

Selection on Observables I: The Assumption

Stat 256, Spring 2026

Chad Hazlett

UCLA

`chazlett@ucla.edu`

Where we are

Unit 1: DAGs, d-separation, backdoor criterion — a *graphical* language for identification.

Unit 2: Potential outcomes, experiments — ignorability from randomization gives

$$\tau_{\text{ATE}} = \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0].$$

Unit 3 (now): Observational studies. No randomization. We need an *assumption* to recover causal effects by conditioning on observed X .

Today: the assumption, its two languages ($\text{CI} \leftrightarrow \text{backdoor}$), and stratification as the first estimator. Wednesday: matching, weighting, regression.

Setting

- Units: $i = 1, \dots, N$
- Treatment: $D_i \in \{0, 1\}$, **not** randomly assigned
- Potential outcomes: Y_{0i}, Y_{1i} ; observed $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$
- Pre-treatment covariates: X_i (a vector)
- Quantity of interest: $\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}]$ (or τ_{ATT})

The Assumption: Conditional Ignorability

Identification Assumption (Conditional Ignorability / Selection on Observables)

$$\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp D_i \mid X_i$$

Many names for the same thing:

- Conditional ignorability (CI)
- Selection on observables (SOO)
- No unmeasured confounders
- Conditional exchangeability
- “Backdoor criterion satisfied by X ” (DAG language)

Read it as: *among units with the same X , treatment assignment is as good as random.*

CI and the backdoor criterion: two languages, one object

In Unit 1 we said: if X satisfies the backdoor criterion relative to (D, Y) in a DAG, the effect of D on Y is identified by adjusting for X :

$$P(Y \mid do(D=d)) = \sum_x P(Y \mid D=d, X=x) P(X=x).$$

In PO language we just said CI means $\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp D_i \mid X_i$.

Claim: these are the *same statement*, written in different alphabets. Let's see why.

Informal argument

Backdoor-satisfying X blocks all non-causal paths from D to Y :

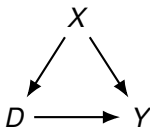
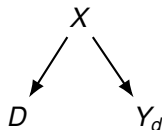
- Any association between D and Y that remains *within a stratum of X* can only come from the causal arrow $D \rightarrow Y$.
- Equivalently: within a stratum of X , treatment carries no information about the potential outcomes beyond its causal effect.
- That is conditional ignorability: $\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp D_i \mid X_i$.

So “block the backdoors” (DAG side) and “ D is as-if randomized within strata of X ” (PO side) say the same thing. To make this precise we need to be able to see potential outcomes on the graph.

View 1: the test graph $G_{\underline{D}}$

Take your DAG. Delete all arrows *out of* D . Call this $G_{\underline{D}}$.

Original DAG

 $G_{\underline{D}}$: delete $D \rightarrow Y$ 

In $G_{\underline{D}}$, the only way D and Y associate is through X . So: $D \perp\!\!\!\perp Y \mid X$ in $G_{\underline{D}}$.

$G_{\underline{D}}$ is a “null graph”: $Y = Y_1 = Y_0$. So we actually have:

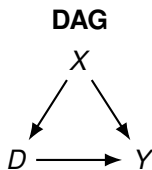
$$Y_d \perp\!\!\!\perp D \mid X.$$

Thus conditional ignorability is certified.

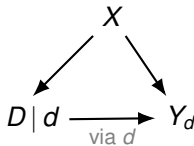
View 2: SWIGs (Richardson & Robins, 2013)

The core difficulty is that we need to reference both the realized D and the potential outcome Y_d even when $d \neq D$.

The SWIG (single world intervention graph) does just that:



SWIG: intervene $do(D = d)$



The “ $D | d$ ” node splits: the left side is the “natural” D you observe; the right side is the set value d that flows into Y_d .

