THE UNIVERSITY *of* EDINBURGH

Methods for Causal Inference
Lecture 11: Front-Door Criterion

Ava Khamseh

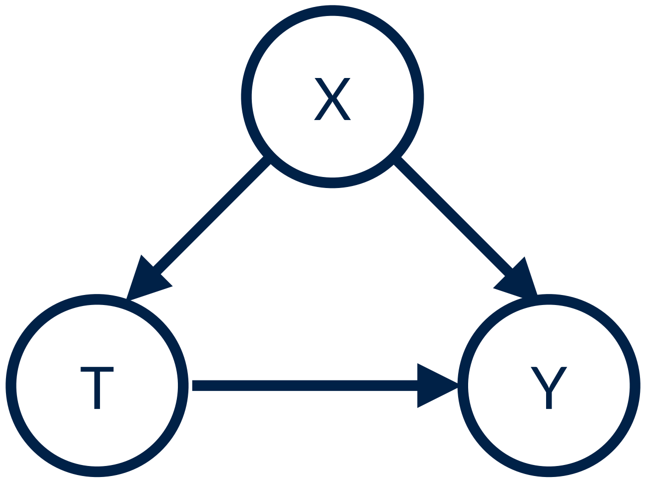
School of Informatics

2023-2024

The adjustment formula

T: Drug usage
X: Sex
Y: Recovery

Use Pm as an intermediate tool



To know how effective the drugs is in the population, compare the **hypothetical interventions** by which

(i) the drug is administered uniformly to the entire population do(T=1) **vs**

(ii) complement, i.e., everyone is prevented from taking the drug do(T=0)

Aim: Estimate the difference (**Average Causal Effect ACE, aka ATE**)

$$\boxed{p(Y=1|do(T=1)) - p(Y=1|do(T=0))}$$

Pearl, Causal Inference in Statistics (2016)

The Backdoor Criterion

Under what conditions does a causal model permit computing the causal effect of one variable on another, from **data** obtained from **passive observations**, with **no intervention**? i.e., Under what conditions is the structure of a causal graph sufficient of computing a causal effect from a given data set? **Identifiability**

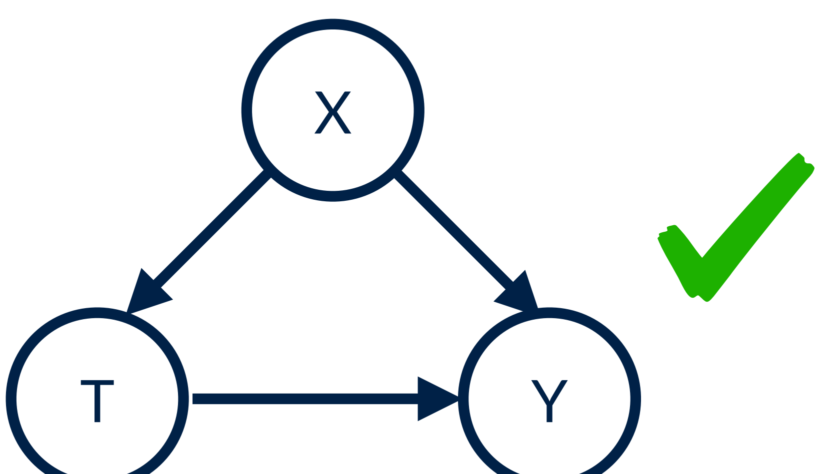
Backdoor Criterion: Given an ordered pair of variables (T,Y) in a DAG G, a set of variables X satisfies the backdoor criterion relative to (T,Y) if:

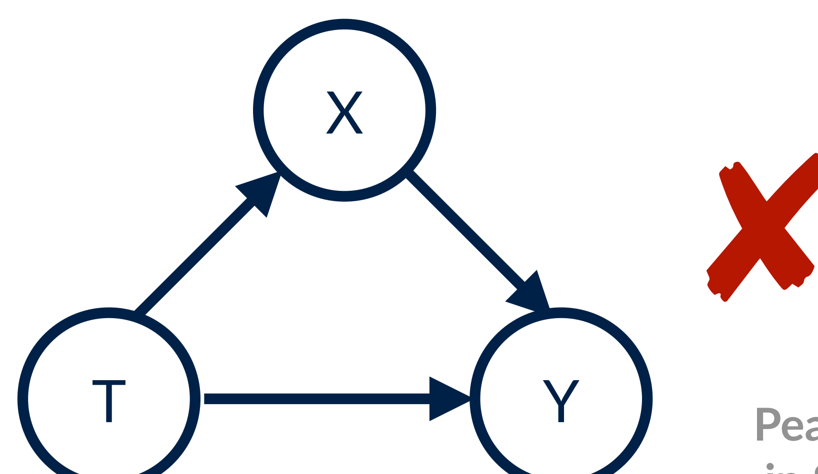
(i) no node in X is a descendent of T

(ii) X block every path between T and Y that contains an arrow into T

If X satisfies the backdoor criterion then the causal effect of T on Y is given by:

$$p(Y = y|do(T = t)) = \sum_x p(Y = y|T = t, X = x)p(X = x)$$





Pearl, Causal Inference in Statistics (2016)

