# PLSC 30600

## Week 3: MAR estimation, reweighting, and imputation

Molly Offer-Westort

Department of Political Science,
University of Chicago

Winter 2026

# Science table: MAR and regression estimation

| $i$ | $X_{[1]i}$ | $X_{[2]i}$ | $Y_i(0)$ | $R_i$ | $Y_i^*(0)$ |
|---|---|---|---|---|---|
| 1 | A | 0 | 0 | 1 | 0 |
| 2 | A | 0 | ? | 0 | -99 |
| 3 | B | 0 | 1 | 1 | 1 |
| 4 | B | 0 | ? | 0 | -99 |
| 5 | A | 1 | 0 | 1 | 0 |
| 6 | A | 1 | ? | 0 | -99 |
| 7 | B | 1 | 1 | 1 | 1 |
| 8 | B | 1 | ? | 0 | -99 |
| 9 | A | 0 | 1 | 1 | 1 |
| 10 | B | 1 | ? | 0 | -99 |

# Regression estimation under MAR

- Under MAR, for all $x \in \mathrm{Supp}\,[X_i]$:
$\mathrm{E}\,[Y_i \mid R_i = 0, \boldsymbol{X}_i = \boldsymbol{x}] = \mathrm{E}\,[Y_i^* \mid R_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] = \mathrm{E}\,[Y_i \mid \boldsymbol{X}_i = \boldsymbol{x}]$.

- In our case, with two covariates by MAR:
$$\mathrm{E}\,[Y_i \mid X_{[1]i}, X_{[2]i}] = \mathrm{E}\,[Y_i^* \mid R_i = 1, X_{[1]i}, X_{[2]i}].$$

- Assume a functional form for the CEF, e.g.
$$\mathrm{E}\,[Y_i \mid X_{[1]i}, X_{[2]i}] = \beta_0 + \beta_1 1\{X_{[1]i} = B\} + \beta_2 X_{[2]i}.$$

- Use regression to estimate, predict for all $i$, then average.

# Science table: MAR and regression estimation

| $i$ | $X_{[1]i}$ | $X_{[2]i}$ | $Y_i(0)$ | $R_i$ | $Y_i^*(0)$ |
|---|---|---|---|---|---|
| 1 | A | 0 | 0 | 1 | 0 |
| 2 | A | 0 | ? | 0 | $\hat{\mathrm{E}}[Y_i \mid X_{[1]i} = x_{[}1], X_{[2]i} = x_{[}2]]$ |
| 3 | B | 0 | 1 | 1 | 1 |
| 4 | B | 0 | ? | 0 | $\hat{\mathrm{E}}[Y_i \mid X_{[1]i} = x_{[}1], X_{[2]i} = x_{[}2]]$ |
| 5 | A | 1 | 0 | 1 | 0 |
| 6 | A | 1 | ? | 0 | $\hat{\mathrm{E}}[Y_i \mid X_{[1]i} = x_{[}1], X_{[2]i} = x_{[}2]]$ |
| 7 | B | 1 | 1 | 1 | 1 |
| 8 | B | 1 | ? | 0 | $\hat{\mathrm{E}}[Y_i \mid X_{[1]i} = x_{[}1], X_{[2]i} = x_{[}2]]$ |
| 9 | A | 0 | 1 | 1 | 1 |
| 10 | B | 1 | ? | 0 | $\hat{\mathrm{E}}[Y_i \mid X_{[1]i} = x_{[}1], X_{[2]i} = x_{[}2]]$ |