You can see CI directly as d-separation of D from Y_d in the SWIG:

$$Y_d \perp\!\!\!\perp D \mid X$$

Formal statement, with a nod to the literature

Theorem (backdoor/adjustment criterion $\Rightarrow$ CI). If X satisfies the backdoor criterion relative to (D, Y) in a DAG G , then in any SCM consistent with G ,

$$Y_d \perp\!\!\!\perp D \mid X \quad \text{for each } d.$$

References for more:

- Pedagogical/informal but nice: *What If*, ch. 7.
- Direct, formal: Pearl, Glymour & Jewell (2016), *Primer*, Thm 4.3.1: backdoor-satisfying $Z \Rightarrow Y_x \perp\!\!\!\perp X \mid Z$.
- Also formal: Shpitser, VanderWeele & Robins (2010), Thm 4 states adjustment criterion and ties that to CI on their “counterfactual graph.”¹

¹These are like SWIGs but rather than splitting nodes, you show multiple copies of the graph, with and without intervention, with fused exogenous variables. Similar to “twin networks” we will mention later.

What would make CI true?

CI is a strong assumption. It says that all confounders are observed and included in X .

Credibility depends on the *story*: what determined treatment and outcome, and have you measured all of it?

What sorts of stories work?

- **Design:** a conditionally randomized experiment (e.g., block-randomized with different treatment probabilities per stratum). Then CI holds by construction.
- **Observational:** you must *argue* that X captures all common causes of D and Y . Substantive knowledge, not statistical testing.

CI is **not verifiable from the data**. No amount of checking observed covariates can rule out unobserved confounders.

Common Support / Positivity

CI alone is not enough. Another “identification-like” assumption:

Identification Assumption (Common Support / Overlap)

$$0 < \Pr(D_i = 1 \mid X_i = x) < 1 \quad \text{for all } x \text{ in the support of } X.$$

Weaker version for ATT: only need $\Pr(D = 1 \mid X) < 1$ (need controls for every treated stratum, but not vice versa).

This is somewhat ill-defined:

- What if X is continuous? Then we are assuming some unclear form of smoothness in $\Pr(D \mid X)$
- “Practical” positivity violations: some strata may lack observed treated or control units, even if they are not zero probability.
- An alternative: abandon common support and worry instead about reasoned uncertainty in each $\mathbb{E}[Y_d \mid X]$ as we move away from supporting data.

Identification of the ATE under CI

Quantity: $\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}]$

Step 1: within a stratum of X , CI gives an “experiment”:

$$\begin{aligned}\mathbb{E}[Y_{1i} - Y_{0i} \mid X=x] &= \mathbb{E}[Y_{1i} \mid X=x] - \mathbb{E}[Y_{0i} \mid X=x] \\ &\stackrel{\text{CI}}{=} \mathbb{E}[Y_{1i} \mid D=1, X=x] - \mathbb{E}[Y_{0i} \mid D=0, X=x] \\ &\stackrel{\text{consistency}}{=} \mathbb{E}[Y_i \mid D=1, X=x] - \mathbb{E}[Y_i \mid D=0, X=x].\end{aligned}$$

Step 2: put the strata back together using $\Pr(X)$:

$$\begin{aligned}\tau_{ATE} &= \mathbb{E}_X[\mathbb{E}[Y_{1i} - Y_{0i} \mid X]] \\ &= \sum_x (\mathbb{E}[Y_i \mid D=1, X=x] - \mathbb{E}[Y_i \mid D=0, X=x]) \Pr(X=x).\end{aligned}$$

The RHS is fully observable (given positivity), so the ATE is **identified**.

Exactly what you did with the adjustment formula.²

²Here we are asking for $P(Y \mid do(D=1)) - P(Y \mid do(D=0))$ at once, instead of each $P(Y \mid do(D))$.

Example: Smoking and Mortality (Cochran 1968)

Unadjusted death rates per 1,000 person-years:

	Canada	U.K.	U.S.
Non-smokers	20.2	11.3	13.5
Cigarette only	20.5	14.1	13.5
Cigars/pipes	35.5	20.7	17.4

Pipe/cigar smokers die at much higher rates. Causal effect of smoking?