The Backdoor Criterion

Under what conditions does a causal model permit computing the causal effect of one variable on another, from **data** obtained from **passive observations**, with **no intervention**? i.e., Under what conditions is the structure of a causal graph sufficient of computing a causal effect from a given data set? **Identifiability**

Backdoor Criterion: Given an ordered pair of variables (T,Y) in a DAG G, a set of variables X satisfies the backdoor criterion relative to (T,Y) if:

(i) no node in X is a descendent of T
(ii) X block every path between T and Y that contains an arrow into T

If X satisfies the backdoor criterion then the causal effect of T on Y is given by:

$$p(Y = y | do(T = t)) = \sum_{x} p(Y = y | T = t, X = x) p(X = x)$$

In other words, condition on a set of nodes X such that:

(i) We block all spurious paths between T and Y
(ii) We leave all direct paths from T to Y unperturbed
(iii) We create no new spurious paths (do not unblock any new paths)

Pearl, Causal Inference in Statistics (2016)

Recall ...

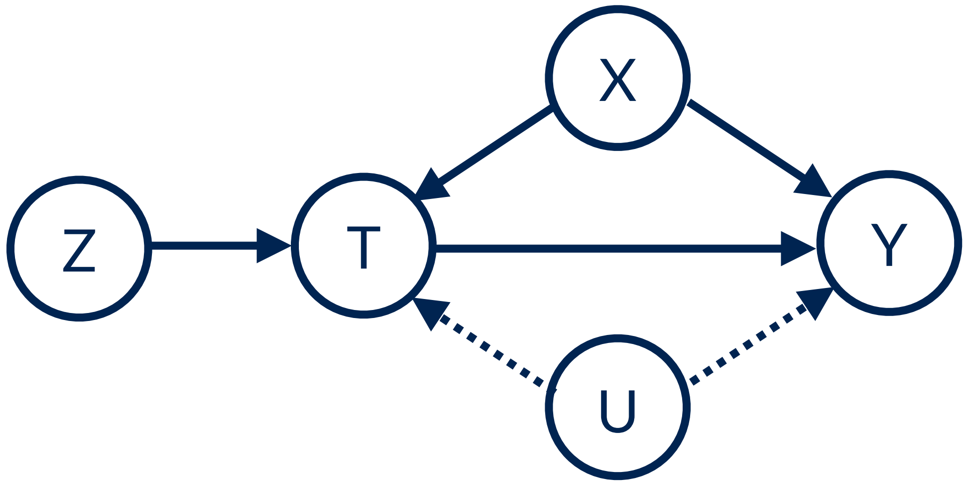
- Backdoor does not exhaust all ways of estimating causal effects from a graph

- Front-door criterion can still be used for patterns that do not satisfy the backdoor criterion

- Example: Smoking and lung cancer (1970), industry argued to prevent antismoking regulation by suggesting that the correlation could be explained by a carcinogenic genotype that induces a craving for nicotine

- Recall sensitivity analysis

- Recall instrumental variable approach



Pearl, Causal Inference in Statistics (2016)

Instrumental Variable assumptions

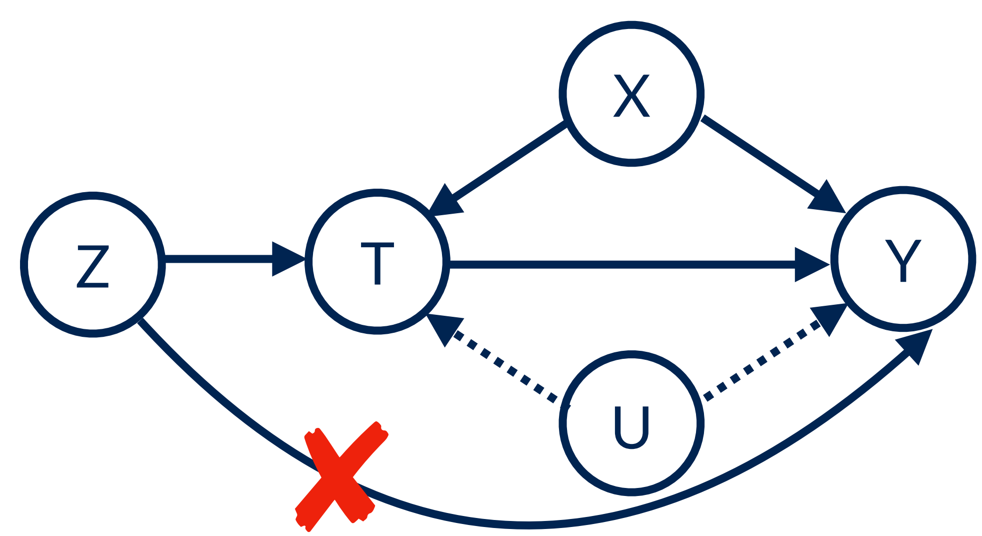
- **SUTVA**: Potential outcomes for each individual i are unrelated to the treatment status of other individuals:

$$Y^{(i)}(\mathbf{Z}, \mathbf{T}) = Y^{(i)}(Z^{(i)}, T^{(i)}) \;,\; |\mathbf{Z}| = |\mathbf{T}| = N \text{ individuals}$$