# Science table: MAR regression plug-in

| $i$ | $X_{[1]i}$ | $X_{[2]i}$ | $Y_i(0)$ | $R_i$ | $Y_i^*(0)$ |
|---|---|---|---|---|---|
| 1 | A | 0 | 0 | 1 | 0 |
| 2 | A | 0 | ? | 0 | $\hat{\beta}_0$ |
| 3 | B | 0 | 1 | 1 | 1 |
| 4 | B | 0 | ? | 0 | $\hat{\beta}_0 + \hat{\beta}_2$ |
| 5 | A | 1 | 0 | 1 | 0 |
| 6 | A | 1 | ? | 0 | $\hat{\beta}_0 + \hat{\beta}_3$ |
| 7 | B | 1 | 1 | 1 | 1 |
| 8 | B | 1 | ? | 0 | $\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_3$ |
| 9 | A | 0 | 1 | 1 | 1 |
| 10 | B | 1 | ? | 0 | $\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_3$ |

$$\mathrm{E}\left[Y_i \mid X_{[1]i}, X_{[2]i}\right] = \beta_0 + \beta_2 1\{X_{[1]i} = B\} + \beta_3 X_{[2]i}.$$

# Code: regression plug-in under MAR

```
> df <- data.frame(
+   X_1 = c("A","A","B","B","A","A","B","B","A","B"),
+   X_2 = c(0,0,0,0,1,1,1,1,0,1),
+   R   = c(1,0,1,0,1,0,1,0,1,0),
+   Ystar = c(0,-99,1,-99,0,-99,1,-99,1,-99)
+ )
> df$X_1 <- factor(df$X_1)
> (fit <- lm(Ystar ~ X_1 + X_2, data = df, subset = R == 1))

Call:
lm(formula = Ystar ~ X_1 + X_2, data = df, subset = R == 1)

Coefficients:
(Intercept)         X_1B          X_2
     0.4286       0.7143      -0.2857
```

# Code: regression plug-in under MAR

```
> df$yhat <- df$Ystar
> df$yhat[which(df$R == 0)] <-
+   predict(fit, newdata = df[which(df$R == 0), ])
> round(df$yhat, 3)

 [1] 0.000 0.429 1.000 1.143 0.000 0.143 1.000 0.857 1.000 0.857

> mean(df$yhat)

[1] 0.6428571
```

# Science table: strong ignorability and regression estimation

| $i$ | $X_{[1]i}$ | $X_{[2]i}$ | $Y_i(0)$ | $Y_i(1)$ | $D_i$ | $Y_i$ |
|---|---|---|---|---|---|---|
| 1 | A | 0 | 0 | ? | 0 | 0 |
| 2 | A | 0 | ? | 1 | 1 | 1 |
| 3 | B | 0 | 1 | ? | 0 | 1 |
| 4 | B | 0 | ? | 1 | 1 | 1 |
| 5 | A | 1 | 0 | ? | 0 | 0 |
| 6 | A | 1 | ? | 1 | 1 | 1 |
| 7 | B | 1 | 1 | ? | 0 | 1 |
| 8 | B | 1 | ? | 0 | 1 | 0 |
| 9 | A | 0 | 1 | ? | 0 | 0 |
| 10 | B | 1 | ? | 1 | 1 | 1 |

# Regression estimation under strong ignorability

- Under strong ignorability, for all $x \in \mathrm{Supp}\left[X_i\right]$:

$$\mathrm{E}\left[Y_i(d) \mid X_i = x\right] = \mathrm{E}\left[Y_i \mid D_i = d, X_i = x\right], \quad d \in \{0,1\}.$$

- We need a treatment indicator in the regression model.

- Example CEF specification:

$$\mathrm{E}\left[Y_i \mid D_i, X_{[1]i}, X_{[2]i}\right] = \beta_0 + \beta_1 D_i + \beta_2 1\{X_{[1]i} = B\} + \beta_3 X_{[2]i}.$$

- Predict $\widehat{Y}_i(0)$ and $\widehat{Y}_i(1)$, then average differences.

