16. Treatments in many time periods. What to do.

Ian Lundberg
Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

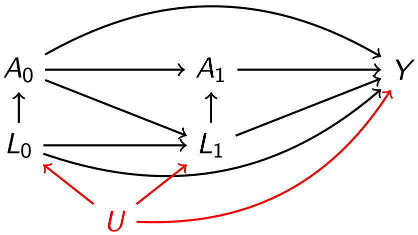
18 Oct 2022

Learning goals for today

At the end of class, you will be able to:

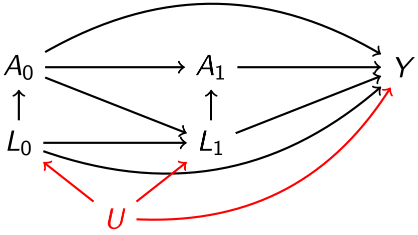
1. Reason about the sequential ignorability assumption
2. Apply inverse probability weighting to treatments over time

Identification: The adjustment set



A joint adjustment set for $\bar{A}$ is doomed

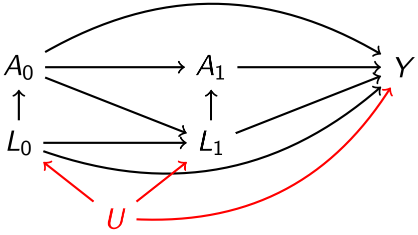
Identification: The adjustment set



A joint adjustment set for $\bar{A}$ is doomed

- What happens if you adjust for L_1?

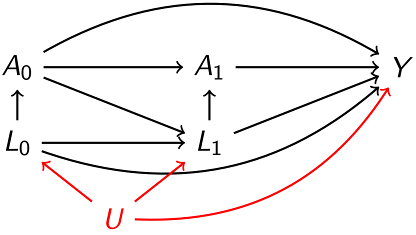
Identification: The adjustment set



A joint adjustment set for $\bar{A}$ is doomed

- What happens if you adjust for L_1?
  - You block a causal path: $A_0 \rightarrow \boxed{L_1} \rightarrow Y$
  - You open a backdoor path: $A_0 \rightarrow \boxed{L_1} \leftarrow U \rightarrow Y$

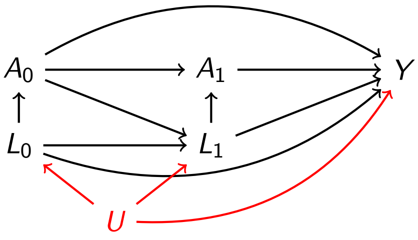
Identification: The adjustment set



A joint adjustment set for $\bar{A}$ is doomed

- What happens if you adjust for L_1?
  - You block a causal path: $A_0 \rightarrow \boxed{L_1} \rightarrow Y$
  - You open a backdoor path: $A_0 \rightarrow \boxed{L_1} \leftarrow U \rightarrow Y$
- What happens if you don't adjust for L_1?

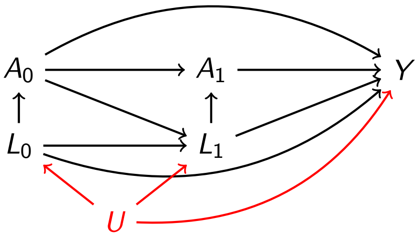
Identification: The adjustment set



A joint adjustment set for $\bar{A}$ is doomed

- What happens if you adjust for L_1?
  - You block a causal path: $A_0 \to \boxed{L_1} \to Y$
  - You open a backdoor path: $A_0 \to \boxed{L_1} \leftarrow U \to Y$
- What happens if you don't adjust for L_1?
  - A backdoor path remains: $A_1 \leftarrow L_1 \to Y$

Identification: The adjustment set



A joint adjustment set for $\bar{A}$ is doomed

- What happens if you adjust for L_1?
  - You block a causal path: $A_0 \rightarrow \boxed{L_1} \rightarrow Y$
  - You open a backdoor path: $A_0 \rightarrow \boxed{L_1} \leftarrow U \rightarrow Y$
- What happens if you don't adjust for L_1?
  - A backdoor path remains: $A_1 \leftarrow L_1 \rightarrow Y$

To proceed, we need a different adjustment set in each time period

Notation

- $\bar{A}_k = (A_0, A_1, \ldots, A_k)$ treatments up to time k
- $\bar{L}_k = (L_0, L_1, \ldots, L_k)$ confounders up to time k
- $g()$ treatment strategy
- Y^g potential outcome under that strategy