- Non-zero average/relevant: Treatment assignment Z associated with the treatment $\mathbb{E}\left[\left(T^{(i)}|z=1\right) - \left(T^{(i)}|z=0\right)\right]$
- Treatment assignment Z is random (Z and Y do not share a cause).

$$\left(Y^{(i)}|z=1,t\right) = \left(Y^{(i)}|z=0,t\right)$$

- **Exclusion Restriction**: Any effect of Z on Y is via an effect of Z on T, i.e., Z should not affect Y when T is held constant



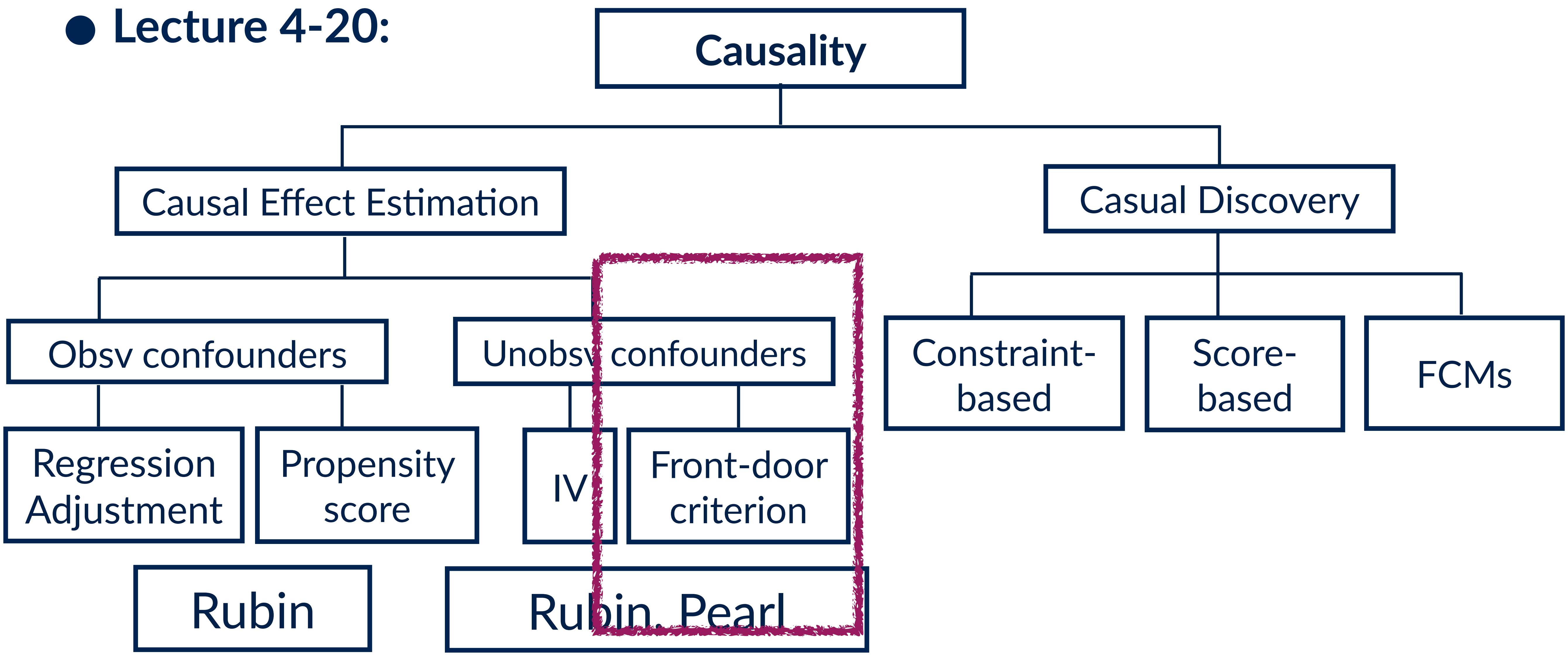
- **Monotonicity** (increasing encouragement “dose” increases probability of treatment, no defiers):

$$\left(T^{(i)}|z=1\right) \geq \left(T^{(i)}|z=0\right)$$

Rubin 1996

Overview of the course

- **Lecture 1**: Introduction & Motivation, why do we care about causality? Why deriving causality from observational data is non-trivial.
- **Lecture 2:** Recap of probability theory, variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM
- **Lecture 4-20:**



