Covariate Adjustment

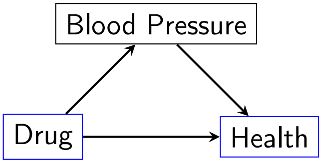
N. Koch, M. Vázquez, A. Nava, F. Otto

ETH Zürich

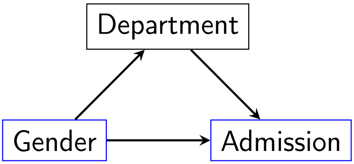
March 13, 2023

Causal Questions

(1) Drug Treatment



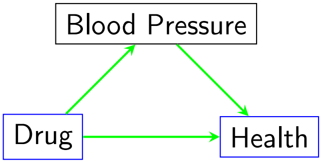
(2) College Admissions



Causal Questions

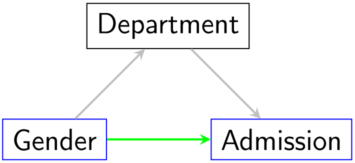
(1) Drug Treatment

- Does the drug improve health overall?
- Effect on mediator is relevant
- ▷ *Total causal effect (TCE)*



(2) College Admissions

- Are females being discriminated against?
- Effect on mediator is *not* relevant
- ▷ *Direct causal effect (DCE)*



Total vs Direct Causal Effect

Definition (Total causal effect, TCE)

Given a causal model over $(X, Y, \mathbf{W})$, there is a total causal effect of X on Y if any of the following equivalent conditions hold:

1. $\exists x' : \; p(y \mid do(X := x')) \neq p(y)$
2. $\exists x', \tilde{x} : \; p(y \mid do(X := x')) \neq p(y \mid do(X := \tilde{x}))$
3. *(others...)*

Definition (Direct causal effect, DCE)

(later...)

Note: which type of causal effect we are interested in is context-dependent

Example: Kidney Stone Treatment

- $S \in \{\text{small}, \text{large}\}$; $T \in \{A, B\}$; $R \in \{0, 1\}$
  - Treatment A: open surgery
  - Treatment B: non-invasive treatment

	Treatment A	Treatment B
Small	**93%** (81/87)	87% (234/270)
Large	**73%** (192/263)	69% (55/80)
Total	78% (273/350)	**83%** (289/350)
	80% (562/700)	

⇝ **Q:** What is the chance of recovery for a *given* patient if we assign them to treatment A *as opposed to* treatment B?

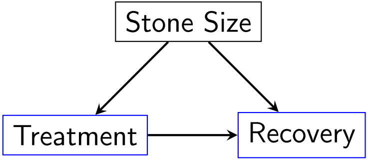
$$\mathbb{P}(R = 1 \mid \mathrm{do}(T := A)) \overset{?}{\lessgtr} \mathbb{P}(R = 1 \mid \mathrm{do}(T := B))$$

Example: Kidney Stone Treatment (cont.)

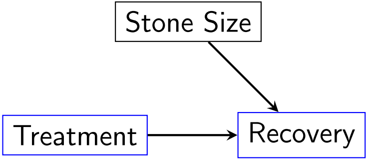
Assumption (Autonomy)

Intervening (only) on x_i *leaves* $p\big(x_j \mid pa(x_j)\big)$, $i \neq j$, *unchanged.*

observational distribution



interventional distribution

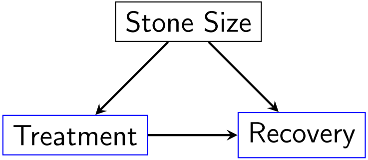


Example: Kidney Stone Treatment (cont.)

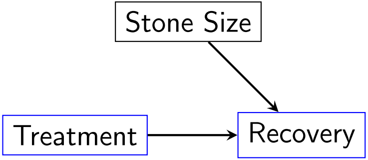
Assumption (Autonomy)

Intervening (only) on x_i leaves $p\big(x_j \mid pa(x_j)\big)$, $i \neq j$, unchanged.

observational distribution



interventional distribution



$$
\begin{aligned}
\mathbb{P}(S) &= \mathbb{P}\big(S \mid \mathrm{do}(T := \cdot)\big) \\
\mathbb{P}(R \mid S, T = \cdot) &= \mathbb{P}\big(R \mid S, \mathrm{do}(T := \cdot)\big) \\
\mathbb{P}(R \mid T = \cdot) &\neq \mathbb{P}\big(R \mid \mathrm{do}(T := \cdot)\big)
\end{aligned}
$$

Example: Kidney Stone Treatment (cont.)