# Definition: regression estimator for causal inference

## Definition (Regression Estimator for Causal Inference)

Let $Y_i(0)$, $Y_i(1)$, and $D_i$ be random variables with $\mathrm{Supp}[D_i] = \{0, 1\}$. Let $Y_i = Y_i(1) \cdot D_i + Y_i(0) \cdot (1 - D_i)$ and $\tau_i = Y_i(1) - Y_i(0)$, and let $\boldsymbol{X}_i$ be a random vector. Given $n$ i.i.d. observations of $(Y_i, D_i, \boldsymbol{X}_i)$, the regression estimator for $\mathrm{E}[\tau_i]$ is

$$\widehat{\mathrm{E}}[\tau_i] = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathrm{E}}[Y_i \mid D_i = 1, \boldsymbol{X}_i] \; - \; \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathrm{E}}[Y_i \mid D_i = 0, \boldsymbol{X}_i],$$

where $\widehat{\mathrm{E}}[Y_i \mid D_i = d, \boldsymbol{X}_i = \boldsymbol{x}]$ is an estimator of the CEF.

# Science table: strong ignorability and regression imputation

| $i$ | $X_{[1]i}$ | $X_{[2]i}$ | $Y_i(0)$ | $Y_i(1)$ | $D_i$ | $Y_i$ |
|---|---|---|---|---|---|---|
| 1 | A | 0 | 0 | $\hat{m}_1(x)$ | 0 | 0 |
| 2 | A | 0 | $\hat{m}_0(x)$ | 1 | 1 | 1 |
| 3 | B | 0 | 1 | $\hat{m}_1(x)$ | 0 | 1 |
| 4 | B | 0 | $\hat{m}_0(x)$ | 1 | 1 | 1 |
| 5 | A | 1 | 0 | $\hat{m}_1(x)$ | 0 | 0 |
| 6 | A | 1 | $\hat{m}_0(x)$ | 1 | 1 | 1 |
| 7 | B | 1 | 1 | $\hat{m}_1(x)$ | 0 | 1 |
| 8 | B | 1 | $\hat{m}_0(x)$ | 0 | 1 | 0 |
| 9 | A | 0 | 1 | $\hat{m}_1(x)$ | 0 | 0 |
| 10 | B | 1 | $\hat{m}_0(x)$ | 1 | 1 | 1 |

$$\hat{m}_d(x) = \hat{\mathrm{E}}\left[Y_i \mid D_i = d, X_{[1]i} = x_{[1]}, X_{[2]i} = x_{[2]}\right], \quad d \in \{0, 1\}.$$

# Science table: strong ignorability regression plug-in

| $i$ | $X_{[1]i}$ | $X_{[2]i}$ | $Y_i(0)$ | $Y_i(1)$ | $D_i$ | $Y_i$ |
|---|---|---|---|---|---|---|
| 1 | A | 0 | 0 | $\hat{\beta}_0 + \hat{\beta}_1$ | 0 | 0 |
| 2 | A | 0 | $\hat{\beta}_0$ | 1 | 1 | 1 |
| 3 | B | 0 | 1 | $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2$ | 0 | 1 |
| 4 | B | 0 | $\hat{\beta}_0 + \hat{\beta}_2$ | 1 | 1 | 1 |
| 5 | A | 1 | 0 | $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3$ | 0 | 0 |
| 6 | A | 1 | $\hat{\beta}_0 + \hat{\beta}_3$ | 1 | 1 | 1 |
| 7 | B | 1 | 1 | $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$ | 0 | 1 |
| 8 | B | 1 | $\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_3$ | 0 | 1 | 0 |
| 9 | A | 0 | 1 | $\hat{\beta}_0 + \hat{\beta}_1$ | 0 | 0 |
| 10 | B | 1 | $\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_3$ | 1 | 1 | 1 |

$$\mathrm{E}\left[Y_i \mid D_i, X_{[1]i}, X_{[2]i}\right] = \beta_0 + \beta_1 D_i + \beta_2 1\{X_{[1]i} = B\} + \beta_3 X_{[2]i}.$$

