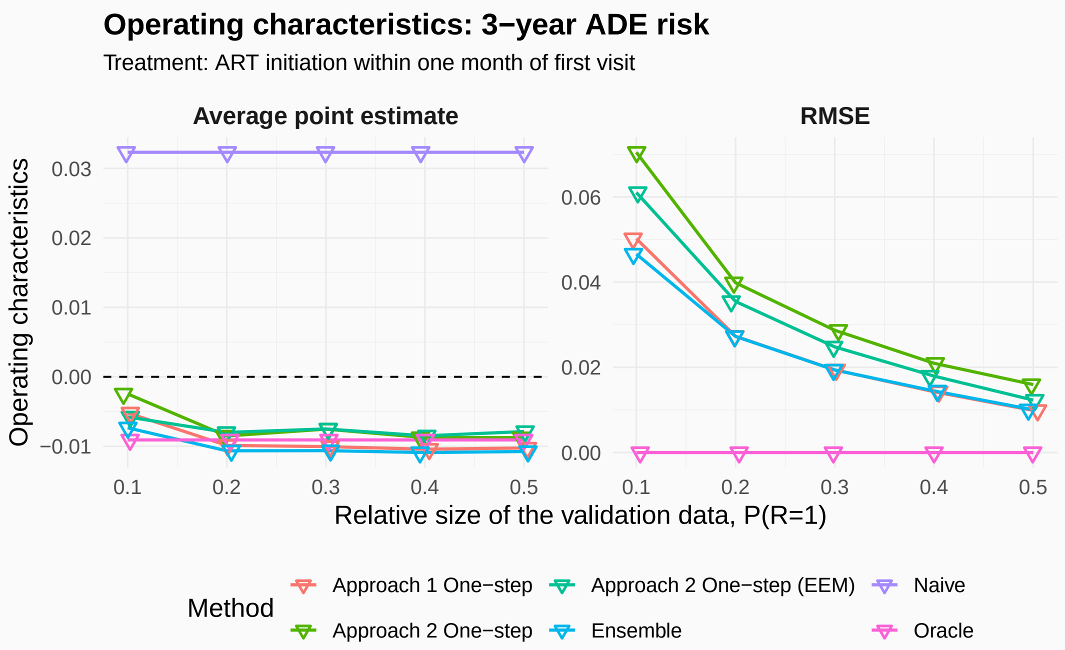
Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	1
Y_3	Y_3^*	A_3	A_3^*	X_3	1
	Y_4^*		A_4^*	X_4	0
	Y_5^*		A_5^*	X_5	0
	Y_6^*		A_6^*	X_6	0
	Y_7^*		A_7^*	X_7	0
	Y_8^*		A_8^*	X_8	0
	Y_9^*		A_9^*	X_9	0
	Y_{10}^*		A_{10}^*	X_{10}	0

VCCC application

"Reveal" increasingly larger shares of the validated data $\mathbb{P}(R = 1)$, over-sampling those with $A^* = 1$ and $Y^* = 1$

- E.g. 5%, 10%, $15\%, \ldots$
- Implement proposed estimators at each share

Y	Y^*	A	A^*	X	R
Y_1	Y_1^*	A_1	A_1^*	X_1	1
Y_2	Y_2^*	A_2	A_2^*	X_2	1
Y_3	Y_3^*	A_3	A_3^*	X_3	1
Y_4	Y_4^*	A_4	A_4^*	X_4	1
	Y_5^*		A_5^*	X_5	0
	Y_6^*		A_6^*	X_6	0
	Y_7^*		A_7^*	X_7	0
	Y_8^*		A_8^*	X_8	0
	Y_9^*		A_9^*	X_9	0
	Y_{10}^*		A_{10}^*	X_{10}	0


