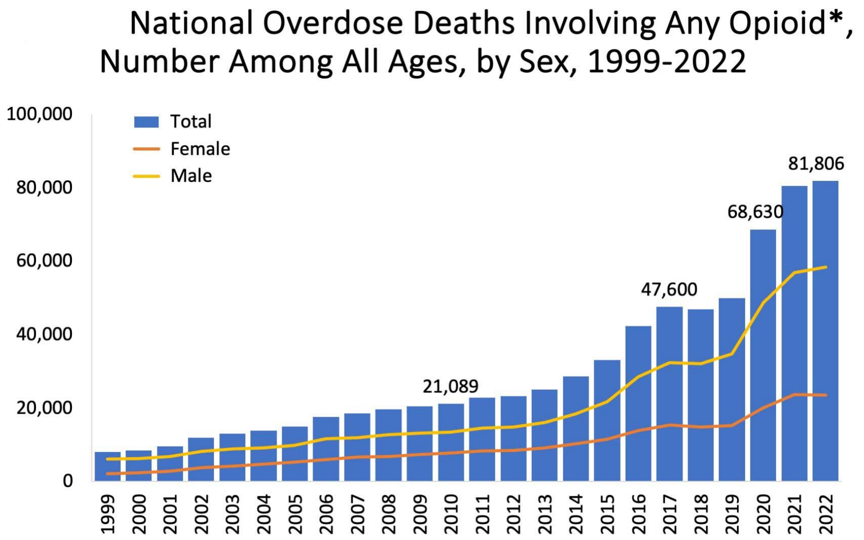
A Cautionary Tale on Integrating Studies with Disparate Outcome Measures for Causal Inference

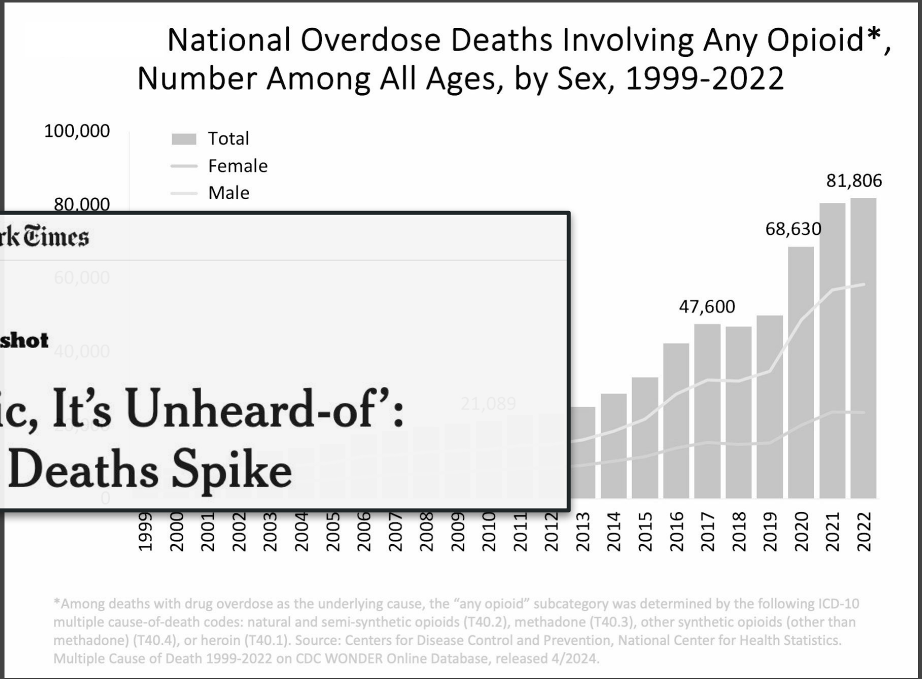
Harsh Parikh, Trang Nguyen, Elizabeth Stuart, Kara Rudolph, Caleb Miles

More than **80,000 deaths** in 2022 and 2023 were linked to **Opioids** in the US



*Among deaths with drug overdose as the underlying cause, the "any opioid" subcategory was determined by the following ICD-10 multiple cause-of-death codes: natural and semi-synthetic opioids (T40.2), methadone (T40.3), other synthetic opioids (other than methadone) (T40.4), or heroin (T40.1). Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Multiple Cause of Death 1999-2022 on CDC WONDER Online Database, released 4/2024.

More than **80,000 deaths** in 2022 and 2023 were linked to **Opioids** in the US




The New York Times

TheUpshot

'It's Huge, It's Historic, It's Unheard-of': Drug Overdose Deaths Spike

Medication such as **Buprenorphine** & **Naltrexone** are used to treat OUD

Medication such as **Buprenorphine** & **Naltrexone** are used to treat OUD

Major challenge: **Severe withdrawal symptoms**

Medication such as **Buprenorphine** & **Naltrexone** are used to treat OUD

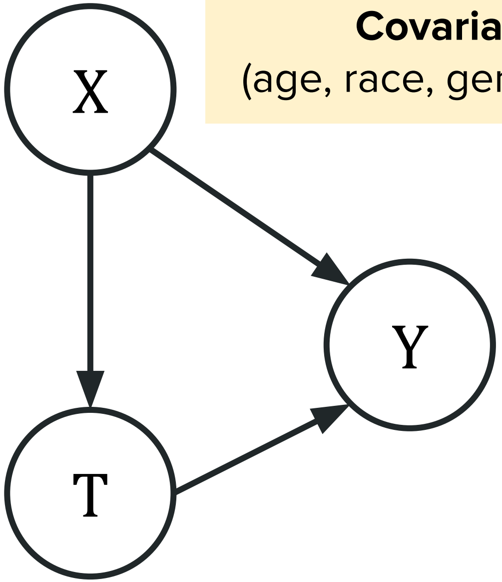
Major challenge: **Severe withdrawal symptoms**

Severity of **withdrawal** symptoms are *associated* with **OUD relapse**

Which of the medications for OUD result in least severe withdrawal symptoms? 🤕

X:BOT Study

Cohort Size: 540 patients



Pretreatment Covariates
(age, race, gender, etc.)

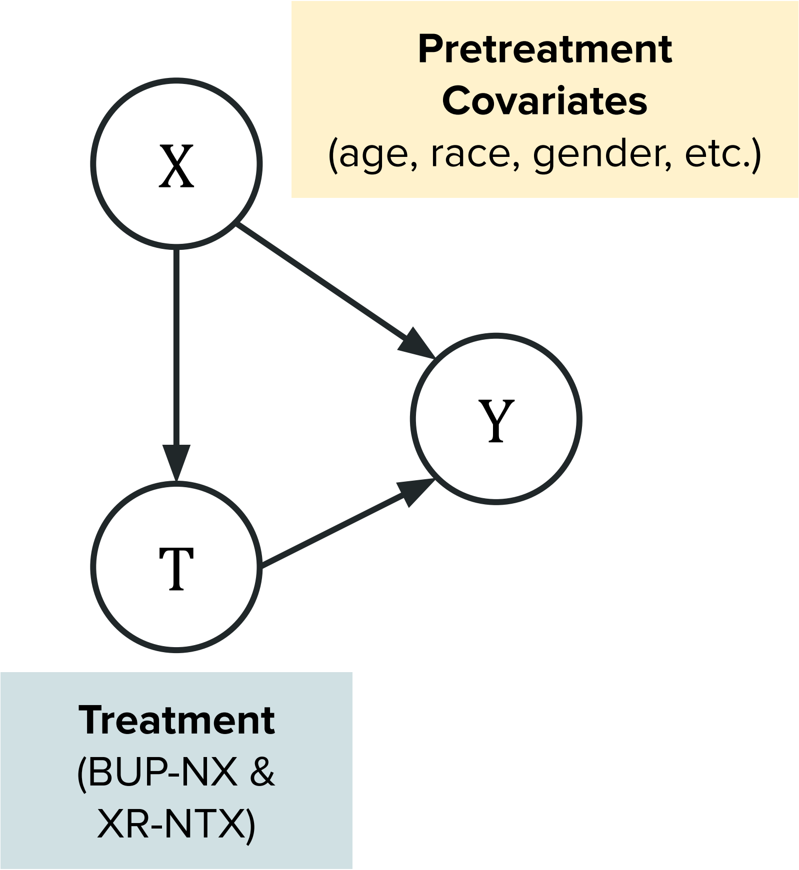
X:BOT Study

Cohort Size: 540 patients

Randomized (1:1) to receive:

T=0: Buprenorphine Naloxone (BUP-NX)

T=1: Extended Release Naltrexone (XR-NTX)



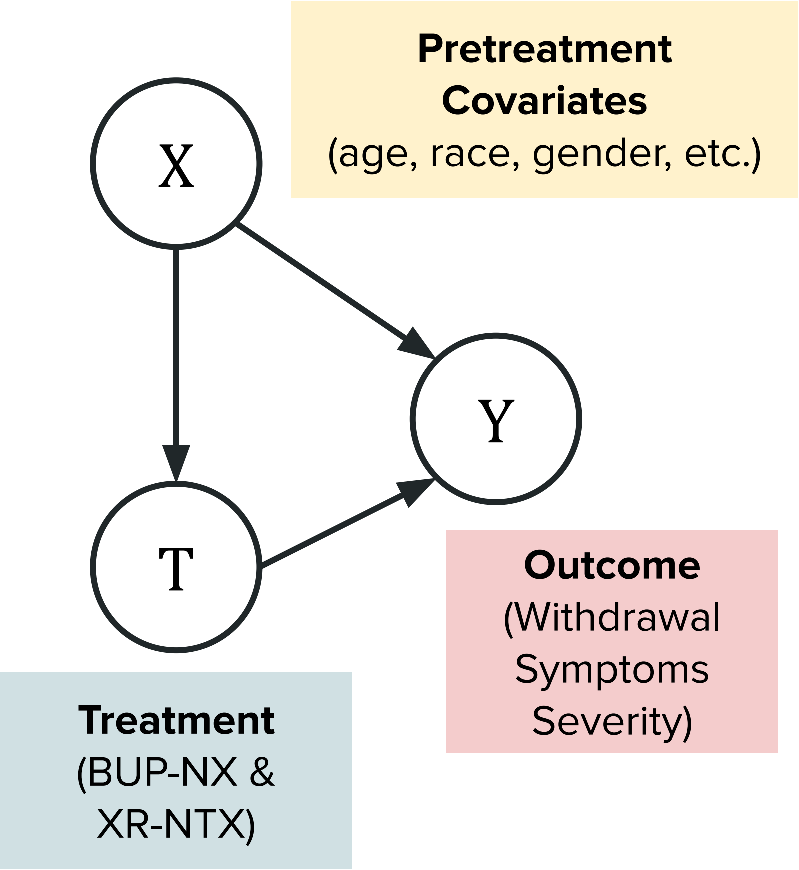
X:BOT Study

Cohort Size: 540 patients

Randomized (1:1) to receive:

> *T=0:* Buprenorphine Naloxone (BUP-NX)
>
> *T=1:* Extended Release Naltrexone (XR-NTX)

Outcome: Max. Subjective Opioid Withdrawal Scale (SOWS) Score b/w 10th and 14th Week



X:BOT Study

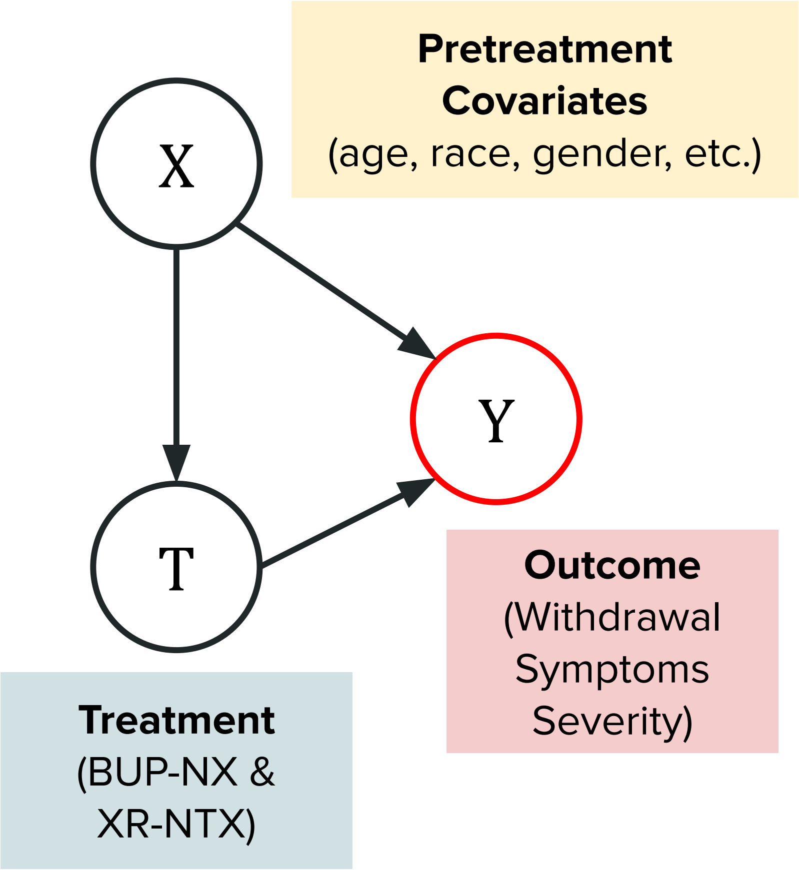
Cohort Size: 540 patients

Randomized (1:1) to receive:

T=0: Buprenorphine Naloxone (BUP-NX)

T=1: Extended Release Naltrexone (XR-NTX)

Outcome: Max. Subjective Opioid Withdrawal Scale (SOWS) Score b/w 10th and 14th Week



X:BOT Study

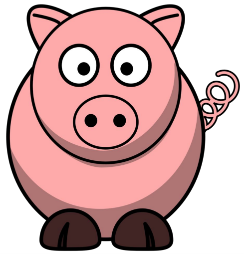
Cohort Size: 540 patients

Randomized (1:1) to receive:

T=0: Buprenorphine Naloxone (BUP-NX)

T=1: Extended Release Naltrexone (XR-NTX)

Outcome: Max. Subjective Opioid Withdrawal Scale (SOWS) Score b/w 10th and 14th Week

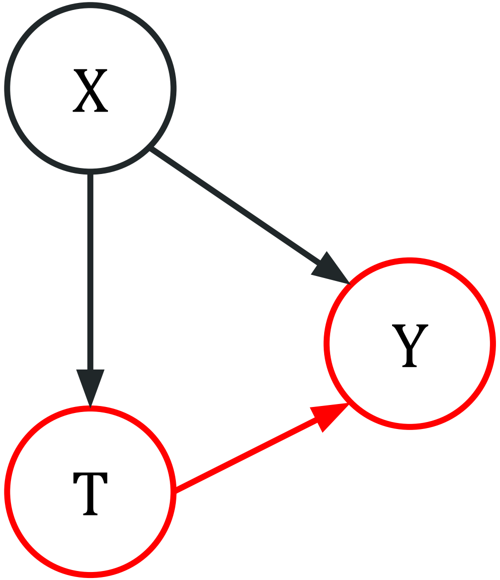


SOWS:

- Score measured using 16 symptoms such as nausea, vomiting, stomach cramp, etc.
- Ranges between 0 and 64
- Higher means worse

