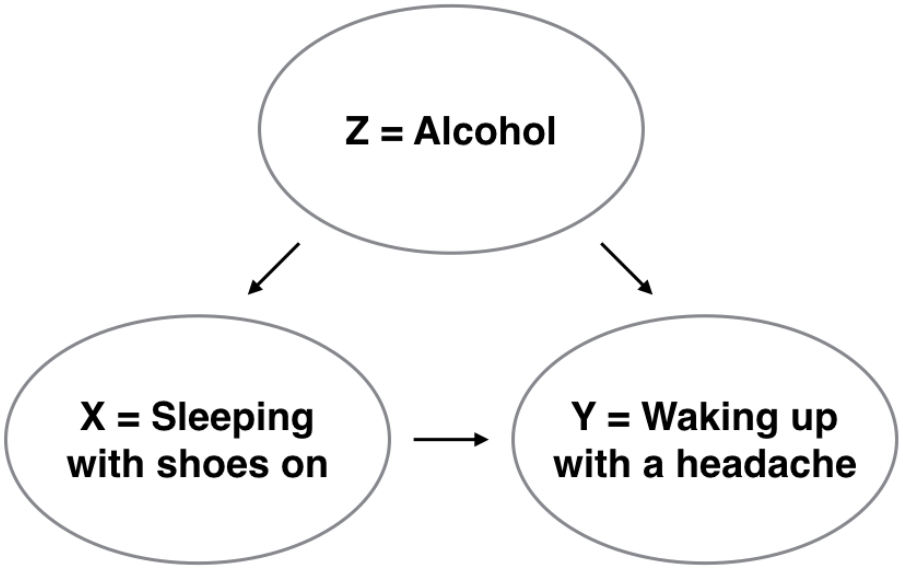
Causal inference

How to go beyond association to estimate cause-effect relations and do intervention planning?



Beate Sick

Topics of today

- Humans and scientists want/need to understand the “WHY”
- Correlation versus association
- Pearl’s ladder of causation
- Causal graphical models
- Backdoor criterion: how to set up an regression model for intervention planning

We need to understand causality to plan intervention

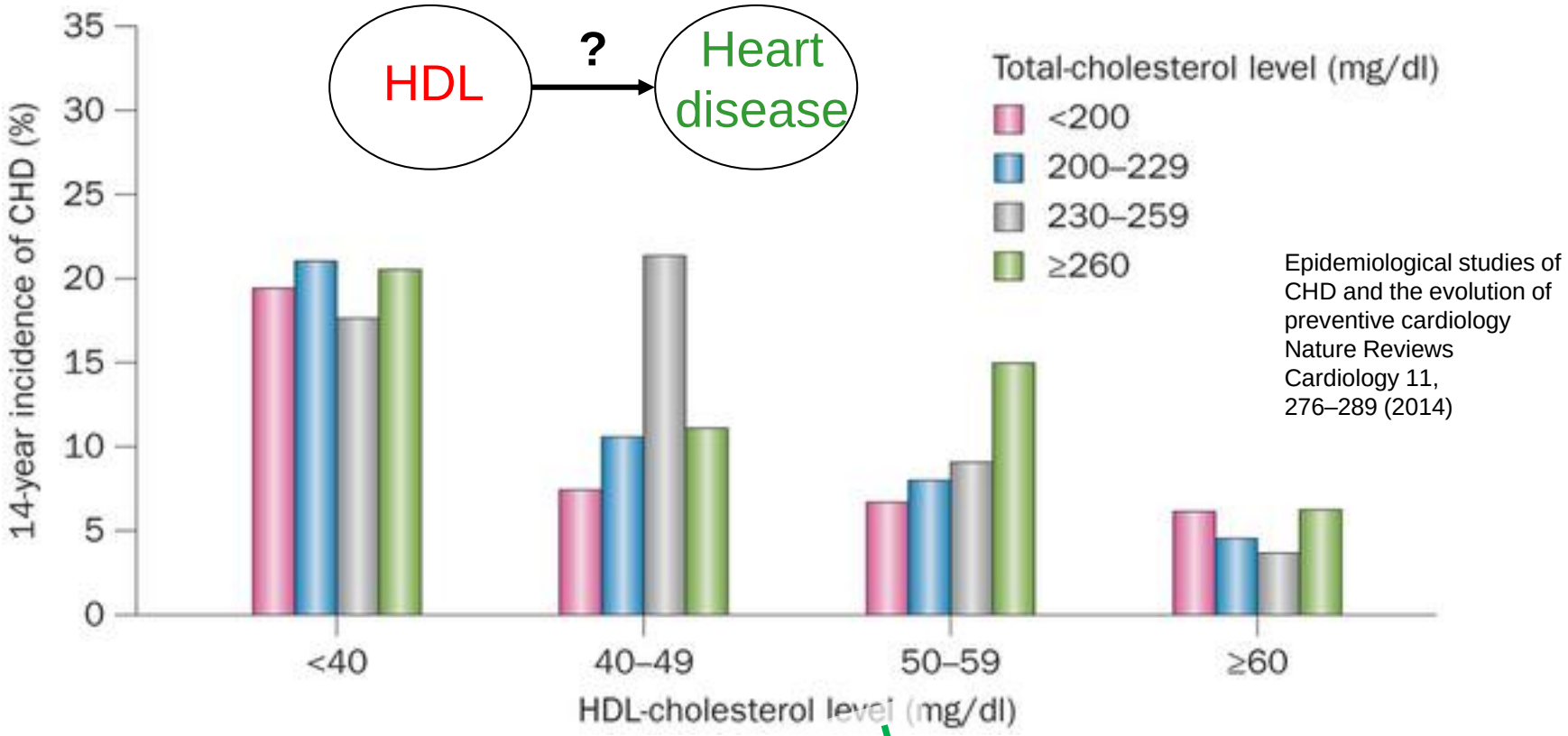
Do violent video games cause violence among young people?