Or could something else explain this?

The Confounder: Age

Mean ages by smoking group:

	Canada	U.K.	U.S.
Non-smokers	54.9	49.1	57.0
Cigarette only	50.5	49.8	53.2
Cigars/pipes	65.9	55.7	59.7

Pipe/cigar smokers are *much older* on average.

Age is associated with both smoking type and death rate — it is a confounder.
Can we adjust for it?

Subclassification Reveals the Effect

Fabricated data inspired by the U.S. figures (death rates per 1,000 person-years):

Age group	Rate (Cig)	Rate (Non)	Effect	N (Cig)	N (Non)
< 45	5	3	+2	40	10
45–54	10	5	+5	30	20
55–64	20	10	+10	20	30
65+	45	23	+22	10	40
Pooled	13.5	13.5	0	100	100

Within every age group, smokers die at higher rates. But the pooled difference is **zero** — because smokers are younger.

Subclassification (each age stratum has $N_j = 50$):

$$\hat{\tau}_{ATE} = \frac{1}{4}(2 + 5 + 10 + 22) = 9.75$$

The confounding hid an effect of nearly 10 deaths per 1,000 person-years.

The Most Important Slide

Pause and ask:

- What **assumption** did we need to interpret this causally?
- State it as CI: $\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp D_i \mid X_i$. What is X here?
- State it in “unobserved confounders” language.
- Do you believe it? What unmeasured U could violate CI?
- Can you *test* CI with data? What can you test?

This is the hard part — not the estimator.

Don't Condition on Post-treatment Variables

You already know why:

- **Selection bias**: if you restrict to a subset of the data defined by a post-treatment variable, you may be selecting on a collider, inducing an association between D and the confounder that wasn't there before, breaking the randomization.
- **Mediation “bias”** (or changing the estimand): if you control for a variable that is on the causal pathway from D to Y , you are blocking part of the causal effect you are trying to estimate, which can lead to an underestimate of the total effect.

What we understood simply as collider bias before should be visible to you as a violation of CI. Conditioning on some Z with $D \rightarrow Z \leftarrow U \rightarrow Y$

- changes the distribution of Y_d (because of the U),
- and does so differently for the treated vs. control (because of the D), so you lose the $Y_d \perp\!\!\!\perp D$ you thought you had.

Estimation under CI: the idea

The ID result says: compute treatment effects *within strata of X* , then average over the population distribution of X .

Subclassification (stratification) does this literally when X is discrete:

- 1 Divide units into J strata defined by X .
- 2 Within-stratum DIM: $\hat{\tau}_j = \bar{Y}_{1j} - \bar{Y}_{0j}$.
- 3 Average: $\hat{\tau}_{ATE} = \sum_{j=1}^J \frac{N_j}{N} \hat{\tau}_j$.

No modeling. Simply the adjustment formula from the DAG (twice).

But: breaks down when X has many components or continuous entries — most strata have 0 treated or 0 controls (curse of dimensionality).

Next time: matching, propensity scores, weighting, regression — all trying to do the same thing when literal stratification fails.

Takeaways

- **CI/SOO is an assumption, not a procedure.** “No unmeasured confounders” is a strong, unverifiable claim about the world.
- **CI and the backdoor criterion are the same statement** in different languages. Adjustment formula = stratification estimator.
- **Credibility comes from the story**, not the data.
- **Positivity** is a separate, ID-like assumption, but can be murky.
- **Don't condition on post-treatment variables.**
- **Once you believe CI**, the estimation question (matching, weighting, regression) is secondary. We turn to that next.

Reminder: Advanced Topic Presentations

Due today: submit *group* (2–3 students) and *top 3 topics* (ranked) using the google form sent via BruinLearn.

- See `PresentationAssignment.pdf` on the course site for the topic list.
- Stick around or email me if you want to talk about this or help find group members.