X:BOT Study

$$Y = \underbrace{\boxed{\theta(X)}}_{\text{Treatment Effect}} T + \underbrace{\boxed{g(X)}}_{\text{Baseline Effect}} + \gamma$$

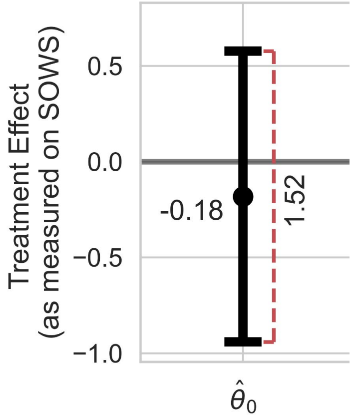


X:BOT Study

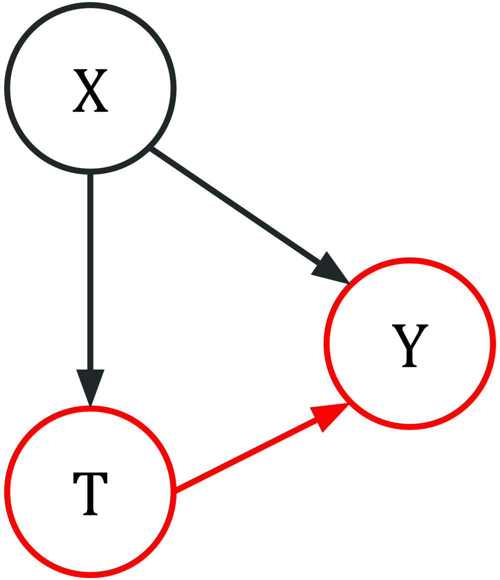
$$Y = \boxed{\theta(X)} T + \boxed{g(X)} + \gamma$$

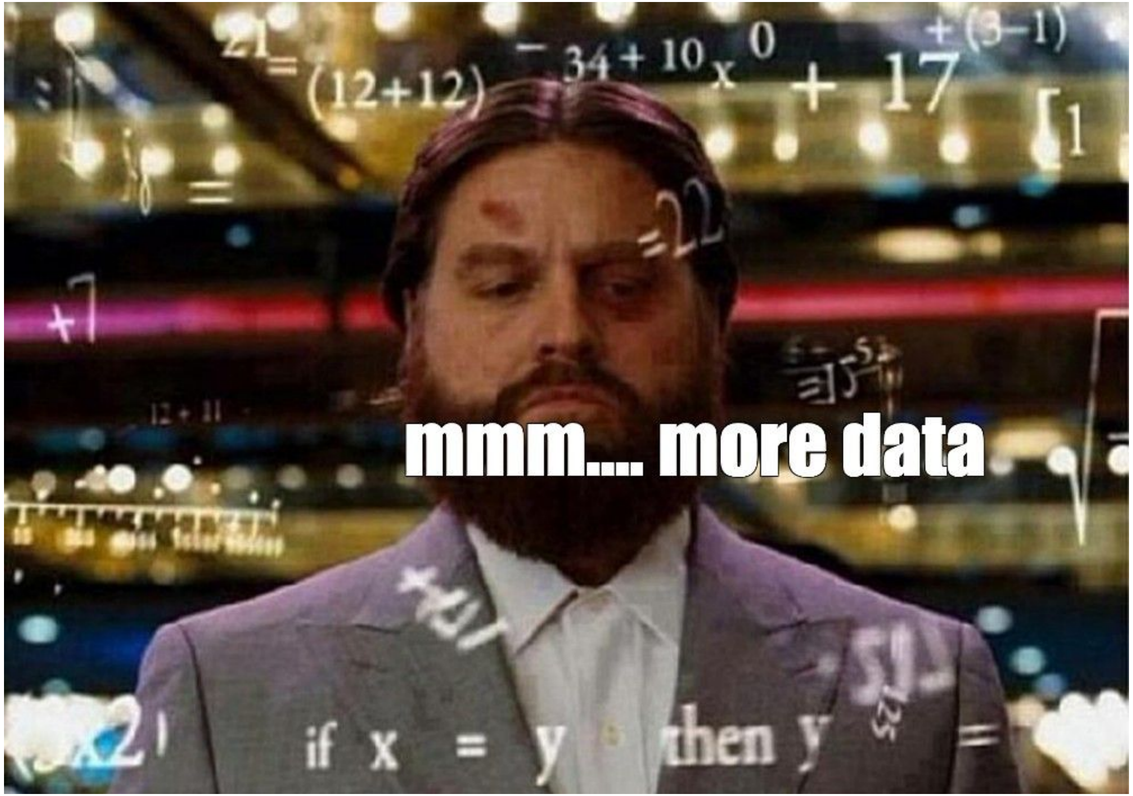
Treatment Effect

Baseline Effect



XR-NTX on-par with BUP-NX





Oh, we have another dataset!

POATS Study

Cohort Size: 360 patients

Treatment:

T=0: Buprenorphine Naloxone (BUP-NX)

Outcome: Max. Clinical Opiate Withdrawal Scale (COWS) Score b/w 10th and 14th Week

COWS:

- Measured using 11 symptoms such as sweating, pupil size, pulse rate, etc.
- Administered by clinician
- Ranges between 0 and 55
- Higher means worse

POATS Study

Coh

Trea

Out

With

10th

pulse

cian

nd 55

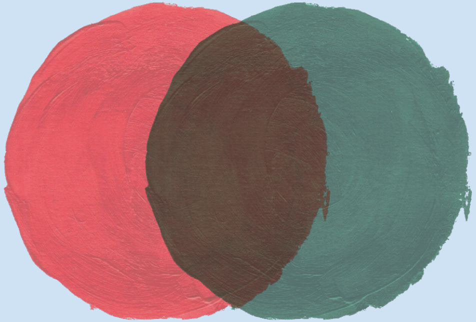
- Higher means worse

Can we integrate POATS data with X:BOT to *improve* the precision of our MOUD effect estimates?

Why Integrate Data Across Studies? 🧐

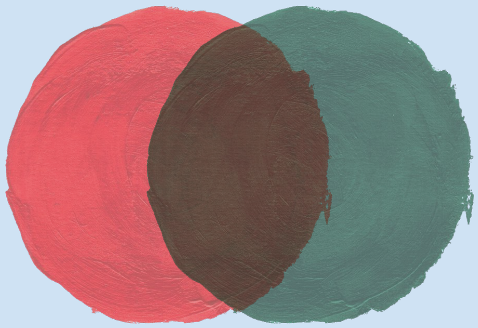
Data Integration & Causal Inference

- Combining datasets often collected under different study designs [Bariembom and Pearl (2016)]



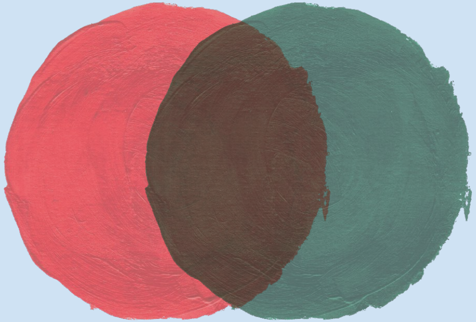
Data Integration & Causal Inference

- Combining datasets often collected under different study designs [Bariembom and Pearl (2016)]
- **Goals:**
    - Generalizability [Stuart et al (2011)]
    - Transportability [Pearl et al (2011)]
    - Efficiency Gain [Yang et al (2020)]
    - Bias/Error Correction [Kallus et al (2018), Parikh et al. (2024)]



Data Integration & Causal Inference

- Combining datasets often collected under different study designs [Bariembom and Pearl (2016)]
- **Goals:**
  - Generalizability [Stuart et al (2011)]
  - Transportability [Pearl et al (2011)]
  - Efficiency Gain [Yang et al (2020)]
  - Bias/Error Correction [Kallus et al (2018), Parikh et al. (2024)]



- Data fusion recognized as one of the *top 10 future directions for causal inference research* in a 2022 Commentary featured in the American Journal of Epidemiology (AJE).