Then ban them!

Does unconditional basic income crank up economy?

Then launch it!

For intervention planning we need to understand the WHY



Epidemiological studies of CHD and the evolution of preventive cardiology Nature Reviews Cardiology 11, 276–289 (2014)

HDL gives a strong negative association with heart disease in cross-sectional studies and is the strongest predictor of future events in prospective studies.

Prediction worked

Roche tested the effect of drug "dalcetrapib" in phase III on 15'000 patients which proved to boost HDL ("good cholesterol") but failed to prevent heart diseases. Roche stomped the failed trial on May 2012 and immediately lost $5billion of its market capitalization.

Planned intervention failed

Probabilistic versus causal reasoning

Traditional statistics, machine learning, Bayesian networks

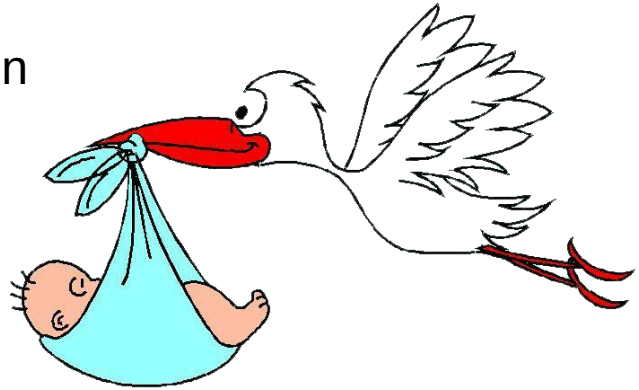
- About associations (are stork population and human birth number per year are associated?)
- The dream is a models for the joined distribution of the data
- Conditional distribution are modeled by regression or classification

  (if we observe a certain number of storks, what is our best estimate of human birth rate?)

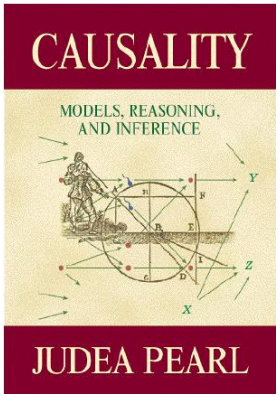
Causal models

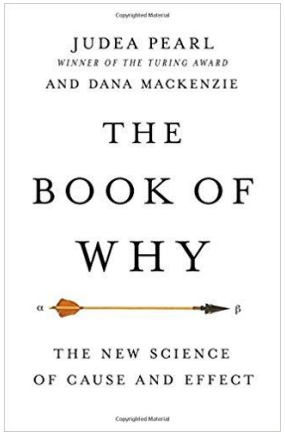
- About causation (do storks do affect human birth rate?)
- The dream is a models for the data generation
- Predict results of interventions

  (if we change the number of storks, what will happen with the human birth rate?)



Judea Pearl broke with the taboo of causal reasoning





ACM Turing Award 2011: “For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning."