Pearl's Front-Door Criterion: An Example

- Fig (a): The graph does not satisfy the backdoor, since the quantity we need to condition on to block the path, i.e. the genotype, is unobserved
- Fig (b): Additional measurement available: tar deposits in patients lungs
- Fig (b) still does not satisfy the backdoor criterion but we can determine the causal effect:

$$p(Y = y|do(X = x))$$

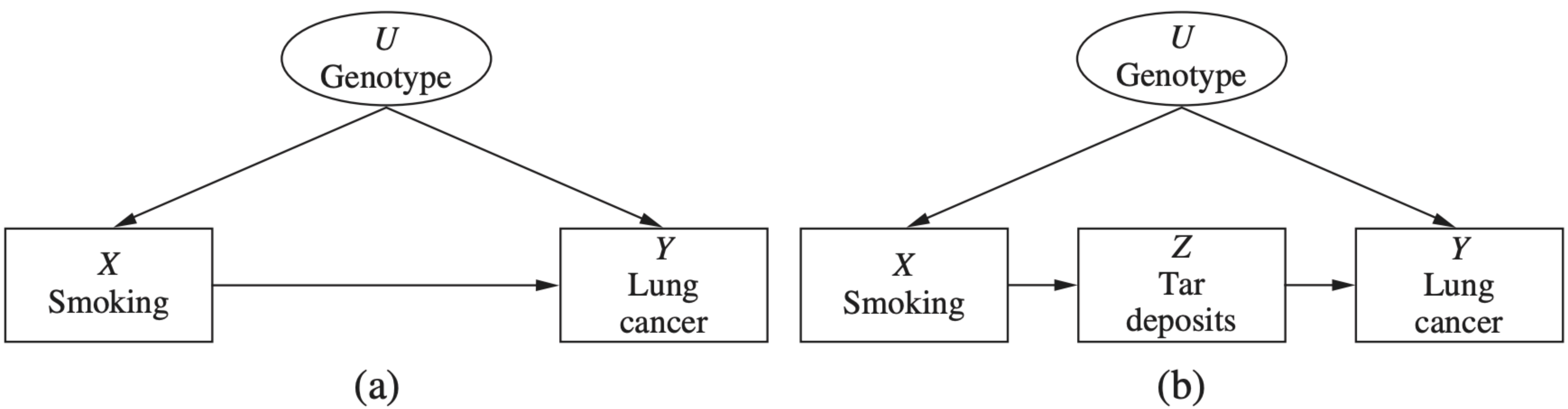


Figure 3.10 A graphical model representing the relationships between smoking (X) and lung cancer (Y), with unobserved confounder (U) and a mediating variable Z

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion: A crafted example

Interpretation 1: Tobacco industry

Beneficial effect of smoking:

15% of smokers have developed lung cancer vs 90.25% of non-smokers

within tar and non-tar subgroups, smokers have a much lower percentage of cancer than non-smokers (numbers in the table are engineered to illustrate the point that observations are not to be trusted)

Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar 400		No tar 400		All subjects 800	
	Smokers	Nonsmokers	Smokers	Nonsmokers	Smokers	Nonsmokers
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323 (85%)	1 (5%)	18 (90%)	38 (10%)	341 (85%)	39 (9.75%)
Cancer	57 (15%)	19 (95%)	2 (10%)	342 (90%)	59 (15%)	361 (90.25%)

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion: A crafted example

Interpretation 2: Anti-smoking lobbyists

Smoking **increases** the risk of lung cancer

If one chooses to smoke, then one's chances of building tar deposits are 95% (380/400) vs 5% (20/400) for the non-smokers.

To evaluate effect of tar, look at **smokers and non-smokers separately**. Tar has harmful effects in both groups: in smokers it increases risk of cancer from 10% to 15% and in non-smokers 90% to 95%. Therefore: Smoking -> tar -> cancer.

Regardless of any natural craving, avoid harmful tar by not smoking.

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323 (85%)	18 (90%)	1 (5%)	38 (10%)	324 (81%)	56 (19%)
Cancer	57 (15%)	2 (10%)	19 (95%)	342 (90%)	76 (19%)	344 (81%)

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion

X -> Z is **identifiable**, since no back path from X and Z: $X \leftarrow U \rightarrow Y \leftarrow Z$

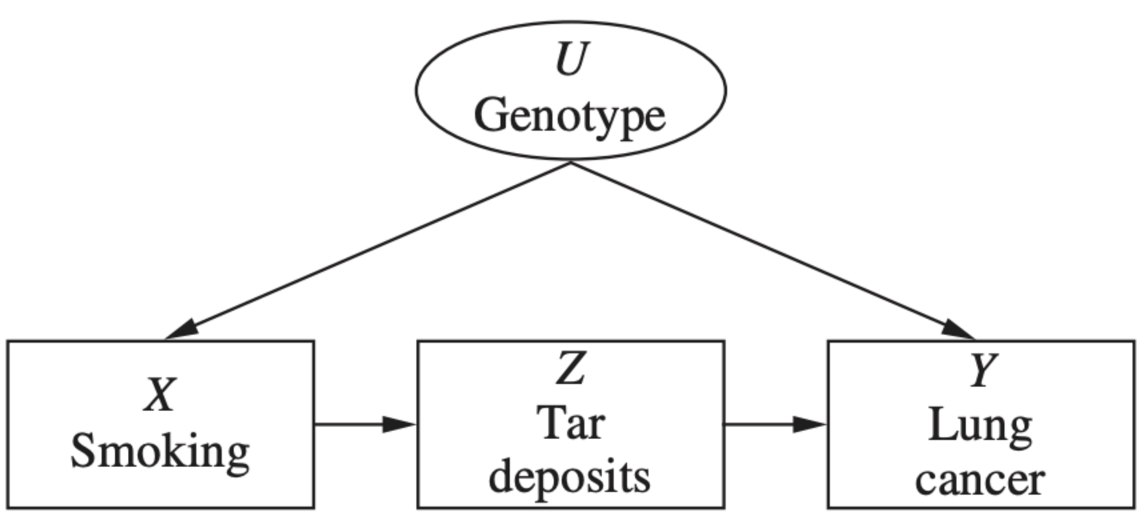
$$p(Z = z | do(X = x)) = p(Z = z | X = x) \qquad *$$