Identification: Sequential ignorability

Identification: Sequential ignorability

Potential outcome under assignment rule g	is independent of	treatment at time k	given	treatments up to $k-1$ followed rule g	and confounders up to time k
Y^g	$\perp\!\!\!\perp$	A_k	$\mid$	$\bar{A}_{k-1} = g(\bar{A}_{k-2}, \bar{L}_{k-1}),$	$\bar{L}_k$

for all assignment rules g and time periods $k = 1, \ldots, K$

Identification: Sequential ignorability

Potential outcome under assignment rule g	is independent of	treatment at time k	given	treatments up to $k-1$ followed rule g	and confounders up to time k
Y^g	$\perp\!\!\!\perp$	A_k	$\mid$	$\bar{A}_{k-1} = g(\bar{A}_{k-2}, \bar{L}_{k-1}),$	$\bar{L}_k$

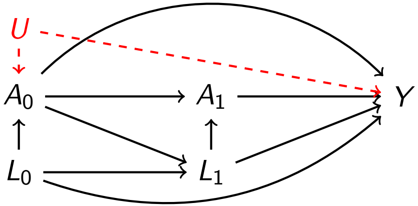
for all assignment rules g and time periods $k = 1, \ldots, K$

$$
\begin{tikzcd}
A_0 \arrow[r] \arrow[rr, bend left] \arrow[dr] & A_1 \arrow[r] & Y \\
L_0 \arrow[u] \arrow[r] \arrow[urr, bend right] & L_1 \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

Identification: Sequential ignorability

Potential outcome under assignment rule g	is independent of	treatment at time k	given	treatments up to $k-1$ followed rule g	and confounders up to time k
Y^g	$\perp\!\!\!\perp$	A_k	$\mid$	$\bar{A}_{k-1} = g(\bar{A}_{k-2}, \bar{L}_{k-1}),$	$\bar{L}_k$

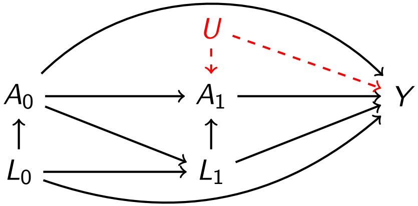
for all assignment rules g and time periods $k = 1, \ldots, K$



Identification: Sequential ignorability

$$\overset{\text{Potential outcome under assignment rule } g}{Y^g} \quad \overset{\text{is independent of}}{\perp\!\!\!\perp} \quad \overset{\text{treatment at time } k}{A_k} \quad \overset{\text{given}}{\Big|} \quad \overset{\text{treatments up to } k-1 \text{ followed rule } g}{\bar{A}_{k-1} = g(\bar{A}_{k-2}, \bar{L}_{k-1}),} \quad \overset{\text{and confounders up to time } k}{\bar{L}_k}$$

for all assignment rules g and time periods $k = 1, \ldots, K$



Identification: Sequential ignorability

Potential outcome under assignment rule g	is independent of	treatment at time k	given	treatments up to $k-1$ followed rule g	and confounders up to time k
Y^g	$\perp\!\!\!\perp$	A_k	$\mid$	$\bar{A}_{k-1} = g(\bar{A}_{k-2}, \bar{L}_{k-1}),$	$\bar{L}_k$

for all assignment rules g and time periods $k = 1, \ldots, K$

$$
\begin{tikzcd}
A_0 \arrow[r] \arrow[dr] \arrow[rr, bend left] & A_1 \arrow[r] & Y \\
L_0 \arrow[u] \arrow[r] \arrow[urr, bend right] & L_1 \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

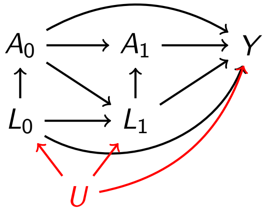
Estimation: Two strategies

1. Inverse probability weighting ($+$ marginal structural models)
2. Structural nested mean models (coming next class)

Inverse probability weighting: DAG motivation

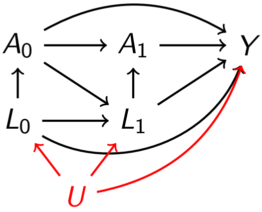
Inverse probability weighting: DAG motivation