Pearl's ladder of causality

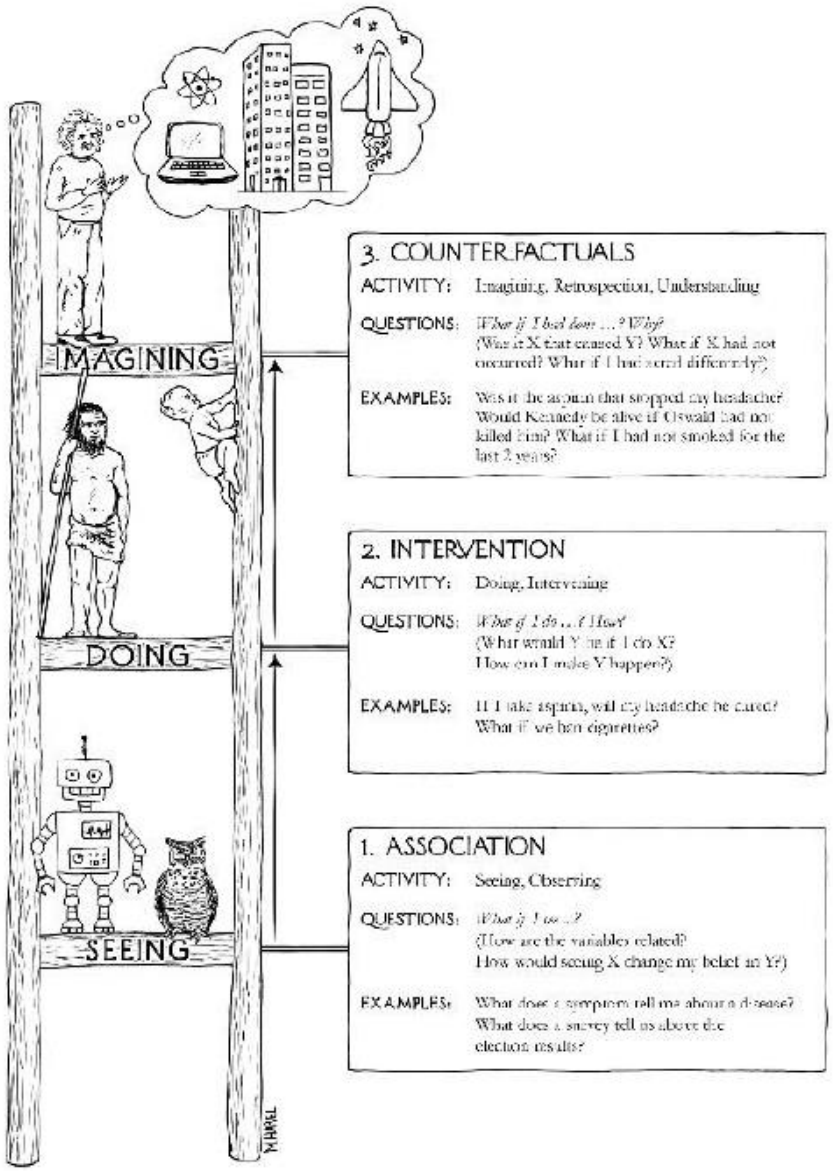


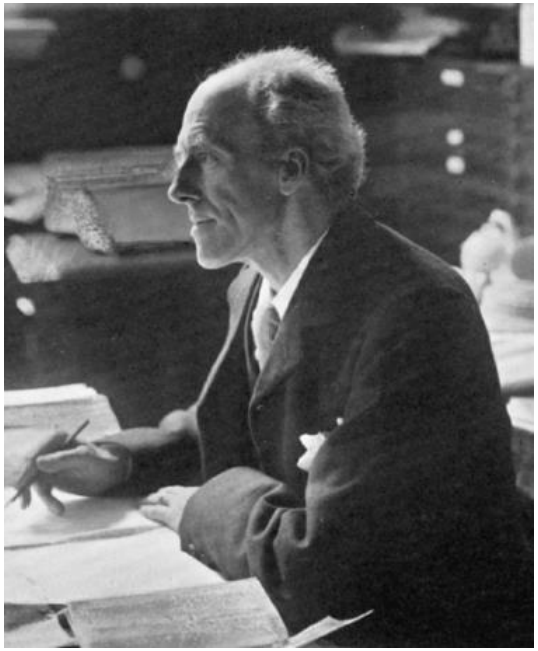
FIGURE 1.2. The Ladder of Causation, with representative organisms at each level. Most animals, as well as present-day learning machines, are on the first

Image credits: "The Book of Why"

With the correlation statistics was born and abandoned causality as "unscientific"

"the ultimate scientific statement of description of the relation between two things can always be thrown back upon… a contingency table [or correlation]."