Claim: can compute total causal effect of treatments using observational distributions only, namely

$$(*) \qquad \mathbb{P}\big(R = 1 \mid \mathrm{do}(T := \cdot)\big) \; = \; \sum_s \mathbb{P}(R = 1 \mid S = s, T = \cdot)\,\mathbb{P}(S = s)$$

Example: Kidney Stone Treatment (cont.)

Claim: can compute total causal effect of treatments using observational distributions only, namely

$$(*) \quad \mathbb{P}\big(R = 1 \mid \mathrm{do}(T := \cdot)\big) = \sum_{s} \mathbb{P}(R = 1 \mid S = s, T = \cdot)\, \mathbb{P}(S = s)$$

Proof:

$$\begin{aligned}
(*) &= \sum_{s\in\{0,1\}} \mathbb{P}\big(R = 1, S = s \mid \mathrm{do}(T := A)\big) \\
&= \sum_{s\in\{0,1\}} \mathbb{P}\big(R = 1 \mid S = s, \mathrm{do}(T := A)\big)\, \mathbb{P}\big(S = s \mid \mathrm{do}(T := A)\big) \\
&= \sum_{s\in\{0,1\}} \mathbb{P}(R = 1 \mid S = s, T = A)\, \mathbb{P}(S = s)
\end{aligned}$$

Example: Kidney Stone Treament (cont.)

Claim: can compute total causal effect of treatments using observational distributions only, namely

$$(*) \qquad \mathbb{P}\big(R = 1 \mid \mathrm{do}(T := \cdot)\big) \;=\; \sum_{s} \mathbb{P}(R = 1 \mid S = s, T = \cdot)\, \mathbb{P}(S = s)$$

Example: Kidney Stone Treament (cont.)

Claim: can compute total causal effect of treatments using observational distributions only, namely

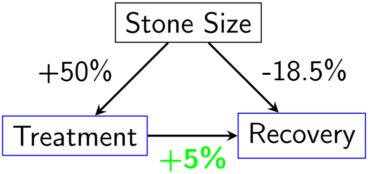
$$(*) \quad \mathbb{P}\big(R = 1 \mid \mathrm{do}(T := \cdot)\big) = \sum_s \mathbb{P}(R = 1 \mid S = s, T = \cdot)\, \mathbb{P}(S = s)$$

Note:

$$\begin{aligned} (*) &= \sum_{s\in\{0,1\}} \mathbb{P}(R = 1 \mid S = s, T = \cdot)\, \mathbb{P}(S = s) \\ &\neq \sum_{s\in\{0,1\}} \mathbb{P}(R = 1 \mid S = s, T = \cdot)\, \mathbb{P}(S = s \mid T = \cdot) \\ &= \mathbb{P}(R = 1 \mid T = \cdot) \end{aligned}$$

$\rightarrow$ we have to *adjust* treatment effects for stone size

Example: Kidney Stone Treatment (cont.)



- $\mathbb{P}\big(R = 1 \mid \mathrm{do}(T := A)\big) \;\approx\; 83\% \;\neq\; 78\% \;\approx\; \mathbb{P}(R = 1 \mid T = A)$
- $\mathbb{P}\big(R = 1 \mid \mathrm{do}(T := B)\big) \;\approx\; 78\% \;\neq\; 83\% \;\approx\; \mathbb{P}(R = 1 \mid T = B)$

Denote by $\tau_{T,R}$ the total causal effect of T on R. Then,

$\rightsquigarrow$ $\tau_{T,R} \;\approx\; 83\% - 78\% \;=\; \mathbf{5\%}$

i.e., treatment A works better than treatment B

Total Causal Effect

Definition (Total causal effect, TCE)

Given a causal model over $(X, Y, \mathbf{W})$, there is a total causal effect of X on Y if any of the following equivalent conditions hold:

1. $\exists x' : \; p\big(y \mid do(X := x')\big) \; \neq \; p(y)$
2. $\exists x', \tilde{x} : \; p\big(y \mid do(X := x')\big) \; \neq \; p\big(y \mid do(X := \tilde{x})\big)$

How should we *quantify* a potential TCE?