# Code: regression plug-in under ignorability

```
> df <- data.frame(
+   X_1 = c("A","A","B","B","A","A","B","B","A","B"),
+   X_2 = c(0,0,0,0,1,1,1,1,0,1),
+   D   = c(0,1,0,1,0,1,0,1,0,1),
+   Y   = c(0,1,1,1,0,1,1,0,0,1)
+ )
> df$X_1 <- factor(df$X_1)
> (fit_ign <- lm(Y ~ D + X_1 + X_2, data = df))

Call:
lm(formula = Y ~ D + X_1 + X_2, data = df)

Coefficients:
(Intercept)            D         X_1B          X_2
     0.3143       0.3571       0.3571      -0.1429
```

# Code: regression plug-in under ignorability

```
> df$yhat_1 <- df$Y
> df$yhat_0 <- df$Y
> df$yhat_1[which(df$D == 0)] <-
+   predict(fit_ign,
+           newdata = transform(df[which(df$D == 0), ], D = 1))
> df$yhat_0[which(df$D == 1)] <-
+   predict(fit_ign,
+           newdata = transform(df[which(df$D == 1), ], D = 0))
```

# Code: regression plug-in under ignorability

```
> round(df[, c("yhat_0", "yhat_1")], 3)

   yhat_0 yhat_1
1   0.000  0.671
2   0.314  1.000
3   1.000  1.029
4   0.671  1.000
5   0.000  0.529
6   0.171  1.000
7   1.000  0.886
8   0.529  0.000
9   0.000  0.671
10  0.529  1.000

> mean(df$yhat_1 - df$yhat_0)

[1] 0.3571429
```

# Directed Acyclic Graphs

- DAGs: what problem are we solving?
- We keep writing assumptions like:

$$Y_i(d) \perp\!\!\!\perp D_i \mid X_i \qquad \text{or} \qquad Y_i \perp\!\!\!\perp R_i \mid X_i .$$

- A DAG is a compact way to encode *which conditional independences are plausible* based on the data-generating process.
- Main use today: selecting (and *not* selecting) adjustment variables.

Greenland et al. (1999)

# Association vs causation: $\Pr[Y \mid D]$ vs $\Pr[Y \mid do(D)]$

- $\Pr[Y = y \mid D = d]$: observational association.

- $\Pr[Y = y \mid do(D = d)]$: distribution of $Y$ *under intervention* setting $D := d$.

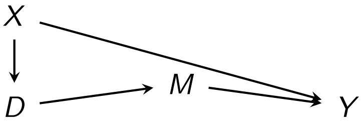
- **Graph surgery:** $do(D = d)$ deletes all arrows *into* $D$ (intervention breaks causes of $D$).

- Conditioning does *not* change the graph; it filters/stratifies the observed data.

Pearl et al. (2016)

# DAG primitives: nodes, arrows, paths, ancestors

- Node = variable.
- Arrow $A \rightarrow B$ means $A$ is a direct cause of $B$.
- A *path* is any sequence of adjacent arrows (ignore direction).
- A *directed path* is a causal pathway (all arrows forward).
- $A$ is an *ancestor* of $B$ if there is a directed path $A \rightarrow \cdots \rightarrow B$. (ancestor/descendant : parent/child)



$D \rightarrow M \rightarrow Y$ is a directed path;
$D \leftarrow X \rightarrow Y$ is a backdoor path.

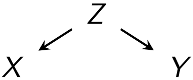
# Chain, fork, collider

**Chain**

$$X \longrightarrow Z \longrightarrow Y$$

Conditioning on $Z$ blocks the path.

**Fork (confounder)**



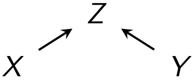
$Z$

$X$

$Y$

Conditioning on $Z$ blocks the backdoor.

**Collider**



$Z$

$X$

$Y$

Path blocked unless you condition on $Z$ (or a descendant).