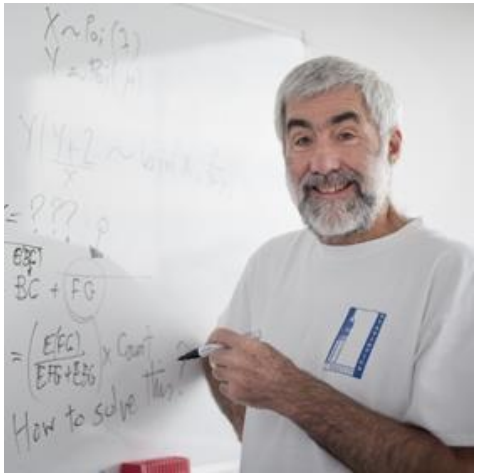
Karl Pearson (1895-1936), The Grammar of Science



Pearl's rephrasing of Pearson's statment:
"**data is all there is to science**".

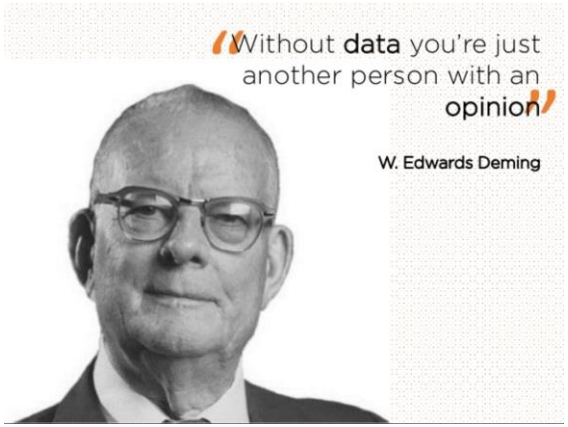
However, Pearson himself wrote several papers about "spurious correlation" vs "organic correlation" (meaning organic=causal?) and started the culture of "think: 'caused by', but say: 'associated with' "…

Quotes of data scientists



“Considerations of causality should be treated as they have always been in statistics: preferably not at all."

Terry Speed, president of the Biometric Society 1994



In God we trust. All others must bring data.

W. Edwards Deming (1900-1993), statistician and father of the total quality management

See also http://bigdata-madesimple.com/30-tweetable-quotes-data-science/

Pearl's statements

Observing [and statistics and AI] entails detection of regularities

We developed [AI] tools that enabled machines to reason with uncertainty [Bayesian networks].. then I left the field of AI

Mathematics has not developed the asymmetric language required to capture our understanding that if *X* causes *Y* .

As much as I look into what's being done with deep learning, I see they're all stuck there on the level of associations. Curve fitting.

The book of Why
https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515/

Regression Model
What can they tell us?

CPD (conditional probability density) encoded by linear regression

CPD: $Y_{X_i} = (Y|X_i) \sim N(\mu_{x_i}, \sigma^2)$

$Y_x \in \mathbb{R} \quad , \quad \mu_x \in \mathbb{R}$

$$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \varepsilon_i$$

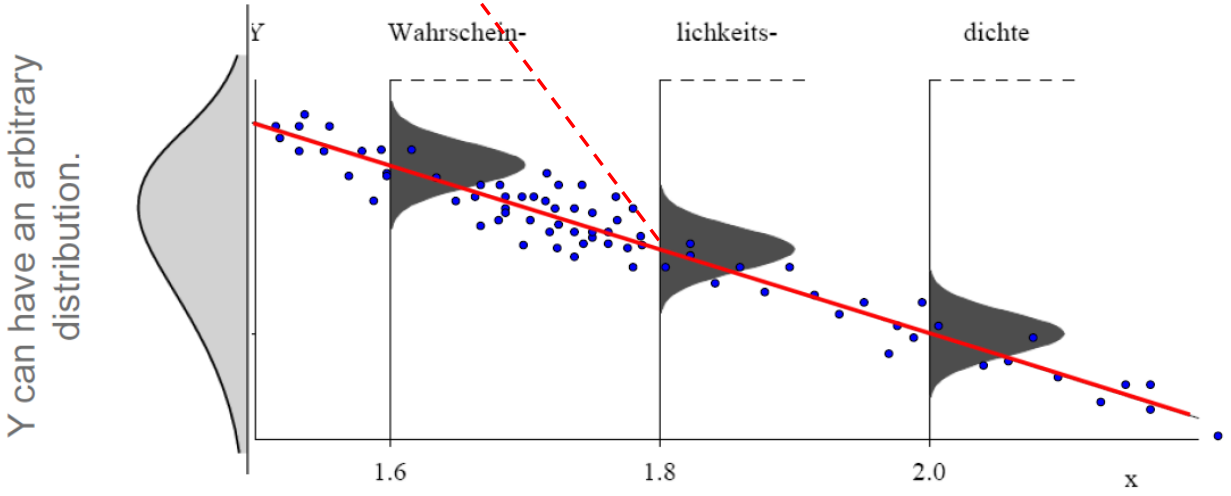
$$E\left(Y_{X_i}\right) = \mu_{x_i} = (\mu|X=x_i) = \beta_0 + \beta_1 \cdot x_{i1}$$