1. $\rightsquigarrow \; p\big(y \mid \mathrm{do}(X := x')\big) \; \leftrightarrow \; p(y)$
2. $\rightsquigarrow \; p\big(y \mid \mathrm{do}(X := x')\big) \; \leftrightarrow \; p\big(y \mid \mathrm{do}(X := \tilde{x})\big)$

Total Causal Effect

$\tau_{X,Y}$: TCE of X on Y

1. from cond. (1):

$$\rightsquigarrow \tau_{X,Y} := \mathbb{E}\big[Y \mid \mathrm{do}(X := x')\big] - \mathbb{E}[Y]$$

2. from cond. (2):
   - for discrete variables:

$$\rightsquigarrow \tau_{X,Y} := \mathbb{E}\big[Y \mid \mathrm{do}(X := x')\big] - \mathbb{E}\big[Y \mid \mathrm{do}(X := \tilde{x})\big]$$

   - for continuous variables:

$$\rightsquigarrow \tau_{X,Y} := \frac{\partial}{\partial x'}\mathbb{E}\big[Y \mid \mathrm{do}(X := x')\big]$$

Problem: definitions generally depend on $x', \tilde{x}$

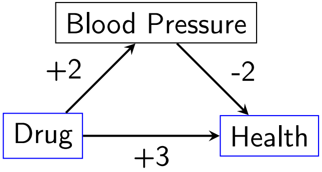
Linear SEMs: Path Method

Proposition (Path Method)

In a linear SEM, the total causal effect of X on Y is the sum of the products of the path coefficients over all directed paths from X to Y.

Consider the SEM over $D, B, H \in \mathbb{R}$

- $D \leftarrow \varepsilon_D$
- $B \leftarrow 2D + \varepsilon_B$
- $H \leftarrow 3D - 2B + \varepsilon_H$



$$\rightsquigarrow \ \tau_{D,H} = 3 + 2 \cdot (-2) = \textbf{(-1)}$$

$\rightarrow$ in linear SEMs, definitions of $\tau_{X,Y}$ coincide and are always constant

Linear SEMs: Regression

Proposition (TCE from regression)

In a linear SEM over $(X, Y, \boldsymbol{W})$, it holds that $\tau_{X,Y} = \gamma$, where γ is the coefficient of X in the linear regression

$$Y = \gamma X + \boldsymbol{\beta}^\top \boldsymbol{Z} + \varepsilon$$

where $\boldsymbol{Z} \subseteq \boldsymbol{W}$ is a valid adjustment set for (X, Y).

$\rightsquigarrow$ if we have $\boldsymbol{Z}$ and a linear SEM, we can

- obtain *unbiased estimates* for $\tau_{X,Y}$
- estimate its *statistical significance*

Valid Adjustment Sets

Definition (Valid adjustment set)

Given a structural causal model over $(X, Y, \boldsymbol{W})$*, a set* $\boldsymbol{Z} \subseteq \boldsymbol{W}$ *is a valid adjustment set for the ordered pair* (X, Y) *if*

$$p(y \mid do(X := x)) = \int_{\boldsymbol{z}} p(y \mid x, \boldsymbol{z})\, p(\boldsymbol{z})\, d\boldsymbol{z}$$

- Valid adjustment sets allow us to estimate total causal effects purely from observational data

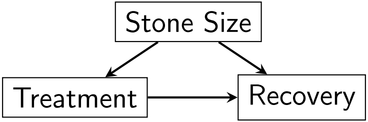
$\rightsquigarrow$ how to identify $\boldsymbol{Z}$ in general?

Parent Adjustment

Proposition (Parent adjustment)

Assume $k \notin PA(i)$. Then:

$$PA(i) \textit{ is a valid adjustment set for } (i, k).$$

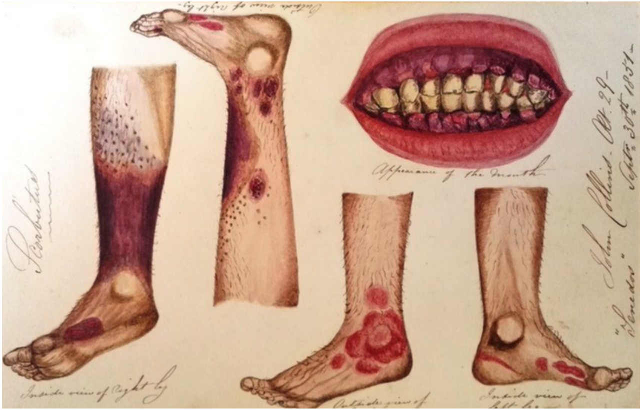


$$p(S, R \mid \mathrm{do}(T)) = \frac{p(S, R, T)}{p(T \mid S)} = p(R \mid T, S)p(S)$$

$$p(R \mid \mathrm{do}(T)) = \sum_s p(R \mid T, S = s)p(S = s)$$

Example: Scurvy

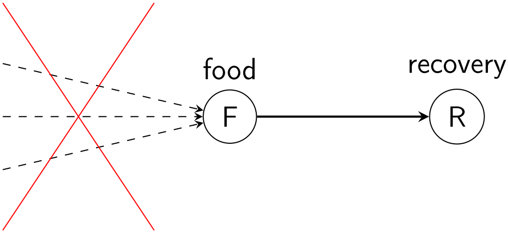
Scurvy disease → James Lind (1747)



Example: Scurvy (cont.)