Z -> Y is **identifiable**, since backdoor from Z to Y:

$$Z \leftarrow X \leftarrow U \rightarrow Y$$

is **blocked** by conditioning on X:

$$p(Y = y | do(Z = z)) = \sum_x p(Y = y | Z = z, X = x) p(X = x) \qquad **$$



Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion

Letting z be the value Z takes when setting X=x, from the graph, we have:

$$p(Y|do(X=x)) = p(Y|do(X=x),Z) = p(Y|do(Z=z))$$

Then summing over all states z of Z:

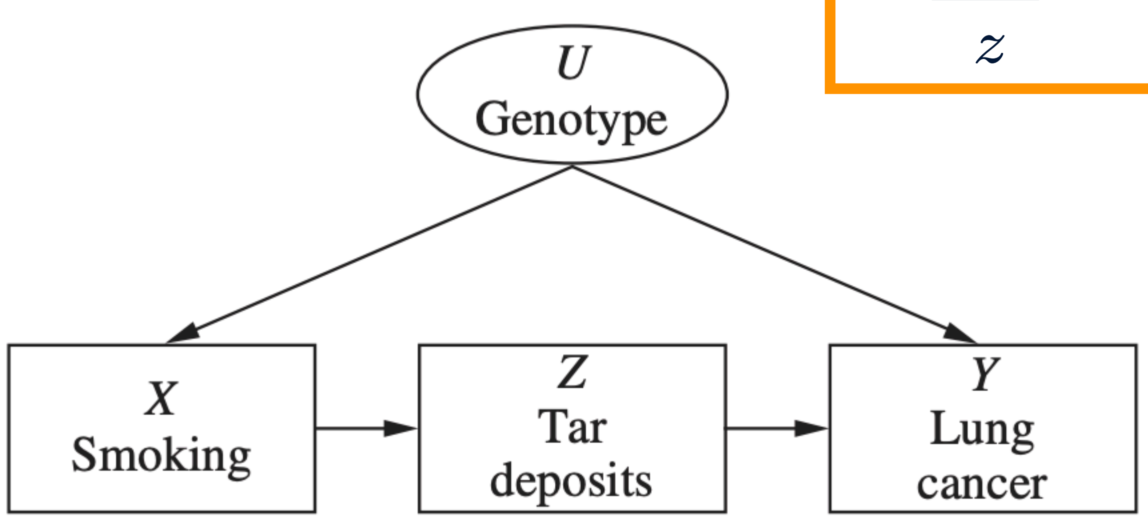
$$p(Y=y|do(X=x)) = \sum_z p(Y=y,z|do(X=x))$$ Total prob rule

Product rule:

$$= \sum_z p(Y=y|z,do(X=x))p(z|do(X=x))$$

Line 1

$$\boxed{= \sum_z p(Y=y|do(Z=z))p(z|do(X=x))}$$



Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion

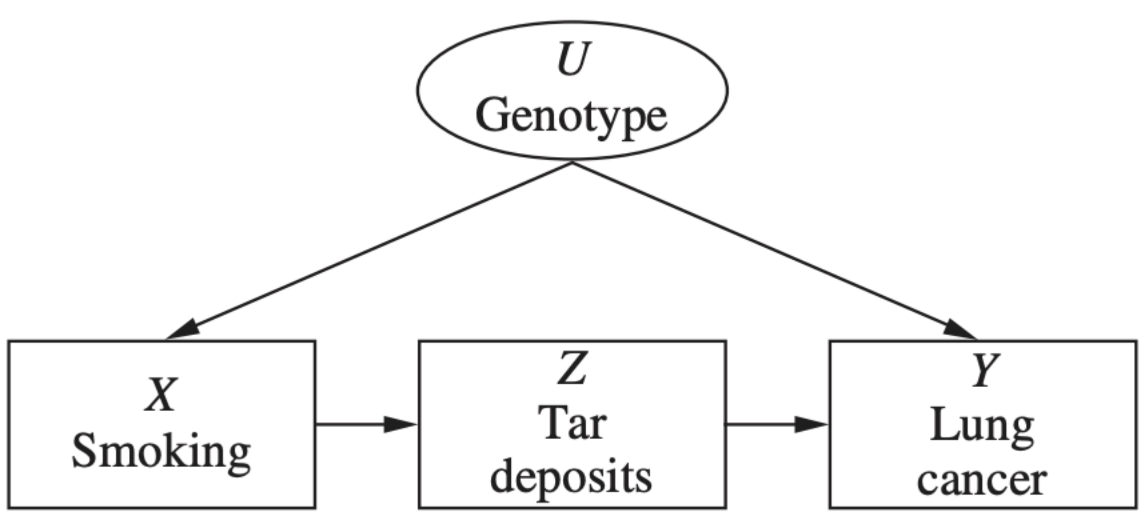
$$p(Z = z|do(X = x)) = p(Z = z|X = x) \quad *$$