$$Var(Y_{X_i}) = Var(Y|X_i) = Var(\varepsilon_i) = \sigma^2$$

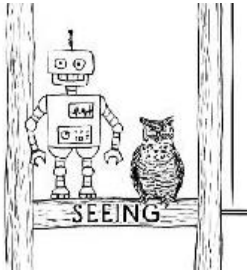
$$\varepsilon_i \sim N(0, \sigma^2)$$

$Y \sim V^{continuous}_{arbirary}$

$(Y|X_i) \sim N(\mu_{x_i}, \sigma^2)$



On the first rung of the ladder
Pure regression can only model associations



$$(\mathrm{Y}_i|\mathbf{X}_i) \sim \mathrm{N}(\mathbf{X}^t{}_i\beta = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{p-1} x_{ip-1},\ \sigma^2)$$

Usual interpretation:
The coefficient β_k gives the change of the expected outcome μ, given the explanatory variable x_k is increased by one unit and all other variables are held constant.

But: How can we increase just one predictor and hold the others constant?

Can a regression model help for intervention planning based on observational data?

Prof. P. Bühlman (ETH): "Pure regression is intrinsically the wrong tool"*

- Interpretation of the coefficients are:
  - purely on association (not on causation)
  - require that all other co-variables do not change when changing the co-variable of interest which is not realistic in many intervention scenarios.

Regression – the "statistical workhorse": the wrong approach

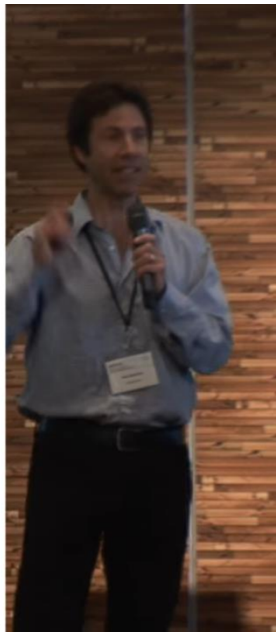
we could use linear model (fitted from n observational data)

$$Y = \sum_{j=1}^{p} \beta_j X_j + \varepsilon,$$
$$\mathrm{Var}(X_j) \equiv 1 \text{ for all } j$$

$|\beta_j|$ measures the effect of variable X_j in terms of "association"

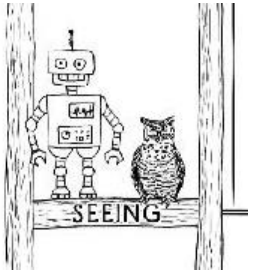
i.e. change of Y as a function of X_j when keeping all other variables X_k fixed

$\rightsquigarrow$ not very realistic for intervention problem
if we change e.g. one gene, some others will also change and these others are not (cannot be) kept fixed



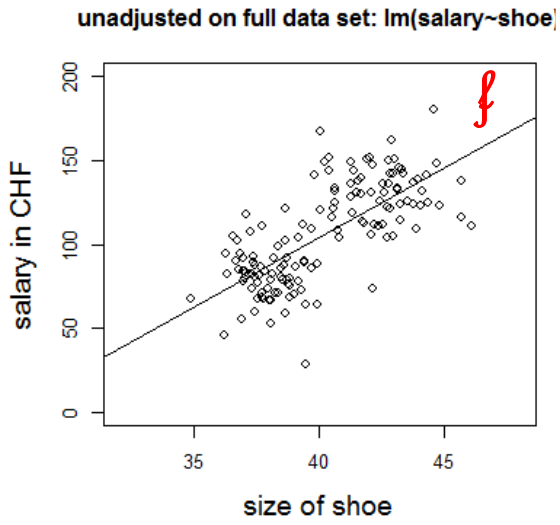
* https://www.youtube.com/watch?v=JBtxRUdmvx4

How we work with rung-1 regression or ML models



xkcd.com

Confounder can introduce spurious association: Adjustment methods can work well (toy example)