- 12 men suffering from similar symptoms, divided in pairs and treated with:
  - ⇝ Cider
  - ⇝ Elixir of vitriol
  - ⇝ Sea water
  - ⇝ Vinegar
  - ⇝ Paste of garlic, horse-radish, mustard seed...
  - ⇝ Oranges and lemons

Example: Scurvy (cont.)



Then:

$$P(R \mid \mathrm{do}(F := f)) = P(R \mid F = f)$$

Backdoor Criterion

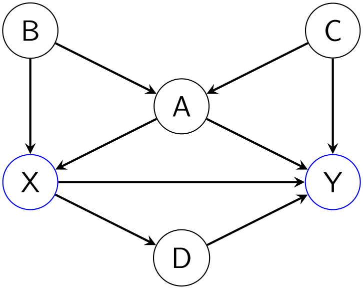
- The backdoor criterion is **sufficient** for adjustment.

Backdoor Criterion (Pearl)

Let $G = (\boldsymbol{V}, \boldsymbol{E})$ be a DAG and $i, k \in \boldsymbol{V}, i \neq k$. A set $\boldsymbol{Z} \subset \boldsymbol{V}$ (not containing i and k) satisfies the ***backdoor criterion*** *relative to (i, k) in G if:*

1. $\boldsymbol{Z} \cap desc(i) = \emptyset$, *and*
2. $\boldsymbol{Z}$ *blocks all "backdoor paths" from i to k in G, i.e., all paths between i and k that start with an arrow into i ($i \leftarrow \ldots k$).*

Example: Backdoor Criterion



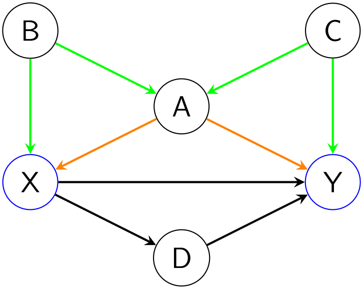
Reminder (Backdoor Criterion)

1. $\boldsymbol{Z} \cap desc(X) = \emptyset$
2. $\boldsymbol{Z}$ *blocks all backdoor paths* $(X \leftarrow \ldots Y)$

Reminder (Block path by Z)

- *Non collider which is in* $\boldsymbol{Z}$.
- *Collider such that neither it nor descendants in* $\boldsymbol{Z}$.

Example: Backdoor Criterion (cont.)



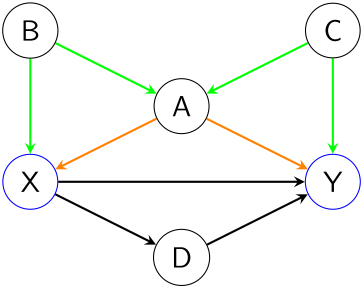
Reminder (Backdoor Criterion)

1. $\boldsymbol{Z} \cap desc(X) = \emptyset$
2. $\boldsymbol{Z}$ *blocks all backdoor paths* $(X \leftarrow \ldots Y)$

Reminder (Block path by Z)

- *Non collider which is in* $\boldsymbol{Z}$.
- *Collider such that neither it nor descendants in* $\boldsymbol{Z}$.

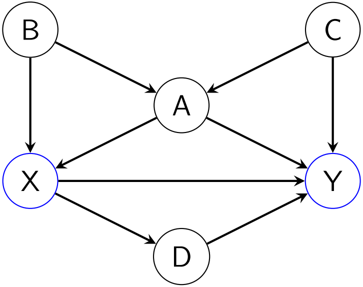
Example: Backdoor Criterion (cont.)



Valid adjustment sets: $\{A, B\}$, $\{A, C\}$ and $\{A, B, C\}$

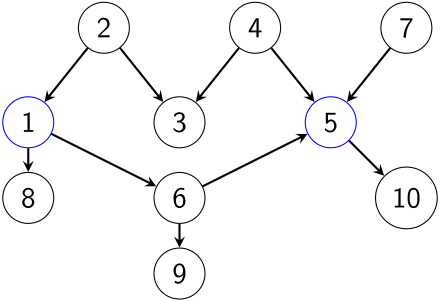
Intuition Backdoor Criterion

- Backdoor paths carry spurious associations from X to Y.
- Paths directed along the arrows from X to Y carry causal associations.
- Blocking backdoor paths ensures that the measured association between X and Y is purely causal.