Operating characteristics: 3−year ADE risk
Treatment: ART initiation within one month of first visit
Average point estimate
RMSE
Operating characteristics
0.03
0.02
0.01
0.00
−0.01
0.06
0.04
0.02
0.00
0.1
0.2
0.3
0.4
0.5
Relative size of the validation data, P(R=1)
Method
Approach 1 One−step
Approach 2 One−step (EEM)
Naive
Approach 2 One−step
Ensemble
Oracle

Roadmap

Discussion

Our work addresses the following high-level question:

How do we estimate causal effects nonparametrically when **(1)** we have error-prone outcome + treatment measurements, **(2)** we have gold-standard treatment + outcome measurements for a *subset* of our data, where **(3)** that subset was collected according to a sampling rule that depends on the initially observed data

We've presented plug-in + debiased one-step estimators that accommodate these sampling schemes (and general two-phase sampling schemes). Currently working on

- Further studying approaches that improve the finite-sample behavior of the one-step estimators
- Development of R package implementing Approach 2 for general missing data patterns

Thank you!

Working paper coming soon!

References

van der Laan, M. J. and Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality*. Springer.

Appendix

Identification

First, let $\boldsymbol{W} = (Y^*, A^*)$. Then, we have

$$\begin{aligned}
\mathbb{E}(Y(a)) &= \mathbb{E}_{\boldsymbol{X}}\mathbb{E}(Y(a)|\boldsymbol{X}) \\
&= \mathbb{E}_{\boldsymbol{W}(a)|\boldsymbol{X}}\left[\mathbb{E}_{\boldsymbol{X}}\mathbb{E}(Y(a)|\boldsymbol{X}, \boldsymbol{W}(a))\right] \\
&= \mathbb{E}_{\boldsymbol{W}(a)|\boldsymbol{X}}\left[\mathbb{E}_{\boldsymbol{X}}\mathbb{E}(Y(a)|\boldsymbol{X}, \boldsymbol{W}(a), A = a, R = 1)\right] \\
&= \mathbb{E}_{\boldsymbol{W}|\boldsymbol{X},A}\left[\mathbb{E}_{\boldsymbol{X}}\mathbb{E}(Y|\boldsymbol{X}, \boldsymbol{W}, A = a, R = 1)\right]
\end{aligned}$$

The problem: the density $p(w|x, a)$ is unidentified

- We only see A in the validation data, which is **not** independent of $\boldsymbol{W}$ so we can't just condition on $R = 1$

Identification

Notice

$$\mathbb{E}(Y(a)) = \sum_w \sum_x \sum_y y \cdot p(y|w, x, a, r = 1) \cdot p(w|x, a) \cdot p(x)$$

Identification

Notice

$$\begin{aligned}
\mathbb{E}(Y(a)) &= \sum_w \sum_x \sum_y y \cdot p(y|w,x,a,r=1) \cdot p(w|x,a) \cdot p(x) \\
&= \sum_w \sum_x \sum_y y \cdot p(y|w,x,a,r=1) \cdot \frac{p(a|x,w)p(w|x)p(x)}{p(a|x)p(x)} \cdot p(x)
\end{aligned}$$

Identification

Notice

$$
\begin{aligned}
\mathbb{E}(Y(a)) &= \sum_w \sum_x \sum_y y \cdot p(y|w,x,a,r=1) \cdot p(w|x,a) \cdot p(x) \\
&= \sum_w \sum_x \sum_y y \cdot p(y|w,x,a,r=1) \cdot \frac{p(a|x,w)p(w|x)p(x)}{p(a|x)p(x)} \cdot p(x) \\
&= \sum_w \sum_x \sum_y y \cdot p(y|w,x,a,r=1) \cdot \frac{p(a|x,w,r=1)p(w|x)}{\sum_{w'} p(a|w',x)p(w'|x)} \cdot p(x)
\end{aligned}
$$