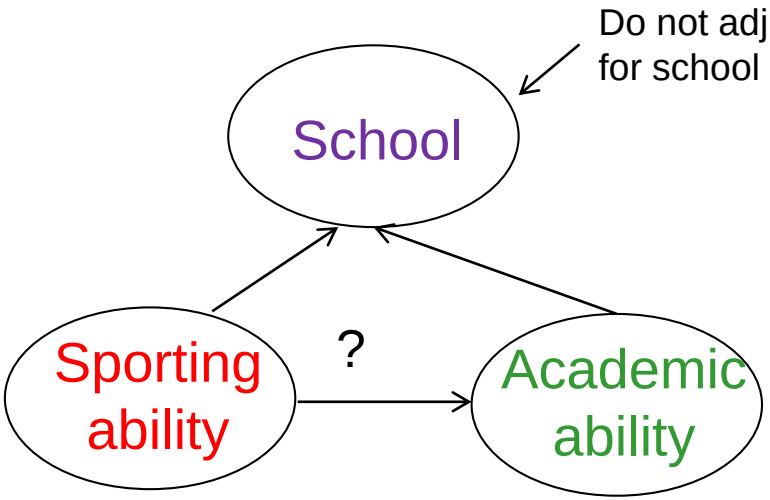
Looking into adjustment methods
Never adjust for a common effect: a toy example

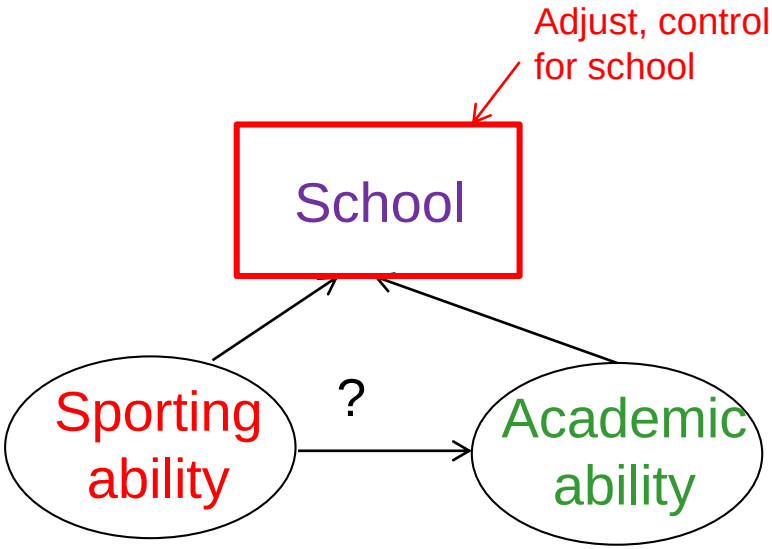
A school accepts pupils who are either good at sport, or good academically, or both

-> School acceptance is associated with sporting and academic abilities

Suppose: in Population sport and academic skills are independent

What happens if we “adjust” for the factor “accepted in school”?