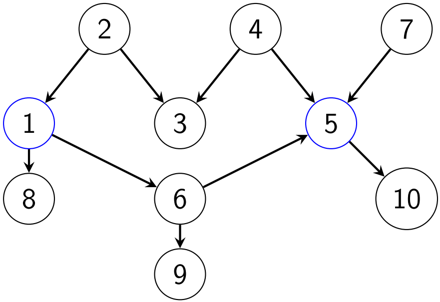


Exercise: Backdoor Criterion

- Interested in the causal effect of X_1 on X_5.
- Select all sets that satisfy the backdoor criterion.



Exercise: Backdoor Criterion (cont.)



- $\{2\}$
- $\{3\}$
- $\{2, 7\}$
- $\{\}$
- $\{7\}$
- $\{2, 3, 4\}$
- $\{8\}$

Reminder (Backdoor Criterion)

1. $\boldsymbol{Z} \cap desc(X) = \emptyset$
2. $\boldsymbol{Z}$ *blocks all backdoor paths* $(X \leftarrow \dots Y)$

Reminder (Block path by Z)

- *Non collider which is in* $\boldsymbol{Z}$.
- *Collider such that neither it nor descendants in* $\boldsymbol{Z}$.

Adjustment Criterion

- The previous criteria were **sufficient** for adjustment.
- The following criterion is **sufficient** and **necessary** for adjustment.

Adjustment Criterion (Shipster et al., Perkovic et al.)

Let $G = (\mathbf{V}, \mathbf{E})$ be a DAG and $i, k \in \mathbf{V}, i \neq k$. A set $\mathbf{Z} \subset \mathbf{V}$ (not containing i and k) satisfies the ***adjustment criterion*** *relative to (i, k) in G if:*

1. *$\mathbf{Z}$ does not contain any descendants of nodes $r \neq i$ on a directed path from i to k in G;*
2. *$\mathbf{Z}$ blocks all paths between i and k in G that are not directed from i to k.*

Adjustment Criterion (cont.)

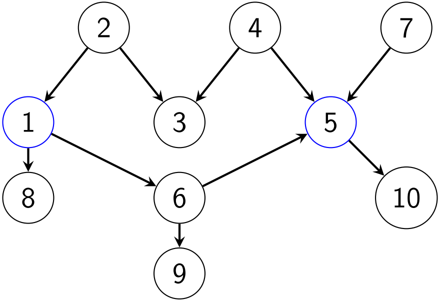
Theorem

$\boldsymbol{Z} \subset \boldsymbol{V}$ satisfies the adjustment criterion relative to (i,k) in a DAG $G = (\boldsymbol{V}, \boldsymbol{E})$ ***if and only if*** *for all p such that (G,p) is a causal Bayesian network, we have:*

$$p(x_k \mid do(x_i)) = \int_{x_{\boldsymbol{z}}} p(x_k \mid x_i, x_{\boldsymbol{z}})\, p(x_{\boldsymbol{z}})\, dx_{\boldsymbol{z}}$$

- **Remark:** it is only sufficient for the identification of **total** causal effects.

Example: Adjustment Criterion



Reminder (Adjust. Criterion)

1. ***Z*** *does not contain descendant of nodes on direct path from i to k*
2. ***Z*** *blocks all paths from i to k that are not directed from i to k*

Reminder (Block path by Z)

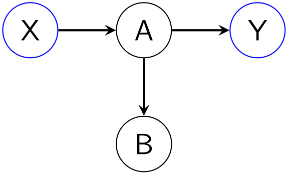
- *Non collider which is in* ***Z****.*
- *Collider such that neither it nor descendants in* ***Z****.*

Determining adjustment sets

The backdoor and the adjustment criterion are (graphical) tools to tackle some problems of "bad controls".

⇝ Should we always adjust for as many variables as possible?

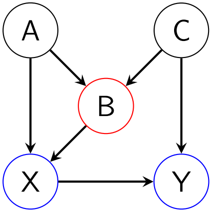
- X: Smoking
- Y: Future miscarriages
- A: Physiological abnormality induced by smoking
- B: Previous miscarriages



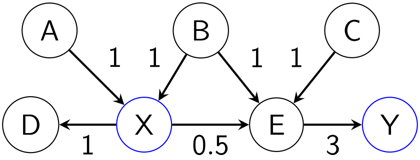
Determining adjustment sets (cont.)

⇝ Is it safe to control only for "pre-treatment" variables?