# d-separation: the blocking definition

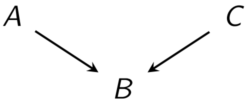
- “Blocking a path” means: given a conditioning set $Z$, that path cannot transmit statistical association between its endpoints under the graphical rules of d-separation.

- A path is **blocked** by a conditioning set $Z$ if *there exists at least one node* on the path such that:
  - the path contains a **chain** or **fork** $A \rightarrow B \rightarrow C$ or $A \leftarrow B \rightarrow C$ with the middle node $B \in Z$; or
  - the path contains a **collider** $A \rightarrow B \leftarrow C$ with $B \notin Z$ *and* no descendant of $B$ in $Z$.

- If *every* path between $X$ and $Y$ is blocked by $Z$, then $X$ and $Y$ are **d-separated** by $Z$.

- Pearl et al. convention: d-separated $\Rightarrow$ *guaranteed* conditional independence (given the model); d-connected $\Rightarrow$ dependence is *typical* (except cancellations).

# Collider intuition: “selection” creates dependence



$A \rightarrow B \leftarrow C$ (collider at $B$)

- Think of:
  - $A$ = ability, $C$ = effort, $B$ = admission to a selective program.

- In the full population, ability and effort can be independent:

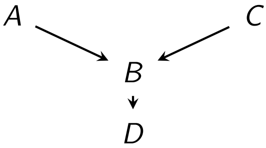
$$A \perp\!\!\!\perp C \quad \Longleftrightarrow \quad \Pr[C \mid A] = \Pr[C].$$

- The collider $B$ blocks the path between $A$ and $C$ *unless we condition on* $B$.

# Conditioning on admission opens the collider path

- Now restrict attention to admitted students: condition on $B = 1$.
- Intuition: admission depends on *either* high ability or high effort (or both).
  - If $A$ is low but $B = 1$, then $C$ must be high to "compensate."
  - If $A$ is high and $B = 1$, then $C$ can be lower and admission still occurs.
- Result: within the selected group $B = 1$, $A$ and $C$ become dependent:
$$\Pr[C \mid A, B = 1] \neq \Pr[C \mid B = 1].$$
- DAG language: conditioning on a collider **opens** the path $A \rightarrow B \leftarrow C$.

# Why descendants of a collider matter (conditioning "leaks" information)



*D* is a descendant of the collider *B*

- Let $D$ = program completion, where $B \to D$.
- Conditioning on $D = 1$ gives (partial) information about $B$.
- This can also induce dependence between $A$ and $C$:

$$\Pr[C \mid A, D = 1] \neq \Pr[C \mid D = 1].$$

- Rule: a collider blocks a path *only if* we do not condition on the collider *and* do not condition on any of its descendants.

# Adjustment for causal effects: backdoor criterion

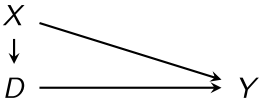
- A **backdoor path** from $D$ to $Y$ is any path that begins with an arrow *into* $D$.

- A set $Z$ satisfies the **backdoor criterion** for $(D, Y)$ if:
  - no element of $Z$ is a descendant of $D$, and
  - $Z$ blocks every backdoor path from $D$ to $Y$.

- If the backdoor criterion holds, then the causal effect is identified by:

$$\Pr[Y = y \mid do(D = d)] = \sum_z \Pr[Y = y \mid D = d, Z = z] \Pr[Z = z].$$

# Example: confounding and a valid adjustment set

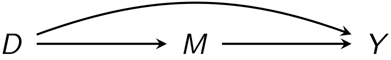


- Backdoor: $D \leftarrow X \rightarrow Y$.

- $Z = \{X\}$ blocks the backdoor and has no descendants of $D$.

- Graphically: $Y(d) \perp\!\!\!\perp D \mid X$ is plausible if this DAG is correct.