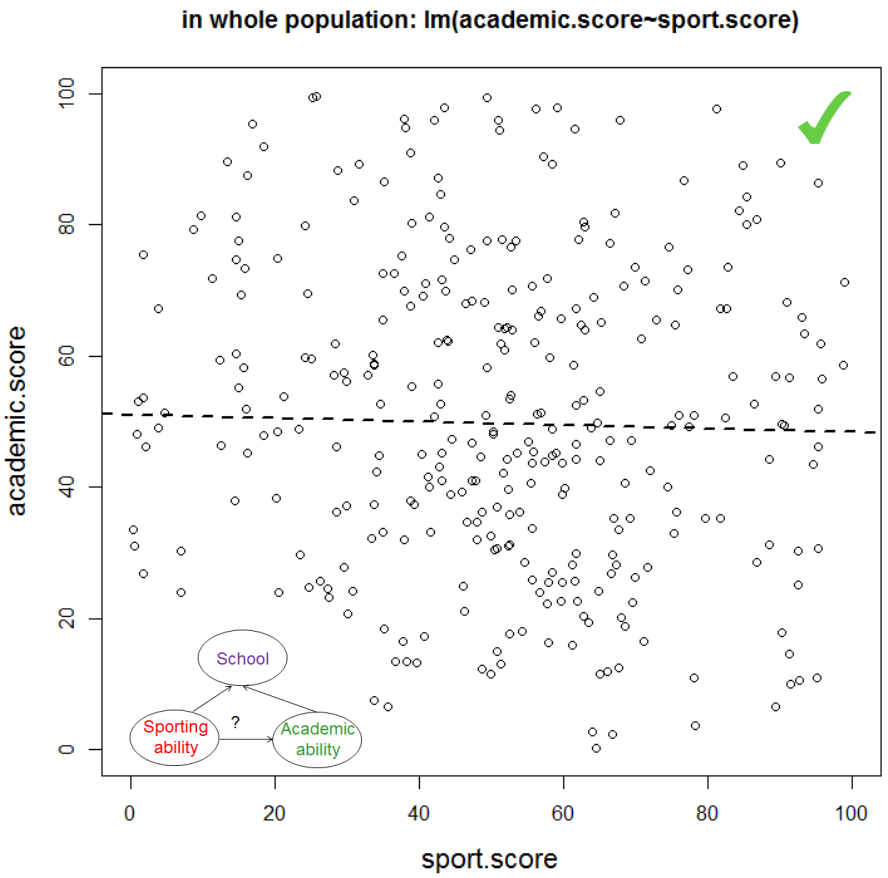


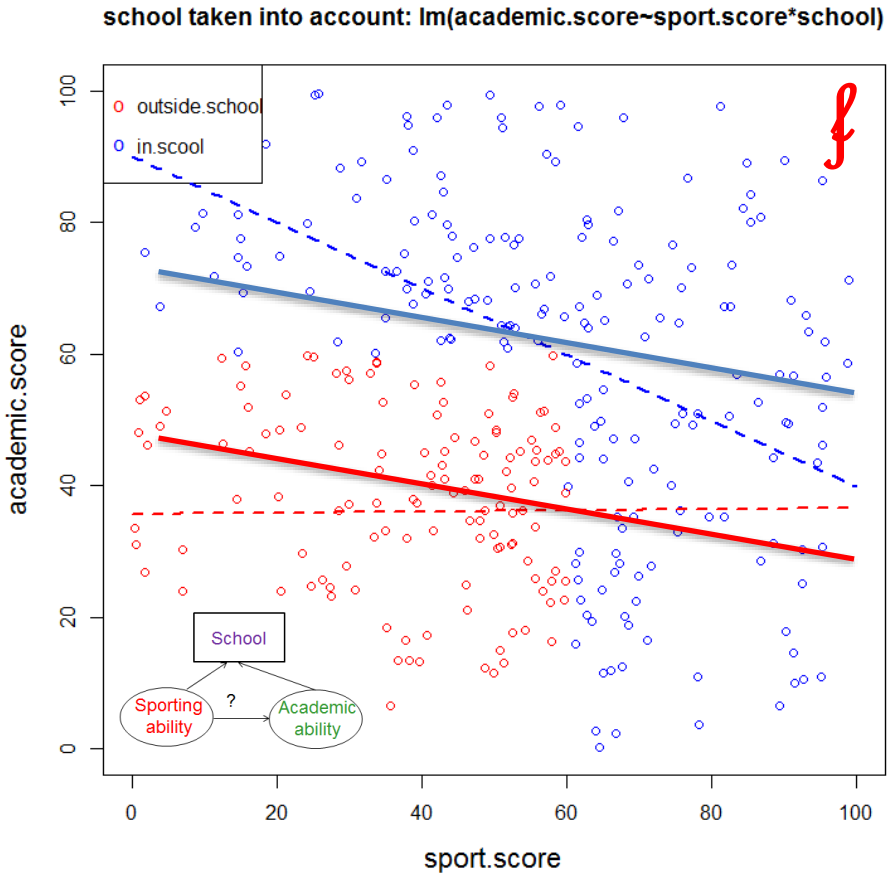



```
m1=lm(academic ~ sport, data=dat)
```

```
m2=lm(academic ~ sport + school, data=dat)
```

Adjusting for associated variables can work out badly
A toy example: effect of sport on academic abilities






```
m1=lm(academic ~ sport, data=dat)
```

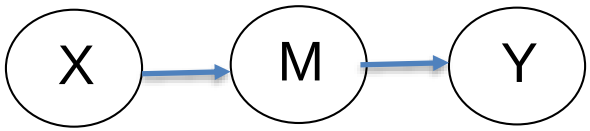
```
m2=lm(academic ~ sport + school, data=dat)
```

In the population there is no association between sport score and academic score, but by controlling for the school-variable we created a spurious association.