$$p(Y = y|do(Z = z)) = \sum_{x'} p(Y = y|Z = z, X = x')p(X = x') \quad **$$

$$p(Y = y|do(X = x)) = \boxed{\sum_{z} p(Y = y|do(Z = z))p(Z = z|do(X = x))}$$

Using $*$ and $**$ summing over all states z of Z:

$$p(Y = y|do(X = x)) = \sum_{z} \sum_{x'} p(Y = y|Z = z, X = x')p(X = x')p(Z = z|X = x)$$



Front-door formula

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion: Which group is right?

$$p(Y=y|do(X=x)) = \sum_z \sum_{x'} p(Y=y|Z=z, X=x')p(X=x')p(Z=z|X=x)$$

$$\begin{aligned} p(Y=1|do(X=1)) = {} & p(Y=1|z=0,x'=0)p(x'=0)p(z=0|x=1) \\ & + p(Y=1|z=0,x'=1)p(x'=1)p(z=0|x=1) \\ & + p(Y=1|z=1,x'=0)p(x'=0)p(z=1|x=1) \\ & + p(Y=1|z=1,x'=1)p(x'=1)p(z=1|x=1) \\ & = 0.5475 \end{aligned}$$

$342/380$

$2/20$

$19/20$

$57/380$

$20/400$

0.5

$380/400$

$$p(Y=1|do(X=0)) = 0.5025$$

Average Causal Effect ACE: $p(Y=1|do(X=1)) - p(Y=1|do(X=0)) = 0.045$

4.5% increase

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323	18	1	38	324	56
	(85%)	(90%)	(5%)	(10%)	(81%)	(19%)
Cancer	57	2	19	342	76	344
	(15%)	(10%)	(95%)	(90%)	(19%)	(81%)

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion: Which group is right?

$$p(Y=y|do(X=x)) = \sum_z \sum_{x'} p(Y=y|Z=z,X=x')p(X=x')p(Z=z|X=x)$$

$$\begin{aligned}
p(Y=1|do(X=1)) &= \underbrace{p(Y=1|z=0,x'=0)}_{342/380}p(x'=0)\underbrace{p(z=0|x=1)}_{20/400} \\
&+ \underbrace{p(Y=1|z=0,x'=1)}_{2/20}p(x'=1)\underbrace{p(z=0|x=1)}_{20/400} \\
&+ \underbrace{p(Y=1|z=1,x'=0)}_{19/20}p(x'=0)\underbrace{p(z=1|x=1)}_{380/400} \\
&+ \underbrace{p(Y=1|z=1,x'=1)}_{57/380}\underbrace{p(x'=1)}_{0.5}\underbrace{p(z=1|x=1)}_{380/400} \\
&= 0.5475
\end{aligned}$$

$$p(Y=1|do(X=0)) = 0.5025$$

Average Causal Effect ACE: $p(Y=1|do(X=1)) - p(Y=1|do(X=0)) = 0.045$

4.5% increase

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323 (85%)	18 (90%)	1 (5%)	38 (10%)	324 (81%)	56 (19%)
Cancer	57 (15%)	2 (10%)	19 (95%)	342 (90%)	76 (19%)	344 (81%)

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion: Which group is right?

$$\boxed{p(Y=y|do(X=x)) = \sum_z \sum_{x'} p(Y=y|Z=z, X=x')p(X=x')p(Z=z|X=x)}$$

$$\begin{aligned}
p(Y=1|do(X=1)) = &\underbrace{p(Y=1|z=0,x'=0)}_{342/380}\underbrace{p(x'=0)}_{0.5}\underbrace{p(z=0|x=1)}_{20/400} \\
&+ \underbrace{p(Y=1|z=0,x'=1)}_{2/20}\underbrace{p(x'=1)}_{0.5}\underbrace{p(z=0|x=1)}_{20/400} \\
&+ \underbrace{p(Y=1|z=1,x'=0)}_{19/20}\underbrace{p(x'=0)}_{0.5}\underbrace{p(z=1|x=1)}_{380/400} \\
&+ \underbrace{p(Y=1|z=1,x'=1)}_{57/380}\underbrace{p(x'=1)}_{0.5}\underbrace{p(z=1|x=1)}_{380/400} \\
&= 0.5475
\end{aligned}$$

$$p(Y=1|do(X=0)) = 0.5025$$

Average Causal Effect ACE: $p(Y=1|do(X=1)) - p(Y=1|do(X=0)) = 0.045$

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323 (85%)	18 (90%)	1 (5%)	38 (10%)	324 (81%)	56 (19%)
Cancer	57 (15%)	2 (10%)	19 (95%)	342 (90%)	76 (19%)	344 (81%)