- X: Smoking
- Y: Adult asthma
- A: Parental smoking
- B: Childhood asthma
- C: Predisposition toward asthma



Example: Determining adjustment sets



Suppose we are interested in the total causal effect of X onto Y ($0.5 \cdot 3 = 1.5$). Lets consider the following three adjustment sets:

- {B,E,A,D,C}
- {B}
- {A,C}

Statistical efficiency in linear SEMs

Goal: Estimate the total causal effects of X on Y in the asymptotically most efficiency way $\rightsquigarrow$ which adjustment set yields the estimator with the smallest asymptotic variance ?

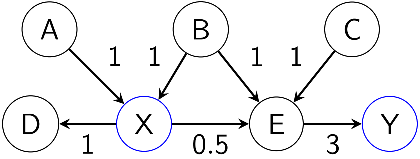
Intuition: for statistically efficient estimators in linear regression setting:

- Avoid variables that are strongly correlated with X;
- Use variables that help to predict Y.

Example: Adjustment Sets and Variance

There are 8 valid adjustment sets **Z**:

- B has to be included;
- E cannot be included;
- A, C, D may be in **Z**.



⇝ Which adjustment set should we use if we want to minimize variance?

Optimal valid adjustment set in Linear SEMs

Let $G = (\mathbf{V},\mathbf{E})$ be a DAG and $i, k \in \mathbf{V}$, $i \neq k$ and $k \in desc(i)$.

Definition (Causal nodes)

$cn(i,k)$: nodes $r \neq i$ on a directed path from i to k in G.

Definition (Forbidden nodes)

$forb(i,k)$: descendants of causal nodes and node i.

Furthermore, let $\hat{\tau}_{ik}^{\mathbf{Z}}$ denote the total causal effect estimator based on $\mathbf{Z}$.

Optimal valid adjustment set in Linear SEMs (cont.)

Theorem

*The **optimal valid adjustment set** is:* $O(i,k) = pa\big(cn(i,k)\big) \setminus forb(i,k))$

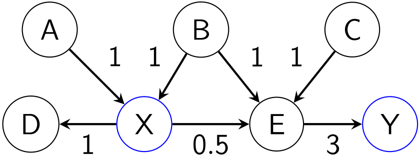
- In other words, for any valid adjustment set $\mathbf{Z}$:

$$\text{a.var}(\hat{\tau}^{\mathbf{O}}_{i,k}) \leq \text{a.var}(\hat{\tau}^{\mathbf{Z}}_{i,k})$$

  where a.var denotes the asymptotic variance.
- If a valid adjustment set exists, $\mathbf{O}$ is one.

Exercise: Adjustment Sets and Variance

- $cn(X, Y) = ?$
- $forb(X, Y) = ?$
- $pa\big(cn(X, Y)\big) = ?$
- $\mathbf{O}(X, Y) = ?$



Reminder (Causal nodes)

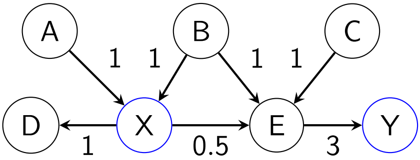
$cn(i, k)$: *nodes* $r \neq i$ *on a directed path from* i *to* k *in* G.

Reminder (Forbidden nodes)

$forb(i, k)$: *descendants of causal nodes and node* i.

Exercise: Adjustment Sets and Variance (cont.)

- $cn(X, Y) = \{E, Y\}$
- $forb(X, Y) = \{X, E, Y\}$
- $pa\big(cn(X, Y)\big) = \{X, B, C, E\}$
- $\mathbf{O}(X, Y) = \{B, C\}$



Intuition: B blocks a backdoor/indirect path, C helps to explain Y and hence reduces the residual variance. Note that A is not included as it is correlated with X and its inclusion would increase the standard error of the estimator.

Reminder: Causal Questions

(1) Drug Treatment

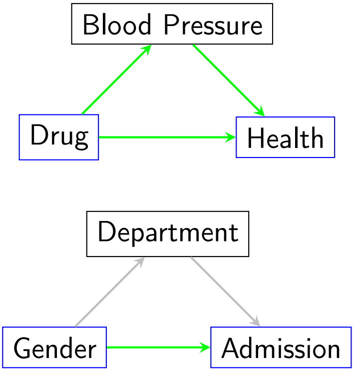
- Does the drug improve health overall?
- Effect on mediator is relevant

▷ *Total causal effect (TCE)*

(2) College Admissions

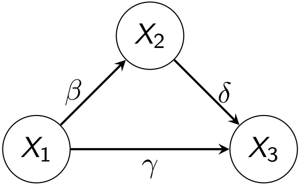
- Are females being discriminated against?
- Effect on mediator is *not* relevant

▷ *Direct causal effect (DCE)*



Controlled Direct Effect

- CDE: one way of defining direct causal effect
- What is the CDE of X_1 on X_3 in this example?