Identification

Notice

$$\begin{aligned}\mathbb{E}(Y(a)) &= \sum_w \sum_x \sum_y y \cdot p(y|w,x,a,r=1) \cdot p(w|x,a) \cdot p(x) \\ &= \sum_w \sum_x \sum_y y \cdot p(y|w,x,a,r=1) \cdot \frac{p(a|x,w)p(w|x)p(x)}{p(a|x)p(x)} \cdot p(x) \\ &= \sum_w \sum_x \sum_y y \cdot p(y|w,x,a,r=1) \cdot \frac{p(a|x,w,r=1)p(w|x)}{\sum_{w'} p(a|w',x)p(w'|x)} \cdot p(x)\end{aligned}$$

Key idea: Use Bayes rule to re-express un-identified density in terms of identifiable ones

- Then, overall expression is identified – just need to define terms and "collapse" the above expectations

Plug-in implementation

Y_i	Y_i^*	A_i	A_i^*	$\boldsymbol{X}_i$	R_i	$\hat{\lambda}_a$	$\hat{\mu}_a$	$\hat{\pi}_a$	$\hat{\eta}_a$
1	0	0	1	3.2	1	NA	NA	NA	NA
0	1	1	1	2.1	1	NA	NA	NA	NA
1	1	0	0	0.3	1	NA	NA	NA	NA
1	0	0	1	1.6	0	NA	NA	NA	NA
0	1	0	1	4.8	0	NA	NA	NA	NA
1	1	0	0	0.9	0	NA	NA	NA	NA

Step 1: use validation data to fit $\hat{\lambda}_a$ and $\hat{\mu}_a$

Plug-in implementation

Y_i	Y_i^*	A_i	A_i^*	$\boldsymbol{X}_i$	R_i	$\hat{\lambda}_a$	$\hat{\mu}_a$	$\hat{\pi}_a$	$\hat{\eta}_a$
1	0	0	1	3.2	1	0.2	0.72	NA	NA
0	1	1	1	2.1	1	0.43	0.18	NA	NA
1	1	0	0	0.3	1	0.64	0.77	NA	NA
1	0	0	1	1.6	0	0.03	0.55	NA	NA
0	1	0	1	4.8	0	0.81	0.11	NA	NA
1	1	0	0	0.9	0	0.39	0.83	NA	NA

Step 1: use validation data to fit $\hat{\lambda}_a$ and $\hat{\mu}_a$

Plug-in implementation

Y_i	Y_i^*	A_i	A_i^*	$\boldsymbol{X}_i$	R_i	$\hat{\lambda}_a$	$\hat{\mu}_a$	$\hat{\pi}_a$	$\hat{\eta}_a$
1	0	0	1	3.2	1	0.2	0.72	`NA`	`NA`
0	1	1	1	2.1	1	0.43	0.18	`NA`	`NA`
1	1	0	0	0.3	1	0.64	0.77	`NA`	`NA`
1	0	0	1	1.6	0	0.03	0.55	`NA`	`NA`
0	1	0	1	4.8	0	0.81	0.11	`NA`	`NA`
1	1	0	0	0.9	0	0.39	0.83	`NA`	`NA`

Step 2: regress $\hat{\lambda}_a$ and $\hat{\mu}_a$ on $\boldsymbol{X}$ and A^* to yield $\hat{\pi}_a$ and $\hat{\eta}_a$

Plug-in implementation

Y_i	Y_i^*	A_i	A_i^*	$\boldsymbol{X}_i$	R_i	$\hat{\lambda}_a$	$\hat{\mu}_a$	$\hat{\pi}_a$	$\hat{\eta}_a$
1	0	0	1	3.2	1	0.2	0.72	0.77	0.23
0	1	1	1	2.1	1	0.43	0.18	0.41	0.26
1	1	0	0	0.3	1	0.64	0.77	0.59	0.80
1	0	0	1	1.6	0	0.03	0.55	0.08	0.53
0	1	0	1	4.8	0	0.81	0.11	0.72	0.18
1	1	0	0	0.9	0	0.39	0.83	0.38	0.79

Step 2: regress $\hat{\lambda}_a$ and $\hat{\mu}_a$ on $\boldsymbol{X}$ and A^* to yield $\hat{\pi}_a$ and $\hat{\eta}_a$