Data Integration for *Efficiency Gain*

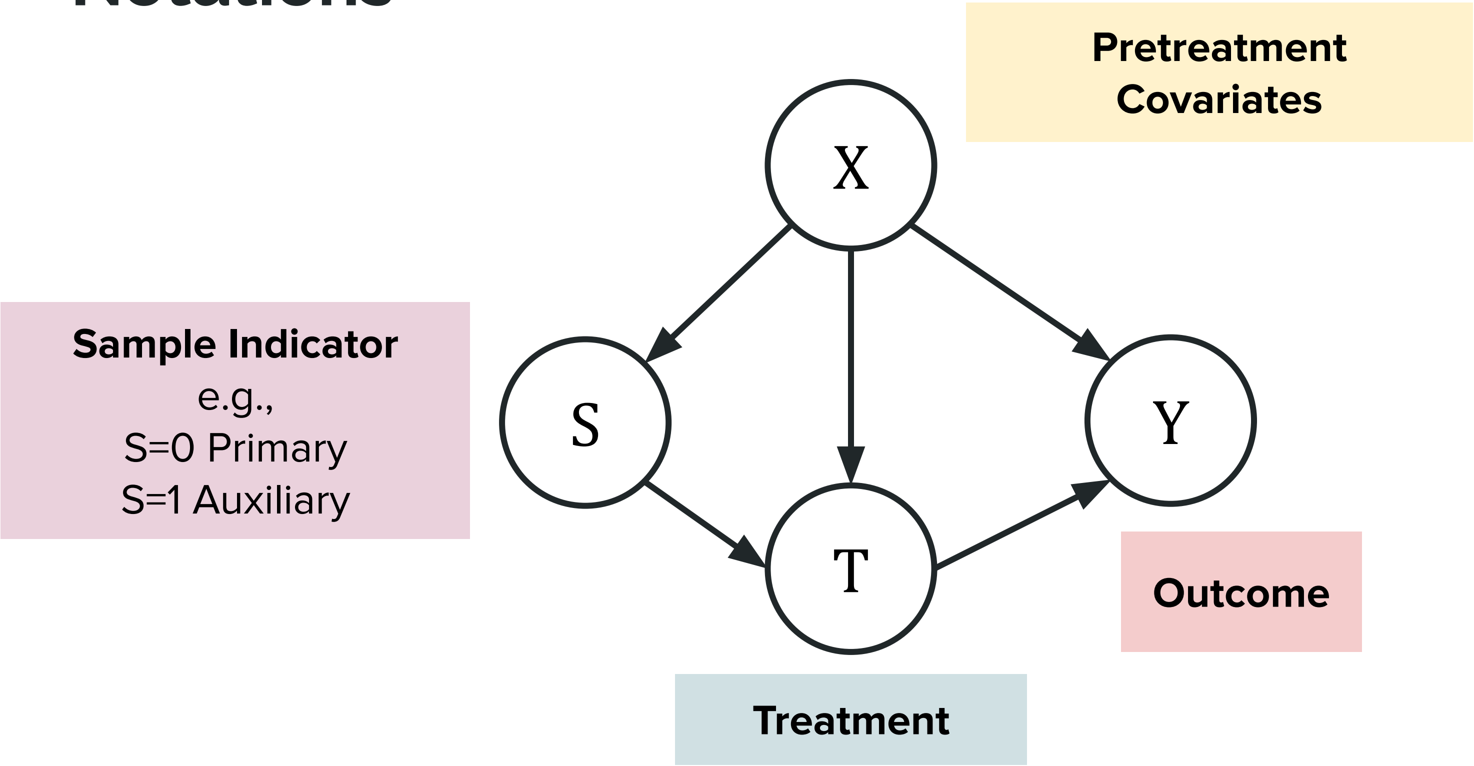
Primary

unit-id	X_1	...	X_p	T	Y
1	x_{11}	…	x_{1p}	t_1	y_1
2	x_{21}	…	x_{2p}	t_2	y_2
…	...	…	…	…	…
n_0	…	…	…	…	…

Auxiliary

unit-id	X_1	...	X_p	T	Y
n_0+1	$X_{n0 + 1, 1}$	…	$x_{n0+1,p}$	t_{n0+1}	y_{n0+1}
…	...	…	…	…	…
$n_0 + n_1$	…	…	…	…	…

Notations

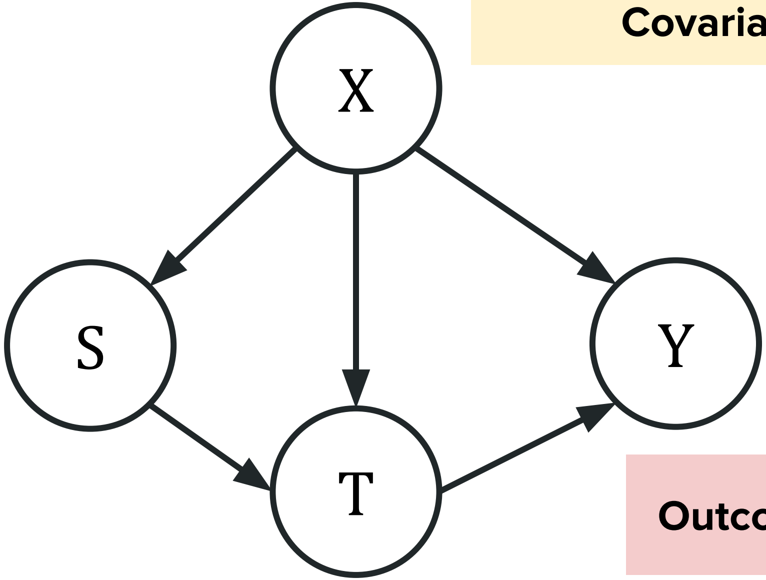


Pretreatment Covariates
X
Sample Indicator
e.g.,
S=0 Primary
S=1 Auxiliary
S
T
Y
Outcome
Treatment