4.5% increase

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion: Which group is right?

$$p(Y=y|do(X=x)) = \sum_z \sum_{x'} p(Y=y|Z=z, X=x')p(X=x')p(Z=z|X=x)$$

$$\begin{aligned} p(Y=1|do(X=1)) = & \boxed{p(Y=1|z=0,x'=0)}p(x'=0)p(z=0|x=1) \\ & + p(Y=1|z=0,x'=1)p(x'=1)p(z=0|x=1) \\ & + p(Y=1|z=1,x'=0)p(x'=0)p(z=1|x=1) \\ & + p(Y=1|z=1,x'=1)p(x'=1)p(z=1|x=1) \\ & = 0.5475 \end{aligned}$$

$342/380$

$2/20$

$19/20$

$57/380$

0.5

$20/400$

$380/400$

$$p(Y=1|do(X=0)) = 0.5025$$

Average Causal Effect ACE: $p(Y=1|do(X=1)) - p(Y=1|do(X=0)) = 0.045$

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323 (85%)	18 (90%)	1 (5%)	38 (10%)	324 (81%)	56 (19%)
Cancer	57 (15%)	2 (10%)	19 (95%)	342 (90%)	76 (19%)	344 (81%)

4.5% increase

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion: Which group is right?

$$p(Y=y|do(X=x)) = \sum_z \sum_{x'} p(Y=y|Z=z, X=x')p(X=x')p(Z=z|X=x)$$

$$\begin{aligned} p(Y=1|do(X=1)) = \; & p(Y=1|z=0,x'=0)p(x'=0)p(z=0|x=1) \\ & + \boxed{p(Y=1|z=0,x'=1)}p(x'=1)p(z=0|x=1) \\ & + p(Y=1|z=1,x'=0)p(x'=0)p(z=1|x=1) \\ & + p(Y=1|z=1,x'=1)p(x'=1)p(z=1|x=1) \\ & = 0.5475 \end{aligned}$$

$342/380$

$2/20$

$19/20$

$57/380$

$20/400$

0.5

$380/400$

$$p(Y=1|do(X=0)) = 0.5025$$

Average Causal Effect ACE: $p(Y=1|do(X=1)) - p(Y=1|do(X=0)) = 0.045$

4.5% increase

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323	18	1	38	324	56
	(85%)	(90%)	(5%)	(10%)	(81%)	(19%)
Cancer	57	2	19	342	76	344
	(15%)	(10%)	(95%)	(90%)	(19%)	(81%)

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion: Which group is right?

$$p(Y=y|do(X=x)) = \sum_z \sum_{x'} p(Y=y|Z=z, X=x')p(X=x')p(Z=z|X=x)$$

$$\begin{aligned} p(Y=1|do(X=1)) = {} & \underbrace{p(Y=1|z=0,x'=0)}_{342/380}\underbrace{p(x'=0)}_{0.5}\underbrace{p(z=0|x=1)}_{20/400} \\ & + \underbrace{p(Y=1|z=0,x'=1)}_{2/20}\underbrace{p(x'=1)}_{0.5}\underbrace{p(z=0|x=1)}_{20/400} \\ & + \underbrace{\boxed{p(Y=1|z=1,x'=0)}}_{19/20}\underbrace{p(x'=0)}_{0.5}\underbrace{p(z=1|x=1)}_{380/400} \\ & + \underbrace{p(Y=1|z=1,x'=1)}_{57/380}\underbrace{p(x'=1)}_{0.5}\underbrace{p(z=1|x=1)}_{380/400} \\ = {} & 0.5475 \end{aligned}$$

$$p(Y=1|do(X=0)) = 0.5025$$

Average Causal Effect ACE: $p(Y=1|do(X=1)) - p(Y=1|do(X=0)) = 0.045$

4.5% increase

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323	18	1	38	324	56
	(85%)	(90%)	(5%)	(10%)	(81%)	(19%)
Cancer	57	2	19	342	76	344
	(15%)	(10%)	(95%)	(90%)	(19%)	(81%)

Pearl, Causal Inference in Statistics (2016)

Pearl's Front-Door Criterion: Which group is right?

$$p(Y=y|do(X=x)) = \sum_z \sum_{x'} p(Y=y|Z=z, X=x')p(X=x')p(Z=z|X=x)$$

$$\begin{aligned}
p(Y=1|do(X=1)) = \; & p(Y=1|z=0,x'=0)p(x'=0)p(z=0|x=1) \\
& + p(Y=1|z=0,x'=1)p(x'=1)p(z=0|x=1) \\
& + p(Y=1|z=1,x'=0)p(x'=0)p(z=1|x=1) \\
& + \boxed{p(Y=1|z=1,x'=1)}p(x'=1)p(z=1|x=1) \\
= \; & 0.5475
\end{aligned}$$