We observe data from this model

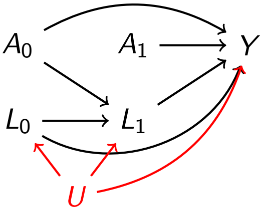


Inverse probability weighting: DAG motivation

We observe data from this model

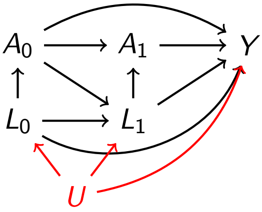


We want this

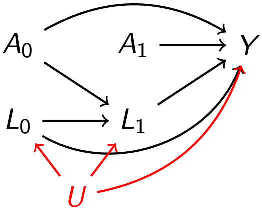


Inverse probability weighting: DAG motivation

We observe data from this model



We want this



1. How would you weight to estimate the effect of A_0?
2. How would you weight to estimate the effect of A_1?

We will combine these

Inverse probability weighting

In time 0, define an inverse probability of treatment weight

$$W^{A_0} = \frac{1}{\mathrm{P}(A_0 \mid L_0)}$$

such that A_0 does not depend on L_0 after weighting

Inverse probability weighting

In time 0, define an inverse probability of treatment weight

$$W^{A_0} = \frac{1}{\mathrm{P}(A_0 \mid L_0)}$$

such that A_0 does not depend on L_0 after weighting

In time 1, do it again

$$W^{A_1} = \frac{1}{\mathrm{P}(A_0 \mid \bar{A}_{k-1}, \bar{L}_1)}$$

Inverse probability weighting

In time 0, define an inverse probability of treatment weight

$$W^{A_0} = \frac{1}{\mathrm{P}(A_0 \mid L_0)}$$

such that A_0 does not depend on L_0 after weighting

In time 1, do it again

$$W^{A_1} = \frac{1}{\mathrm{P}(A_0 \mid \bar{A}_{k-1}, \bar{L}_1)}$$

Continue through all time periods.

Inverse probability weighting

In time 0, define an inverse probability of treatment weight

$$W^{A_0} = \frac{1}{\mathrm{P}(A_0 \mid L_0)}$$

such that A_0 does not depend on L_0 after weighting

In time 1, do it again

$$W^{A_1} = \frac{1}{\mathrm{P}(A_0 \mid \bar{A}_{k-1}, \bar{L}_1)}$$

Continue through all time periods.

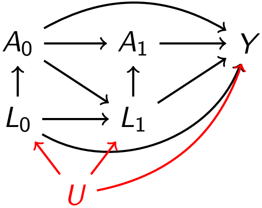
Define the overall weight as the product

$$W^{\bar{A}} = \prod_{k=0}^{K} \frac{1}{\mathrm{P}(A_k \mid \bar{A}_{k-1}, \bar{L}_k)}$$

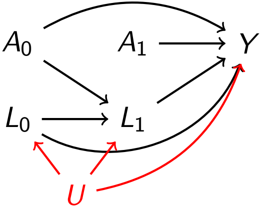
Inverse probability weighting

$$W^{\bar{A}} = \prod_{k=0}^{K} \frac{1}{\mathrm{P}(A_k \mid \bar{A}_{k-1}, \bar{L}_k)}$$

Takes us from this



to this pseudo-population



Inverse probability weighting with marginal structural models

Finally, we can put a model on top of the weighting.

$$\mathrm{E}(Y^{\bar{a}}) = \mathrm{E}(Y \mid \bar{A} = \bar{a}) = h(\bar{a})$$

for some function $h()$ that pools information.

Example: Outcomes depend on the proportion of periods treated

$$h(\bar{a}) = \frac{1}{K+1}\sum_{k=0}^{K} a_k$$

Learning goals for today

At the end of class, you will be able to:

1. Reason about the sequential ignorability assumption
2. Apply inverse probability weighting to treatments over time

Real example: Neighborhood disadvantage

Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. American Sociological Review, 76(5), 713-736.

Real example: Neighborhood disadvantage

Wodtke et al. 2011

How does the neighborhood in which a child lives affect that child's probability of high school completion?

Real example: Neighborhood disadvantage

Wodtke et al. 2011

How does the neighborhood in which a child lives affect that child's probability of high school completion?

- Define a neighborhood as a Census tract

Real example: Neighborhood disadvantage

Wodtke et al. 2011

How does the neighborhood in which a child lives affect that child's probability of high school completion?

- Define a neighborhood as a Census tract
- Score that neighborhood along several dimensions
  - poverty
  - unemployment
  - welfare receipt
  - female-headed households
  - education
  - occupational structure

Real example: Neighborhood disadvantage

Wodtke et al. 2011

How does the neighborhood in which a child lives affect that child's probability of high school completion?