Notations

Pretreatment Covariates

Sample Indicator
e.g.,
S=0 Primary
S=1 Auxiliary

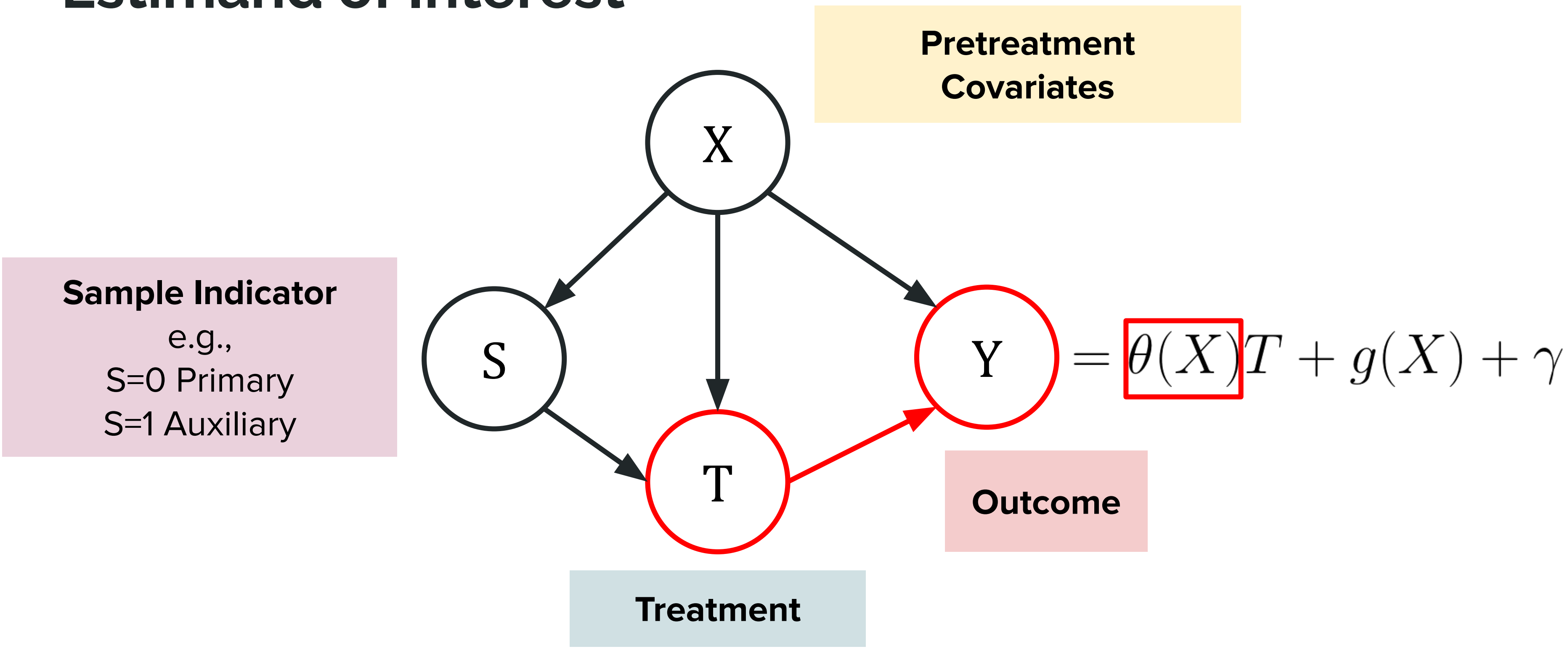


$$Y = \theta(X)T + g(X) + \gamma$$

Outcome

Treatment

Estimand of Interest



Efficiency Gain

$\hat{\theta}_{D_0}$: Estimator using only Primary Data

Efficiency Gain

$\hat{\theta}_{D_0}$: Estimator using only Primary Data

Quantifying Uncertainty: $\mathbb{E}[(\hat{\theta}_{D_0} - \theta)^2] = Var(\hat{\theta}_{D_0})$

Efficiency Gain

$\hat{\theta}_{D_0}$: Estimator using only Primary Data

Quantifying Uncertainty: $\mathbb{E}[(\hat{\theta}_{D_0} - \theta)^2] = Var(\hat{\theta}_{D_0})$

Efficient Estimator: $\hat{\theta}_{D_0}$ such that $Var(\hat{\theta}_{D_0})$ is smallest

Efficiency Gain

$\hat{\theta}_{D_0}$: Estimator using only Primary Data

Quantifying Uncertainty: $\mathbb{E}[(\hat{\theta}_{D_0} - \theta)^2] = Var(\hat{\theta}_{D_0})$

Efficient Estimator: $\hat{\theta}_{D_0}$ such that $Var(\hat{\theta}_{D_0})$ is smallest

$\hat{\theta}_{D_0 \bigoplus D_1}$: Efficient Estimator using both Primary and Auxiliary data

Efficiency Gain

$\hat{\theta}_{D_0}$: Estimator using only Primary Data

Quantifying Uncertainty: $\mathbb{E}[(\hat{\theta}_{D_0} - \theta)^2] = Var(\hat{\theta}_{D_0})$

Efficient Estimator: $\hat{\theta}_{D_0}$ such that $Var(\hat{\theta}_{D_0})$ is smallest

$\hat{\theta}_{D_0 \bigoplus D_1}$: Efficient Estimator using both Primary and Auxiliary data

Consider the efficient estimator $\hat{\theta}_{D_0}$ and $\hat{\theta}_{D_0 \bigoplus D_1}$

Efficiency Gain

$\hat{\theta}_{D_0}$: Estimator using only Primary Data

Quantifying Uncertainty:$\mathbb{E}[(\hat{\theta}_{D_0} - \theta)^2] = Var(\hat{\theta}_{D_0})$

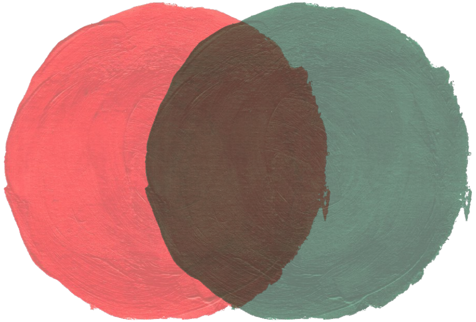
Efficient Estimator: $\hat{\theta}_{D_0}$ such that $Var(\hat{\theta}_{D_0})$is smallest

$\hat{\theta}_{D_0 \bigoplus D_1}$: Efficient Estimator using both Primary and Auxiliary data

Consider the efficient estimator $\hat{\theta}_{D_0}$ and $\hat{\theta}_{D_0 \bigoplus D_1}$

Efficiency Gain : $Var(\hat{\theta}_{D_0 \bigoplus D_1}) < Var(\hat{\theta}_{D_0})$

But we have Disparate Outcomes Measures 😏



Studies with Disparate Outcome Measure

Primary

unit-id	X_1	…	X_p	T	Y	W
1	x_{11}	…	x_{1p}	t_1	y_1	?
2	x_{21}	…	x_{2p}	t_2	y_2	?
…	...	…	…	…	…	?
n_0	…	…	…	…	…	?