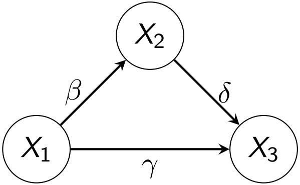


Definition (Controlled direct effect)

$$CDE = \mathbb{E}[X_k \mid do(x_i + 1), do(pa(k) \setminus i)] - \mathbb{E}[X_k \mid do(x_i), do(pa(k) \setminus i)]$$

Controlled Direct Effect - Example

- $X_1 \leftarrow \epsilon_1$
- $X_2 \leftarrow \beta X_1 + \epsilon_2$
- $X_3 \leftarrow \gamma X_1 + \delta X_2 + \epsilon_3$

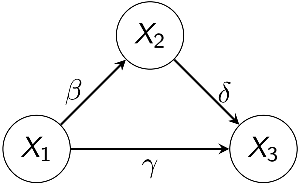


Definition (Controlled direct effect)

$$CDE = \mathbb{E}[X_k \mid do(x_i + 1), do(pa(k) \setminus i)] - \mathbb{E}[X_k \mid do(x_i), do(pa(k) \setminus i)]$$

Controlled Direct Effect - Example (ctd.)

- $X_1 \leftarrow \epsilon_1$
- $X_2 \leftarrow \beta X_1 + \epsilon_2$
- $X_3 \leftarrow \gamma X_1 + \delta X_2 + \epsilon_3$



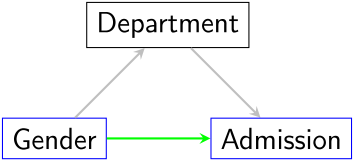
Definition (Controlled direct effect)

$$CDE = \mathbb{E}[X_k \mid do(x_i+1), do(pa(k) \setminus i)] - \mathbb{E}[X_k \mid do(x_i), do(pa(k) \setminus i)]$$

Here:

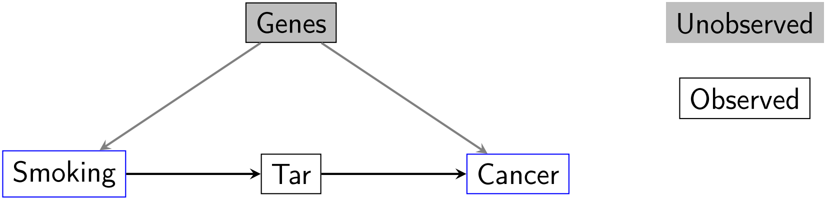
- $\mathbb{E}[X_3 \mid \mathrm{do}(x_1+1), \mathrm{do}(x_2)] - \mathbb{E}[X_3 \mid \mathrm{do}(x_1), \mathrm{do}(x_2)] = \gamma(x_1+1) + \delta x_2 - (\gamma x_1 + \delta x_2) = \gamma$

CDE - Admissions example



- In this case: $\mathbb{E}[A|G = f, D = d] - \mathbb{E}[A|G = m, D = d]$
- Can determine controlled direct effect of gender on admission for each department
- Can become more involved for other mediating variables

Example: Smoking



Problem: Cannot use backdoor or adjustment criterion since genes could not be observed.

Idea: Can we make use of tar being a mediator of the effect of smoking on cancer?

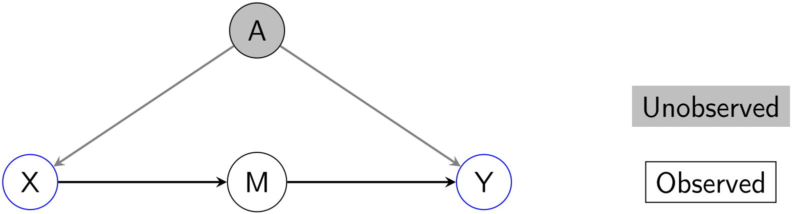
Frontdoor Criterion

Frontdoor Criterion (Pearl)

Let $G = (\boldsymbol{V}, \boldsymbol{E})$ be a DAG and $i, k \in \boldsymbol{V}, i \neq k$. A set $\boldsymbol{M} \subset \boldsymbol{V}$ (not containing i and k) satisfies the ***frontdoor criterion*** *relative to (i, k) in G if:*

1. *$\boldsymbol{M}$ blocks all directed paths from i to k in G, and*
2. *There are no unblocked backdoor paths from i to $\boldsymbol{M}$ in G, and*
3. *i blocks all backdoor paths from $\boldsymbol{M}$ to k in G.*