- Define a neighborhood as a Census tract
- Score that neighborhood along several dimensions
  - poverty
  - unemployment
  - welfare receipt
  - female-headed households
  - education
  - occupational structure
- Scale by the first principle component

Real example: Neighborhood disadvantage

Wodtke et al. 2011

How does the neighborhood in which a child lives affect that child's probability of high school completion?

- Define a neighborhood as a Census tract
- Score that neighborhood along several dimensions
  - poverty
  - unemployment
  - welfare receipt
  - female-headed households
  - education
  - occupational structure
- Scale by the first principle component
- Categorize in 5 quintiles

Real example: Neighborhood disadvantage

Wodtke et al. 2011

How does the neighborhood in which a child lives affect that child's probability of high school completion?

- Define a neighborhood as a Census tract
- Score that neighborhood along several dimensions
  - poverty
  - unemployment
  - welfare receipt
  - female-headed households
  - education
  - occupational structure
- Scale by the first principle component
- Categorize in 5 quintiles

This 5-value treatment is "neighborhood disadvantage"

Real example: Neighborhood disadvantage

Wodtke et al. 2011

Neighborhoods are experienced over time:

$$\bar{a}$$

is a trajectory of neighborhood disadvantage over ages $2, 3, \ldots, 17$

The authors study the effect of neighborhood disadvantage,

$$\begin{aligned} E(Y_{\bar{a}} - Y_{\bar{a}'}) &= E(Y_{\bar{a}}) - E(Y_{\bar{a}'}) \\ &= P(Y_{\bar{a}} = 1) - P(Y_{\bar{a}'} = 1), \end{aligned} \qquad (1)$$

Example:
$\bar{a}$ is residence in the most advantaged neighborhood each year
and
$\bar{a}'$ is residence in the most disadvantaged neighborhood each year

Real example: Neighborhood disadvantage

Wodtke et al. 2011

Problem: Neighborhoods A_1 shape family characteristics L_2, which confound where people live in the future A_2

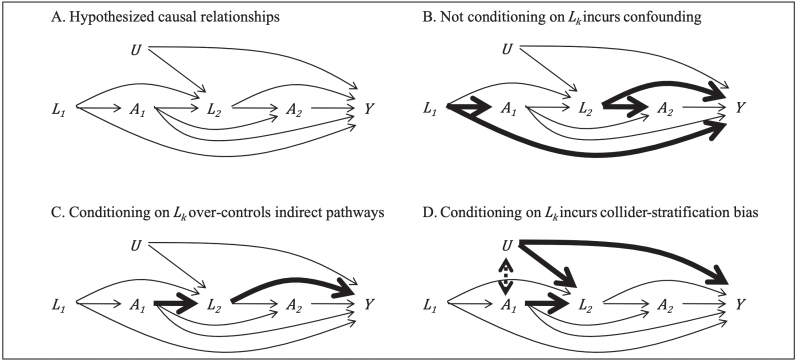


Figure 1. Causal Graphs for Exposure to Disadvantaged Neighborhoods with Two Waves of Follow-up

Note: A_k = neighborhood context, L_k = observed time-varying confounders, U = unobserved factors, Y = outcome.

Table 2. Time-Dependent Sample Characteristics

Variable	Blacks (n = 834)			Nonblacks (n = 1,259)		
	Age 1	Age 10	Age 17	Age 1	Age 10	Age 17
NH disadvantage index, percent						
1st quintile	3.48	3.60	3.48	13.34	19.14	20.65
2nd quintile	3.24	3.72	6.00	19.46	18.67	21.84
3rd quintile	5.28	5.88	7.79	26.13	23.27	22.48
4th quintile	14.87	18.11	18.47	26.13	23.99	21.13
5th quintile	73.14	68.71	64.27	14.93	14.93	13.90
FU head's marital status, percent						
Unmarried	33.93	44.84	52.04	5.88	11.36	15.09
Married	66.07	55.16	47.96	94.12	88.64	84.91
FU head's employment status, percent						
Unemployed	27.22	32.61	33.09	8.10	8.02	9.69
Employed	72.78	67.39	66.91	91.90	91.98	90.31
Public assistance receipt, percent						
Did not receive AFDC	81.06	75.66	82.37	96.27	96.19	97.93
Received AFDC	18.94	24.34	17.63	3.73	3.81	2.07
Homeownership, percent						
Do not own home	69.66	53.48	50.12	40.19	22.32	20.73
Own home	30.34	46.52	49.88	59.81	77.68	79.27
FU income in $1,000s, mean	19.68	25.04	27.45	32.59	46.65	57.50
FU head's work hours, mean	30.08	26.82	27.51	42.65	40.84	40.68
FU size, mean	5.75	5.32	4.81	4.22	4.69	4.33
Cum. residential moves, mean	.32	2.48	3.64	.32	2.16	3.02

Note: NH = neighborhood; FU = family unit. Statistics reported for children not lost to follow-up before age 20 (first imputation dataset).