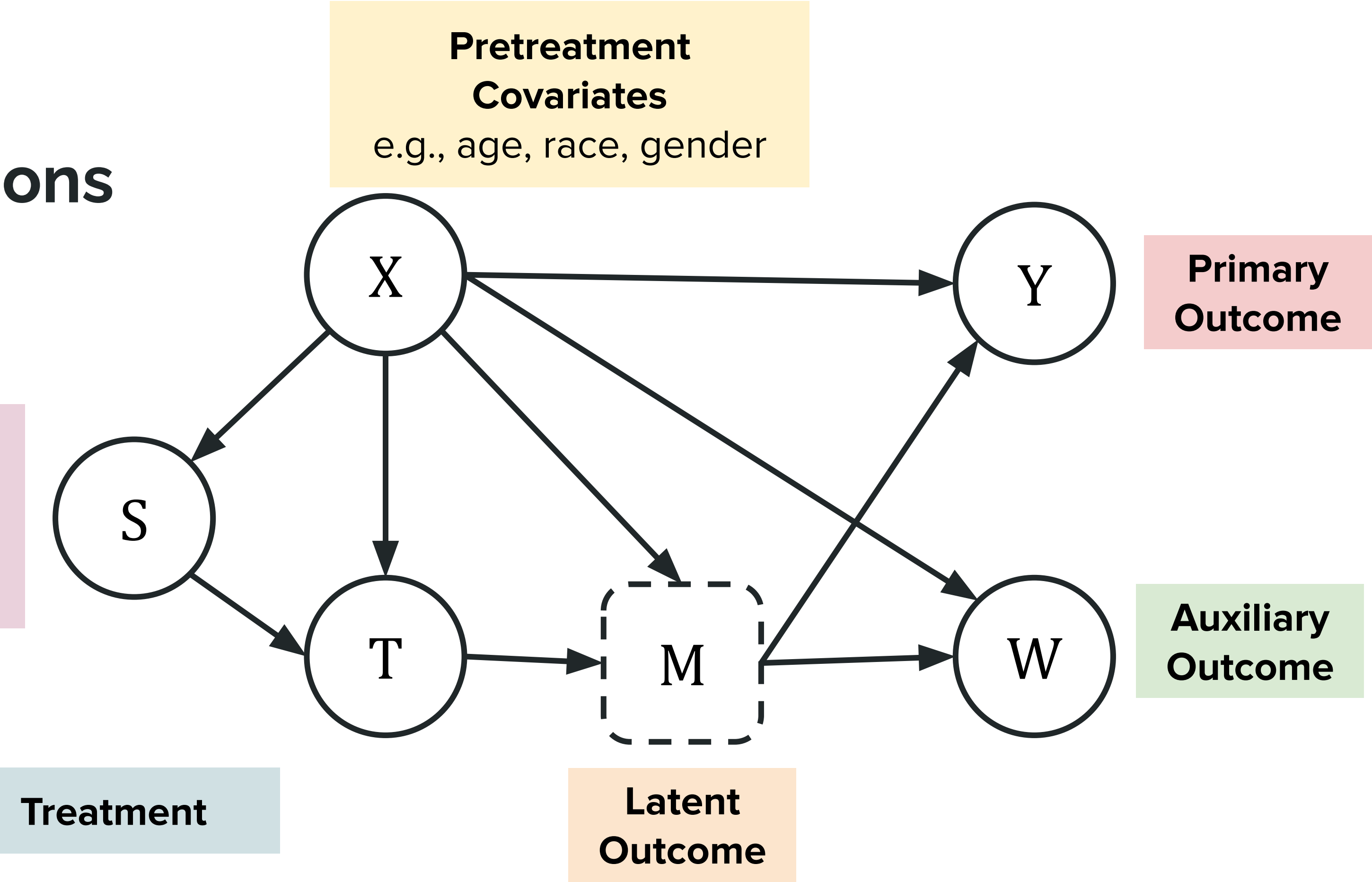
Auxiliary

unit-id	X_1	…	X_p	T	Y	W
n_0+1	$X_{n0 + 1, 1}$	…	$x_{n0+1,p}$	t_{n0+1}	?	w_{n0+1}
…	...	…	…	…	?	…
$n_0 + n_1$	…	…	…	…	?	…

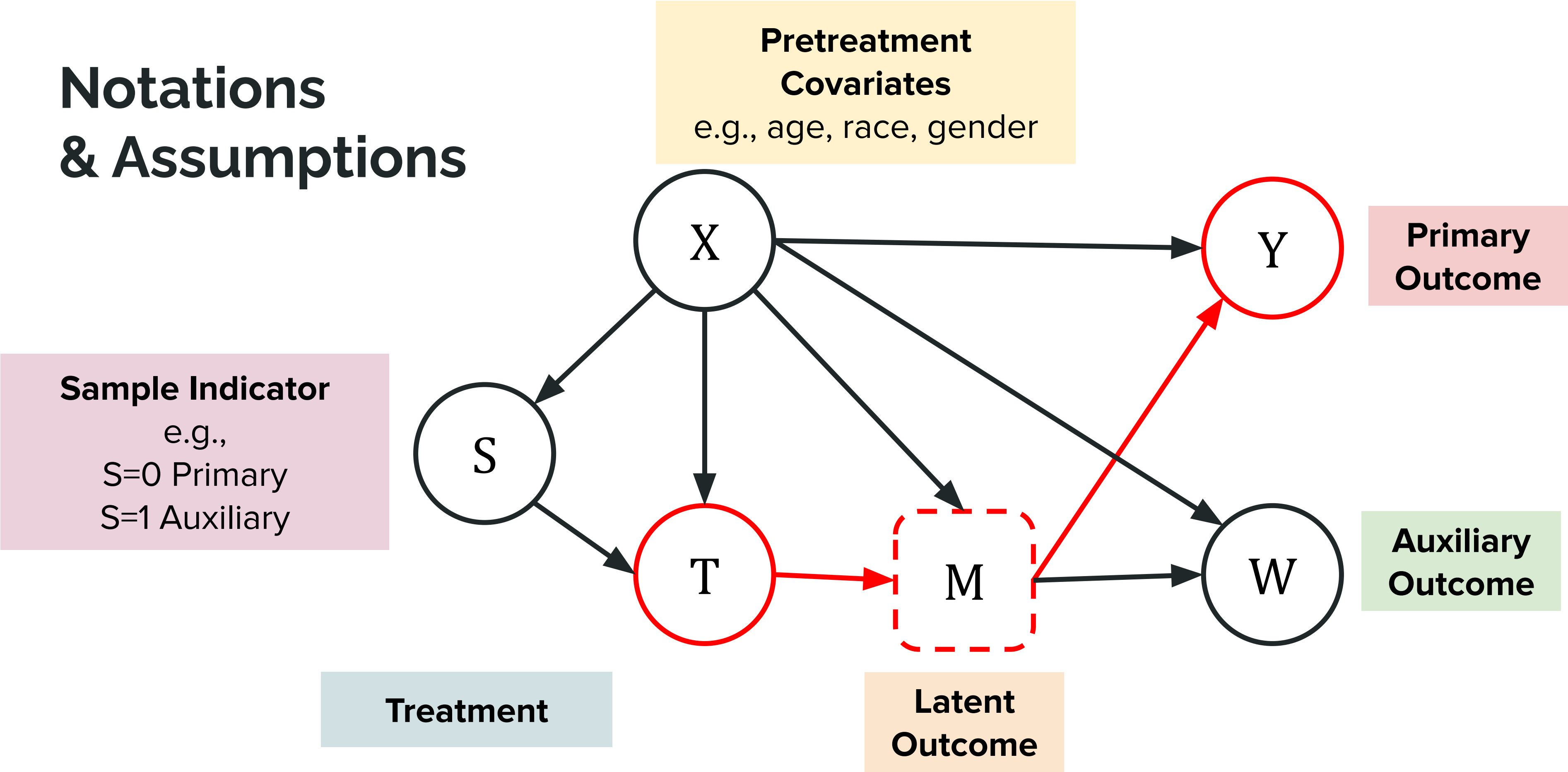
Some Related Literature

- Estimating Long-term Treatment Effects
  - Athey et al (2019), Ghassami et al. (2022)
- Leveraging Surrogate / Proxy Outcomes
  - Wang et al. (2020)
- Correcting for Measurement Errors
  - Ross et al. (2024)
- Meta Learning with Disparate Outcomes
  - Deeks et al. (2019)

Notations & Assumptions



Notations & Assumptions



Structural Assumptions

$$Y = \theta(X)T + g(X) + \gamma$$

$$W = \phi(X)T + f(X) + \delta$$

Estimand of Interest

$$Y = \boxed{\theta(X)}T + g(X) + \gamma$$

$$W = \phi(X)T + f(X) + \delta$$

Recall:

$$Y = \theta(X)T + g(X) + \gamma$$

$$W = \phi(X)T + f(X) + \delta$$

Q: If and when can leveraging auxiliary data with disparate outcome measure yield *efficiency gain*?

$$Y = \theta(X)T + g(X) + \gamma$$

$$W = \phi(X)T + f(X) + \delta$$

Linking Primary & Auxiliary Outcomes

$$\mathbb{E}[Y \mid X] = \alpha(X)\mathbb{E}[W \mid X] + \beta(X)$$

$$Y = \theta(X)T + g(X) + \gamma$$

$$W = \phi(X)T + f(X) + \delta$$

Linking Primary & Auxiliary Outcomes

$$\mathbb{E}[Y \mid X] = \alpha(X)\mathbb{E}[W \mid X] + \beta(X)$$

Assumptions (from strongest to weakest)

$$Y = \theta(X)T + g(X) + \gamma$$

$$W = \phi(X)T + f(X) + \delta$$

Linking Primary & Auxiliary Outcomes

$$\mathbb{E}[Y \mid X] = \alpha(X)\mathbb{E}[W \mid X] + \beta(X)$$

Assumptions (from strongest to weakest)

A.a. α and β are a priori known

Biochemical systems with substantial prior research. Mechanistic parameters connecting intermediate and long term outcomes

$$Y = \theta(X)T + g(X) + \gamma$$

$$W = \phi(X)T + f(X) + \delta$$

Linking Primary & Auxiliary Outcomes

$$\mathbb{E}[Y \mid X] = \alpha(X)\mathbb{E}[W \mid X] + \beta(X)$$

Assumptions (from strongest to weakest)

A.a. α and β are a priori known

A.b. only β is a priori is known; α unknown

Medical outcomes measured using different scales. Baseline levels are well-known (typically zero). Heterogeneous scaling factors (α) often unknown.

$$Y = \theta(X)T + g(X) + \gamma$$

$$W = \phi(X)T + f(X) + \delta$$

Linking Primary & Auxiliary Outcomes

$$\mathbb{E}[Y \mid X] = \alpha(X)\mathbb{E}[W \mid X] + \beta(X)$$

Assumptions (from strongest to weakest)

A.a. α and β are a priori known

A.b. only β is a priori is known; α unknown

A.c. both α and β unknown

Almost every other scenario!