# Post-treatment adjustment changes the estimand (and can bias)



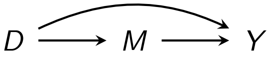
- $M$ is a mediator: conditioning on $M$ blocks part of the causal effect, and gives us a conditional association, not a causal effect.

- If you want the **total effect** ($\Pr[Y|do(D = d)]$), $M$ is not in a backdoor adjustment set.

- More generally: descendants of $D$ are excluded by the backdoor criterion.

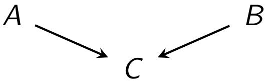
# Mediator vs collider: why the rule differs

**Mediator**



Conditioning on *M* changes the estimand (direct vs total effect).

**Collider**

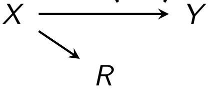


Conditioning on *C* creates bias by opening a path.

- Mediators: conditioning changes *which effect* you estimate.
- Colliders: conditioning breaks identification entirely.

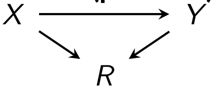
# Missing data as a DAG: MAR vs MNAR

**MAR (OK)**



$Y \perp\!\!\!\perp R \mid X.$

**MNAR (problem)**

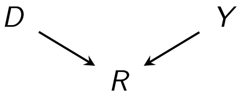


$Y \not\perp\!\!\!\perp R \mid X.$

- $R$ is the response/observation indicator (e.g., $Y$ observed if $R = 1$).

- Under MAR, conditioning on $X$ blocks all paths between $Y$ and $R$.

# Complete-case analysis is conditioning on $R = 1$ (selection)



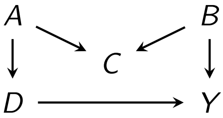
Conditioning on $R = 1$ opens the collider at $R$.

- Even if $D$ and $Y$ are independent marginally, restricting to the observed sample ($R = 1$) can induce $D$–$Y$ association.

- This is the collider lesson applied to missingness/selection.

- Under MAR, we try to block the $Y \to R$ channel *given covariates* (previous slide).

# Greenland et al. (1999): bad controls (conditioning can create confounding)



$C$ is a collider on $A \to C \leftarrow B$.

- Without conditioning on $C$, the collider blocks that path.

- Conditioning on $C$ opens the collider and induces association between $A$ and $B$.

- This opens a new backdoor from $D$ to $Y$ through $A$ and $B$.

- Lesson: “control for $C$” can *increase* bias.

# Why this mistake is tempting: total vs direct effects

- In the DAG, $C$ is *affected by* both $A$ and $B$.
- Researchers often reason:

  *"$C$ is related to both treatment and outcome, so we should control for it."*
- But this mixes up two different causal questions:
  - **Total effect of $D$ on $Y$:** effect through *all* causal paths.
  - **Direct effect of $D$ on $Y$:** effect *not operating through intermediates*.
- Conditioning on $C$ does *not* identify a direct effect here—it creates bias.

# What conditioning on $C$ actually does

- $C$ is a **collider** on the path $D \leftarrow A \rightarrow C \leftarrow B \rightarrow Y$.

- Conditioning on $C$:
    - induces dependence between $A$ and $B$;

    - opens a backdoor path from $D$ to $Y$;

    - violates ignorability for the total effect.

- Formally, after conditioning on $C$:

$$\Pr[Y(d) \mid D, C] \neq \Pr[Y(d) \mid C].$$

- So the resulting estimand is neither:
    - the total effect of $D$ on $Y$, nor
    - a well-defined direct effect.

# Practical workflow for choosing controls

- Step 1: draw a DAG that reflects substantive knowledge.
- Step 2: decide the estimand (total effect? direct effect?).
- Step 3: find a backdoor adjustment set (blocks all arrows-into-$D$ paths, avoids descendants of $D$).
- Step 4: check for **colliders** and **post-treatment variables** you might accidentally condition on.
- Step 5: translate to an estimator (regression / weighting / matching), then diagnose overlap and sensitivity later.