Plug-in implementation

Y_i	Y_i^*	A_i	A_i^*	$\boldsymbol{X}_i$	R_i	$\hat{\lambda}_a$	$\hat{\mu}_a$	$\hat{\pi}_a$	$\hat{\eta}_a$
1	0	0	1	3.2	1	0.2	0.72	0.77	0.23
0	1	1	1	2.1	1	0.43	0.18	0.41	0.26
1	1	0	0	0.3	1	0.64	0.77	0.59	0.80
1	0	0	1	1.6	0	0.03	0.55	0.08	0.53
0	1	0	1	4.8	0	0.81	0.11	0.72	0.18
1	1	0	0	0.9	0	0.39	0.83	0.38	0.79

Step 3: construct $\hat{\psi}_a^{\mathsf{PI}} = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\eta}_a(\boldsymbol{X}_i, A_i^*)}{\hat{\pi}_a(\boldsymbol{X}_i, A_i^*)}$

Equivalence of Approach 1 and Approach 2 EICs

$$\begin{aligned}\phi_a^{\mathsf{obs}}(\boldsymbol{O}) &= \frac{R}{\mathbb{P}(R=1|\boldsymbol{Z})}\chi_a^{\mathsf{full}}(\boldsymbol{O}) \\ &\quad - \left(\frac{R}{\mathbb{P}(R=1|\boldsymbol{Z})} - 1\right)\varphi_a(\boldsymbol{Z})\end{aligned}$$

Equivalence of Approach 1 and Approach 2 EICs

$$\phi_a^{\mathsf{obs}}(\boldsymbol{O}) = \frac{R}{\mathbb{P}(R=1|\boldsymbol{Z})}\left[\frac{I(A=a)}{e_a(\boldsymbol{X})}(Y - m_a(\boldsymbol{X})) + m_a(\boldsymbol{X}) - \psi_a\right]$$
$$- \left(\frac{R}{\mathbb{P}(R=1|\boldsymbol{Z})} - 1\right)\varphi_a(\boldsymbol{Z})$$

Equivalence of Approach 1 and Approach 2 EICs

$$\phi_a^{\mathsf{obs}}(\boldsymbol{O}) = \frac{R}{\mathbb{P}(R=1|\boldsymbol{Z})}\left[\frac{I(A=a)}{\pi_a(\boldsymbol{X})}\left(Y - \frac{\eta_a(\boldsymbol{X})}{\pi_a(\boldsymbol{X})}\right) + \frac{\eta_a(\boldsymbol{X})}{\pi_a(\boldsymbol{X})} - \psi_a\right]$$
$$- \left(\frac{R}{\mathbb{P}(R=1|\boldsymbol{Z})} - 1\right)\varphi_a(\boldsymbol{Z})$$

Equivalence of Approach 1 and Approach 2 EICs

$$\begin{aligned}
\phi_a^{\mathsf{obs}}(\boldsymbol{O}) &= \frac{R}{\mathbb{P}(R=1|\boldsymbol{Z})}\left[\frac{I(A=a)}{\pi_a(\boldsymbol{X})}\left(Y-\frac{\eta_a(\boldsymbol{X})}{\pi_a(\boldsymbol{X})}\right)+\frac{\eta_a(\boldsymbol{X})}{\pi_a(\boldsymbol{X})}-\psi_a\right] \\
&\quad -\left(\frac{R}{\mathbb{P}(R=1|\boldsymbol{Z})}-1\right)\left(\frac{\lambda_a(\boldsymbol{Z})\mu_a(\boldsymbol{Z})}{\pi_a(\boldsymbol{X})}-\frac{\lambda_a(\boldsymbol{Z})\eta_a(\boldsymbol{X})}{\pi_a(\boldsymbol{X})^2}+\frac{\eta_a(\boldsymbol{X})}{\pi_a(\boldsymbol{X})}-\psi_a\right)
\end{aligned}$$