Theoretical Findings

$$\boxed{Var(\hat{\theta}_0) = Var(\hat{\theta}_c) \approx Var(\hat{\theta}_b) > Var(\hat{\theta}_a)}$$

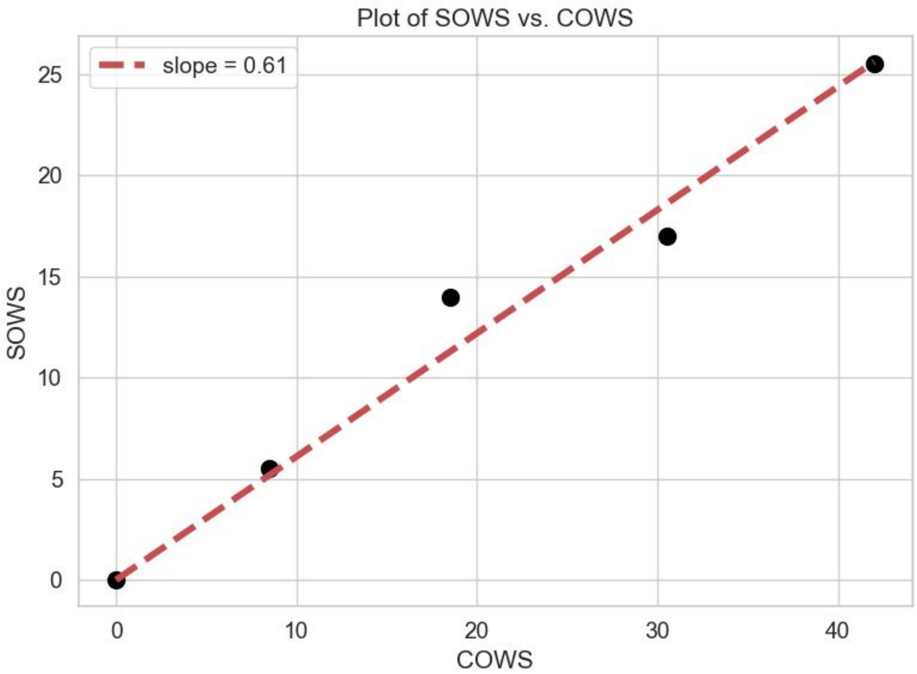
$$\mathbb{E}[Y \mid X] = \alpha(X)\mathbb{E}[W \mid X] + \beta(X)$$

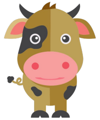
Assumptions (from strongest to weakest)
A.a. α and β are a priori known
A.b. only β is a priori is known; α unknown
A.c. both α and β unknown

X:BOT ⊕ POATS

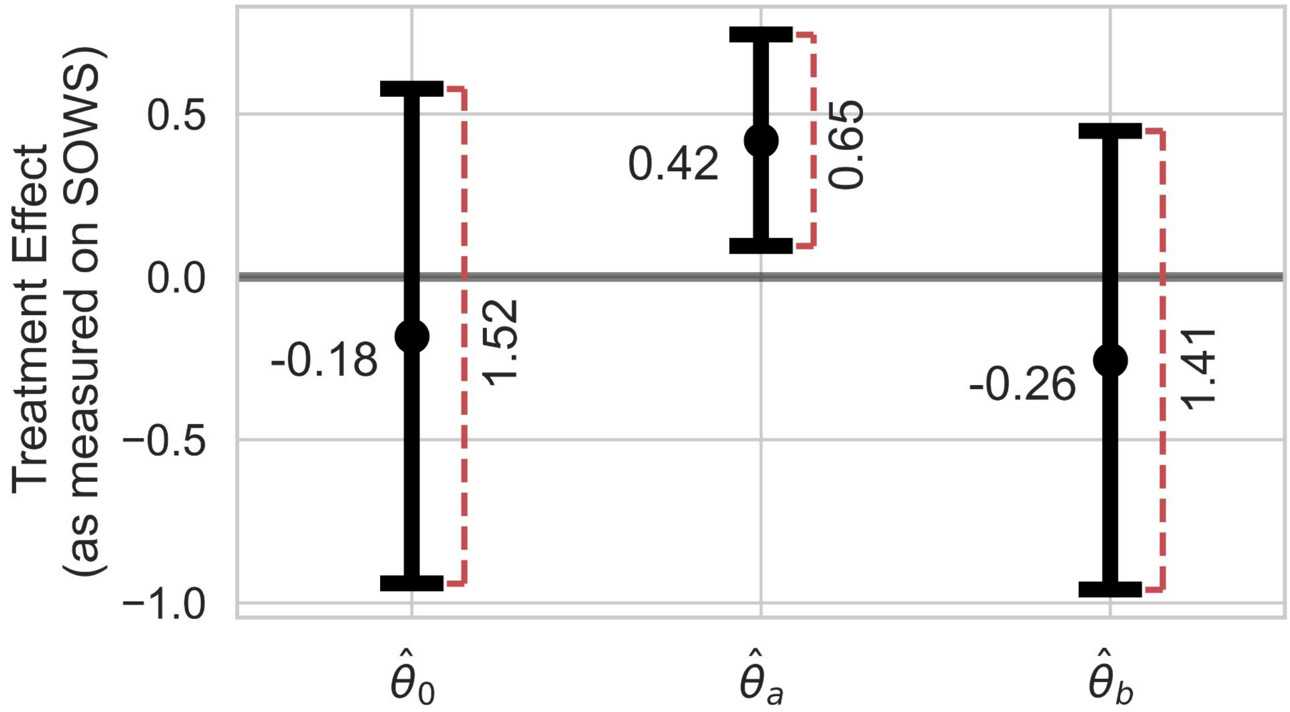
MOUD → Withdrawal Symptoms

COWS v/s SOWS (Building on domain knowledge)