Real example: Neighborhood disadvantage

Wodtke et al. 2011

Problem: Neighborhoods A_1 shape family characteristics L_2, which confound where people live in the future A_2

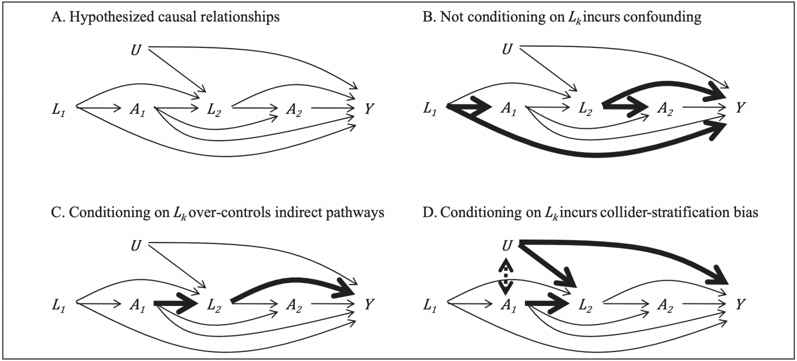


Figure 1. Causal Graphs for Exposure to Disadvantaged Neighborhoods with Two Waves of Follow-up

Note: A_k = neighborhood context, L_k = observed time-varying confounders, U = unobserved factors, Y = outcome.

Real example: Neighborhood disadvantage

Wodtke et al. 2011

Solution: MSM-IPW

$$w_i = \prod_{k=1}^{K} \frac{1}{P(A_k = a_{ki} \mid \bar{A}_{k-1} = \bar{a}_{(k-1)i}, \bar{L}_k = \bar{l}_{ki})}. \quad (4)$$

Also with stabilized weights

$$sw_i = \prod_{k=1}^{K} \frac{P(A_k = a_{ki} \mid \bar{A}_{k-1} = \bar{a}_{(k-1)i}, L_0 = l_0)}{P(A_k = a_{ki} \mid \bar{A}_{k-1} = \bar{a}_{(k-1)i}, \bar{L}_k = \bar{l}_{ki})}, \quad (5)$$

Real example: Neighborhood disadvantage

Wodtke et al. 2011

Marginal structural model: Logit

- 5-category treatment entered numerically
- Baseline covariates included due to stabilized weights
- Weights adjust for time-varying confounding

Real example: Neighborhood disadvantage

Wodtke et al. 2011

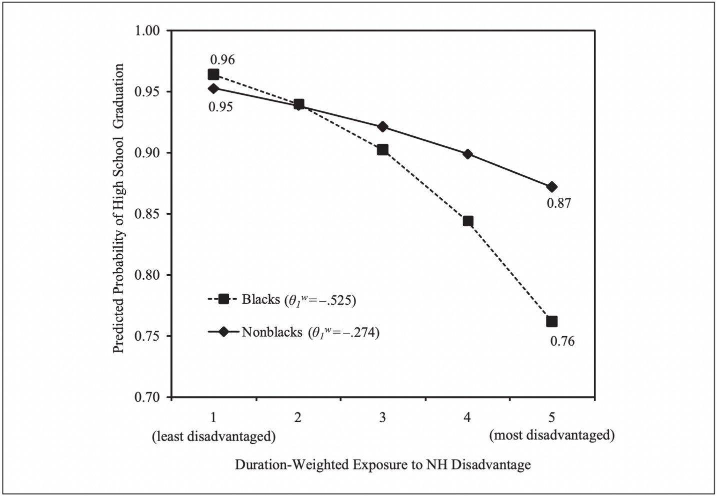


Figure 3. Predicted Probability of High School Graduation by Neighborhood Exposure History

Note: NH = Neighborhood

Learning goals for today

At the end of class, you will be able to:

1. Reason about the sequential ignorability assumption
2. Apply inverse probability weighting to treatments over time

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!