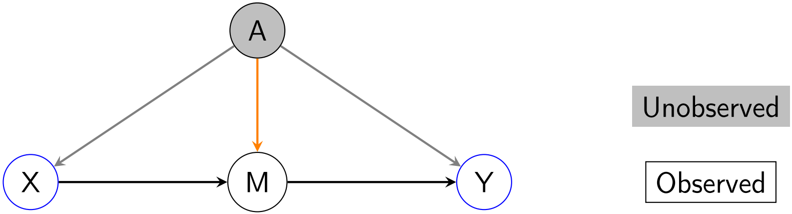
Example: Frontdoor Criterion



Reminder (Frontdoor Criterion)

1. *M blocks all directed paths from X to Y*
2. *There are no unblocked backdoor paths from X to M*
3. *X blocks all backdoor paths from M to Y*

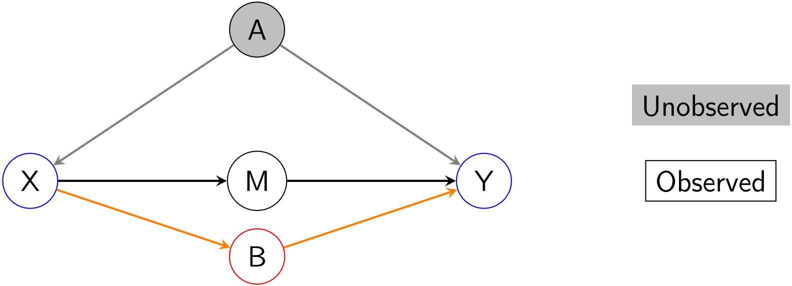
Example: Frontdoor Criterion



Reminder (Frontdoor Criterion)

1. *M blocks all directed paths from X to Y*
2. *There are no unblocked backdoor paths from X to M*
3. *X blocks all backdoor paths from M to Y*

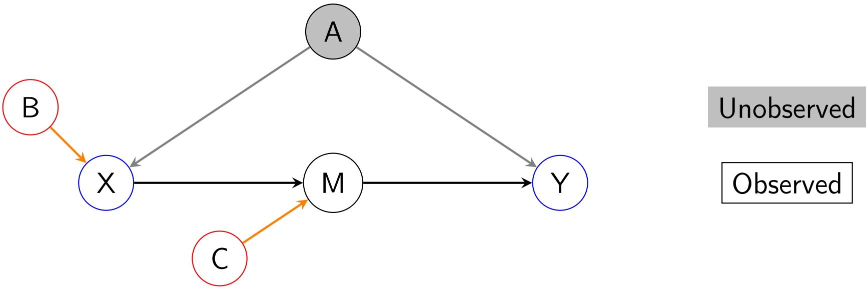
Example: Frontdoor Criterion



Reminder (Frontdoor Criterion)

1. *M blocks all directed paths from X to Y*
2. *There are no unblocked backdoor paths from X to M*
3. *X blocks all backdoor paths from M to Y*

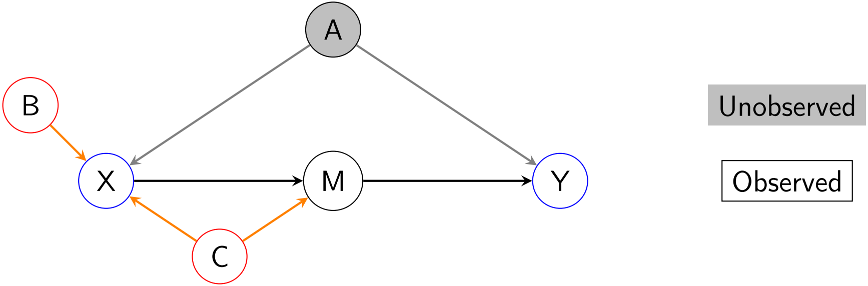
Example: Frontdoor Criterion



Reminder (Frontdoor Criterion)

1. *M blocks all directed paths from X to Y*
2. *There are no unblocked backdoor paths from X to M*
3. *X blocks all backdoor paths from M to Y*

Example: Frontdoor Criterion



Reminder (Frontdoor Criterion)

1. *M blocks all directed paths from X to Y*
2. *There are no unblocked backdoor paths from X to M*
3. *X blocks all backdoor paths from M to Y*

Wrap-Up

The **interventional distribution** $P(Y \mid \mathrm{do}(X))$ is identifiable if:

- There is a valid adjustment set for (X,Y) - **backdoor criterion** and **adjustment criterion**
- If we can apply the **frontdoor criterion**
- Other approaches (e.g. *instrumental variables* - later in the course)

Thank you for your attention.