X:BOT ⊕ POATS

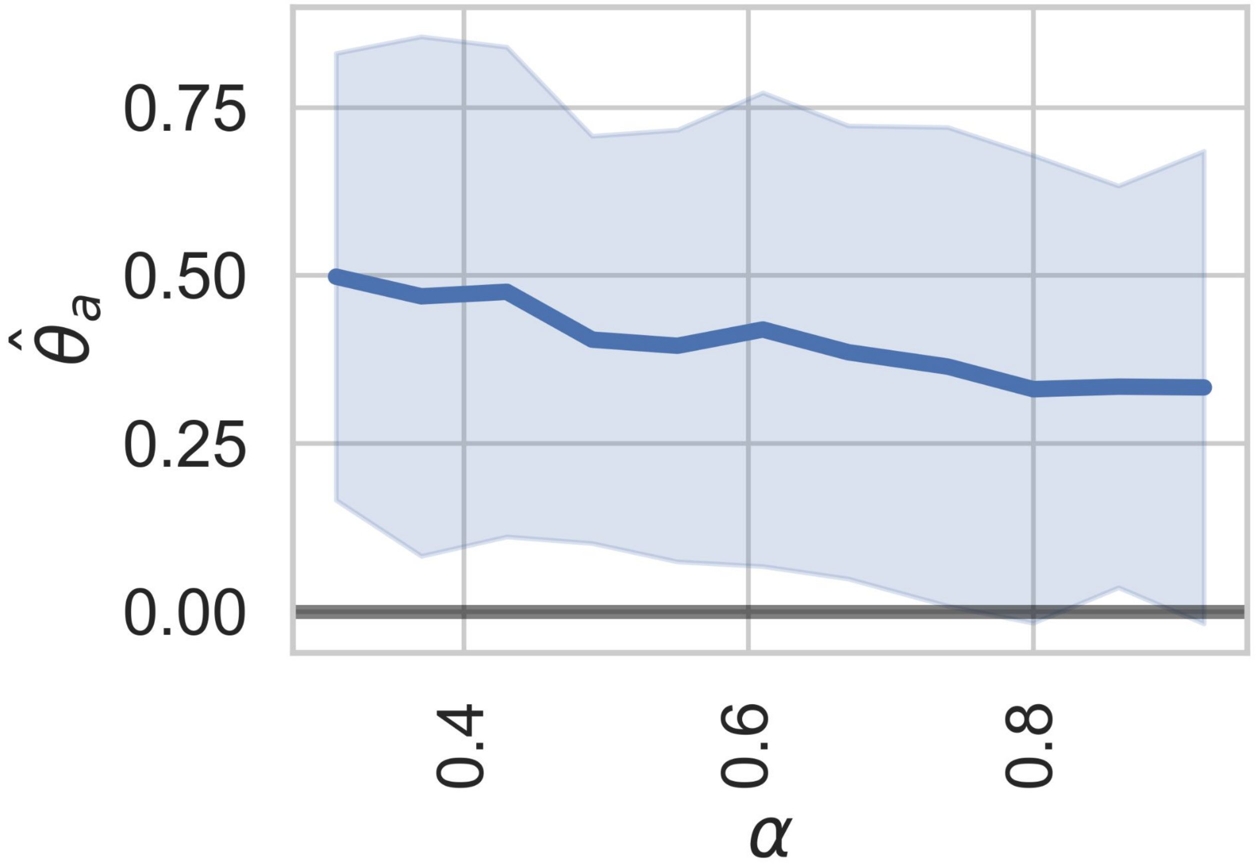


Treatment Effect
(as measured on SOWS)
0.5
0.0
−0.5
−1.0
-0.18
1.52
0.42
0.65
-0.26
1.41
$\hat{\theta}_0$
$\hat{\theta}_a$
$\hat{\theta}_b$

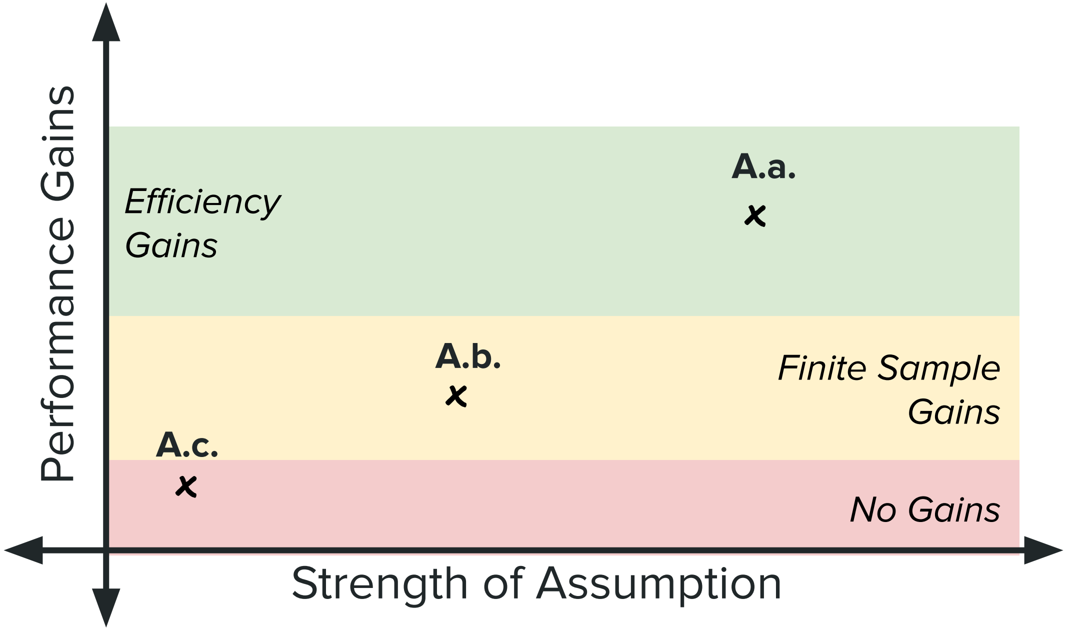
What if we assume A.a. *but* our guess of $\alpha(X)$ is wrong? 😨

What if we assume A.a. *but* our guess of α(X) is wrong? 😨

$$Bias = \mathbb{E}[(\alpha_{mis.}(X) - \alpha_{true}(X))\,\phi(X)]$$



0.75
0.50
0.25
0.00
$\hat{\theta}_a$
0.4
0.6
0.8
α



$$\mathbb{E}[Y \mid X] = \alpha(X)\mathbb{E}[W \mid X] + \beta(X)$$

Assumptions (from strongest to weakest)

A.a. α and β are a priori known

A.b. only β is a priori is known; α unknown

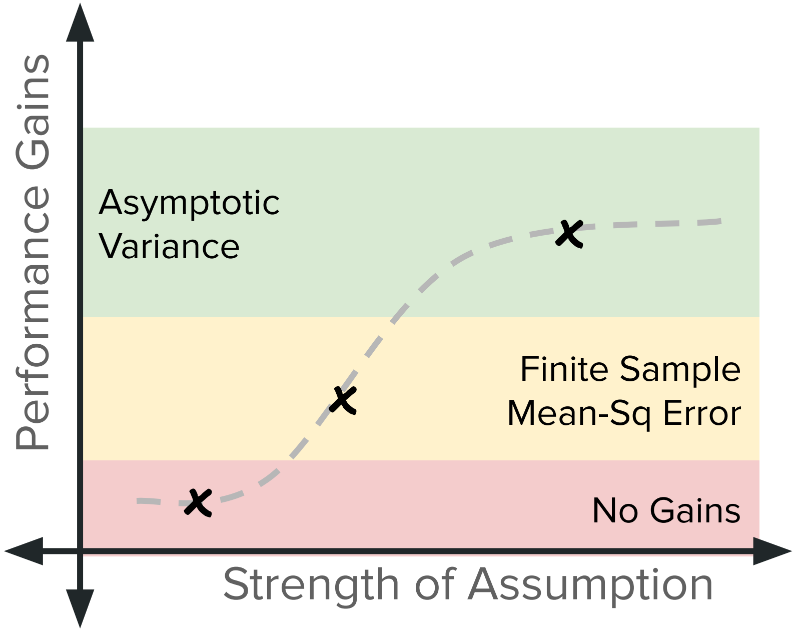
A.c. both α and β unknown

Q: Does Leveraging Auxiliary Study with Disparate Outcome yields Efficiency Gains?

A: It depends on

- Access to background knowledge
- Assumptions one is willing to make
    - Risk of bias due to misspecification

>> *There is no free lunch*


Performance Gains
Asymptotic
Variance
Finite Sample
Mean-Sq Error
No Gains
Strength of Assumption

Thank you!