342/380

2/20

19/20

57/380

20/400

0.5

380/400

$$p(Y=1|do(X=0)) = 0.5025$$

Average Causal Effect ACE: $p(Y=1|do(X=1)) - p(Y=1|do(X=0)) = 0.045$

4.5% increase

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (numbers in thousands)

	Smokers 400		Nonsmokers 400		All subjects 800	
	Tar	No tar	Tar	No tar	Tar	No tar
	380	*20*	*20*	*380*	*400*	*400*
No cancer	323 (85%)	18 (90%)	1 (5%)	38 (10%)	324 (81%)	56 (19%)
Cancer	57 (15%)	2 (10%)	19 (95%)	342 (90%)	76 (19%)	344 (81%)

Pearl, Causal Inference in Statistics (2016)

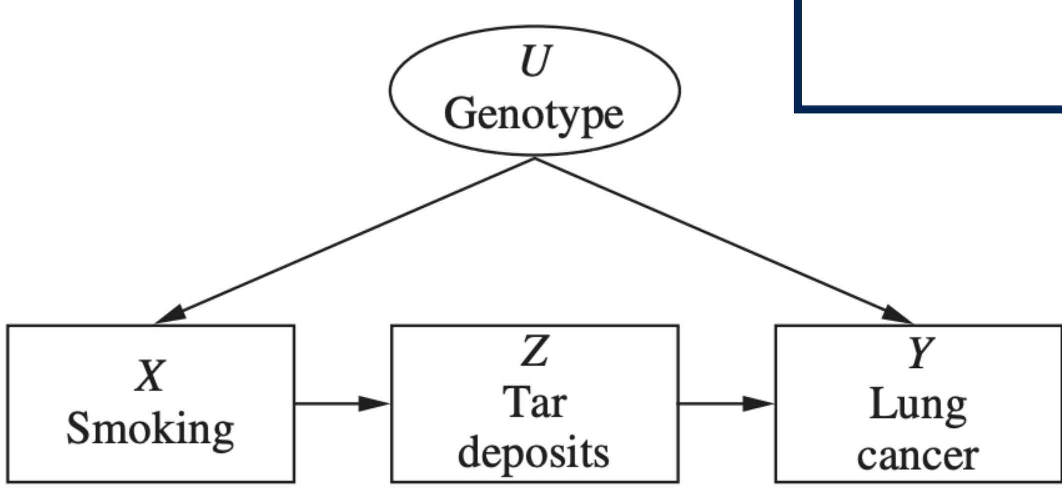
Pearl's Front-Door Adjustment

Front-door criterion: A set of variables Z is said to satisfy the front-door criterion relative to (X,Y) if:

1. Z intercepts all directed paths from X to Y
2. There is no unblocked path from X to Z
3. All backdoor paths from Z to Y are blocked by X

Front-door adjustment: If Z satisfied the front-door criterion relative to (X,Y), and if p(x,z)>0, then the causal effect of X on Y is identifiable and is given by:

$$\boxed{p(y|do(x)) = \sum_z p(z|x) \sum_{x'} p(y|x', z)p(x')}$$



Pearl, Causal Inference in Statistics (2016)

Pearl's Do Calculus

- Do-calculus: Contains, as subsets:
  - Backdoor criterion
  - Front-door criterion

- Allows analysis of more intricate structure beyond back- and front-door

- Uncovers **all** causal effects that can be identified from a given causal graph

- Power of causal graphs is not just representation but towards **discovery** of causal information

Pearl, Causal Inference in Statistics (2016)

Causal Inference

- **Model** a causal inference problem with assumptions manifest in Causal Graphical Models [**Pearl]**

- **Identify** an expression for the causal effect under these assumptions ("causal estimand"), [**Pearl**]

- **Estimate** the expression using statistical methods such as matching or instrumental variables, [**Rubin**'s Potential Outcomes]

- **Verify** the validity of the estimate using a variety of robustness checks.

Causal Inference: Packages and simulations

Simple DoWhy tutorials on my GitHub 'Causality in Biomedicine'

DoWhy tutorials:
https://www.pywhy.org/dowhy/v0.9.1/index.html
https://github.com/py-why/dowhy

CausalGraphicalModels Tutorials:
https://github.com/ijmbarr/causalgraphicalmodels
Adjusting for the wrong variable: http://www.degeneratestate.org/posts/2018/Jul/10/causal-inference-with-python-part-2-causal-graphical-models/
Front-door: http://www.degeneratestate.org/posts/2018/Sep/03/causal-inference-with-python-part-3-frontdoor-adjustment/

Also see ML extensions to DoWhy, e.g. EconML:
https://github.com/microsoft/EconML

Methods for Causal Inference
Lecture 11: Front-Door Criterion

Ava Khamseh

School of Informatics
2023-2024