# Science table: strong ignorability and the propensity score

| $i$ | $X_{[1]i}$ | $X_{[2]i}$ | $Y_i(0)$ | $Y_i(1)$ | $D_i$ | $Y_i$ |
|---|---|---|---|---|---|---|
| 1 | A | 0 | 0 | ? | 0 | 0 |
| 2 | A | 0 | ? | 1 | 1 | 1 |
| 3 | B | 0 | 1 | ? | 0 | 1 |
| 4 | B | 0 | ? | 1 | 1 | 1 |
| 5 | A | 1 | 0 | ? | 0 | 0 |
| 6 | A | 1 | ? | 1 | 1 | 1 |
| 7 | B | 1 | 1 | ? | 0 | 1 |
| 8 | B | 1 | ? | 0 | 1 | 0 |
| 9 | A | 0 | 1 | ? | 0 | 0 |
| 10 | B | 1 | ? | 1 | 1 | 1 |

# Science table: add the (empirical) propensity score

| $i$ | $X_{[1]i}$ | $X_{[2]i}$ | $p_D(X_i)$ | $Y_i(0)$ | $Y_i(1)$ | $D_i$ | $Y_i$ |
|---|---|---|---|---|---|---|---|
| 1 | A | 0 | 0.33 | 0 | ? | 0 | 0 |
| 2 | A | 0 | 0.33 | ? | 1 | 1 | 1 |
| 3 | B | 0 | 0.50 | 1 | ? | 0 | 1 |
| 4 | B | 0 | 0.50 | ? | 1 | 1 | 1 |
| 5 | A | 1 | 0.50 | 0 | ? | 0 | 0 |
| 6 | A | 1 | 0.50 | ? | 1 | 1 | 1 |
| 7 | B | 1 | 0.67 | 1 | ? | 0 | 1 |
| 8 | B | 1 | 0.67 | ? | 0 | 1 | 0 |
| 9 | A | 0 | 0.33 | 1 | ? | 0 | 0 |
| 10 | B | 1 | 0.67 | ? | 1 | 1 | 1 |

| $X_{[1]}$ | $X_{[2]}$ | $n$ | $p_D(X)$ |
|---|---|---|---|
| A | 0 | 3 | 0.33 |
| A | 1 | 2 | 0.50 |
| B | 0 | 2 | 0.50 |
| B | 1 | 3 | 0.67 |

# Science table: hot deck imputation (propensity score)

| $i$ | $X_{[1]i}$ | $X_{[2]i}$ | $p_D(X_i)$ | $Y_i(0)$ | $Y_i(1)$ | $D_i$ | $Y_i$ |
|---|---|---|---|---|---|---|---|
| 1 | A | 0 | 0.33 | 0 | *donor* | 0 | 0 |
| 2 | A | 0 | 0.33 | *donor* | 1 | 1 | 1 |
| 3 | B | 0 | 0.50 | 1 | *donor* | 0 | 1 |
| 4 | B | 0 | 0.50 | *donor* | 1 | 1 | 1 |
| 5 | A | 1 | 0.50 | 0 | *donor* | 0 | 0 |
| 6 | A | 1 | 0.50 | *donor* | 1 | 1 | 1 |
| 7 | B | 1 | 0.67 | 1 | *donor* | 0 | 1 |
| 8 | B | 1 | 0.67 | *donor* | 0 | 1 | 0 |
| 9 | A | 0 | 0.33 | 1 | *donor* | 0 | 0 |
| 10 | B | 1 | 0.67 | *donor* | 1 | 1 | 1 |

- For each missing potential outcome, choose a nearest-neighbor donor in $p_D(X)$ from the opposite treatment arm.

- Impute $\widehat{Y}_i(0)$ or $\widehat{Y}_i(1)$ with that donor's observed outcome.

# References I

Greenland, S., Pearl, J., and Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, 10(1):37–48.

Pearl, J., Glymour, M., and Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.