Looking into adjustment methods
Never adjust for mediator

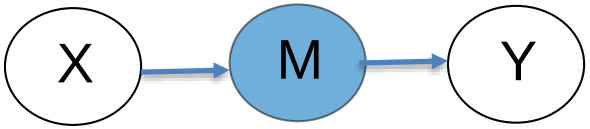
Toy example: a treatment X makes an enzyme M working which reduces pain Y

Not adjusting for M




```
Y ~ X
```

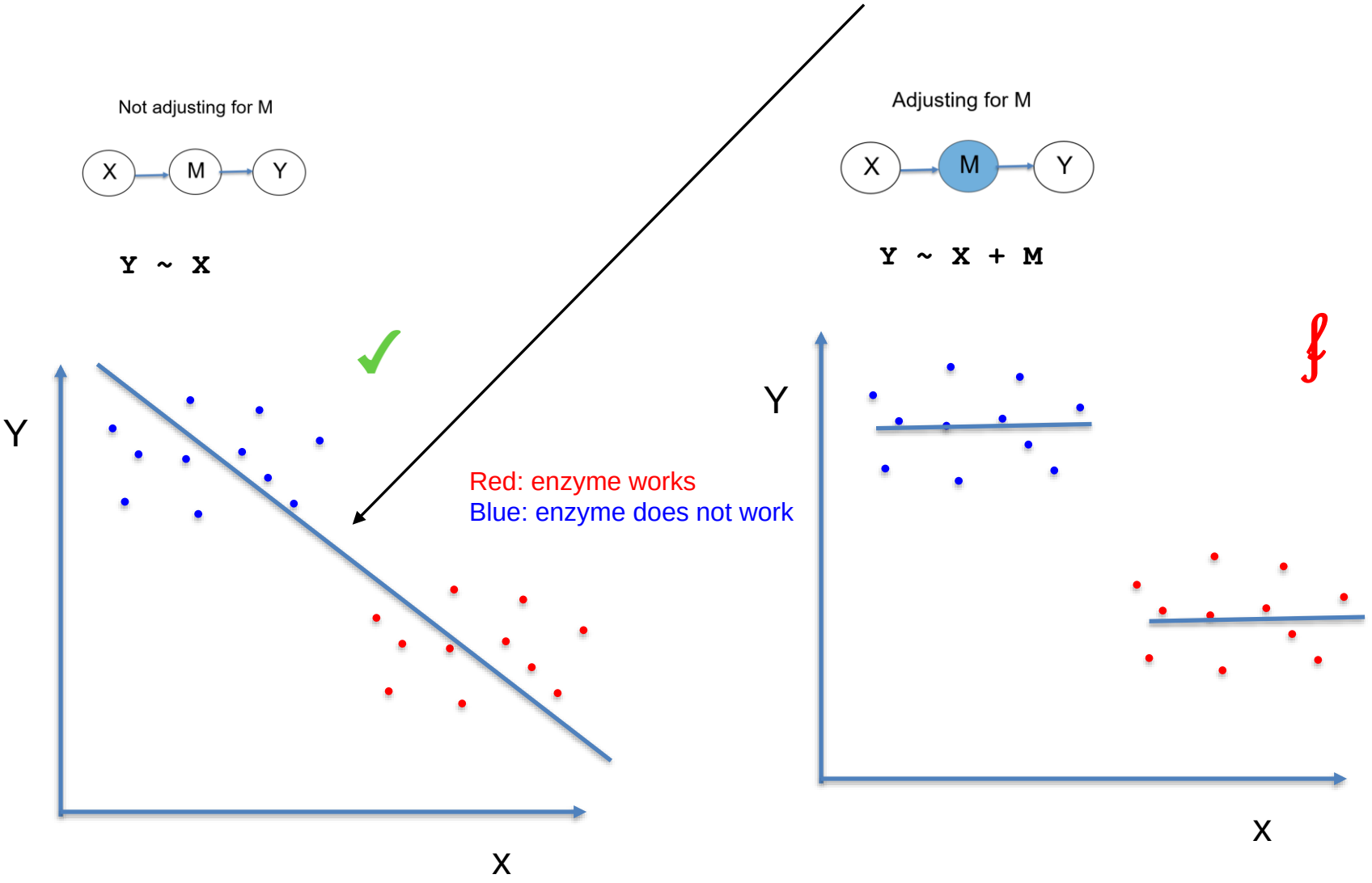
Adjusting for M




```
Y ~ X + M
```

Do not adjust for a mediator

Truth: because of treatment the enzyme starts working and pain Y is reduced!



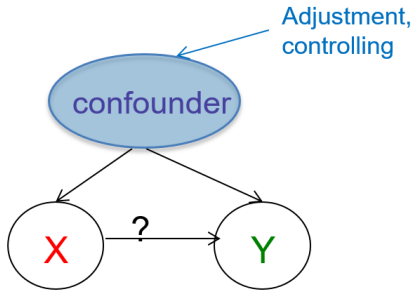
A third variable is associated with X and Y
To adjust or not to adjust – that is the question

Do not adjusting for a mediator!

```
y ~ x
```

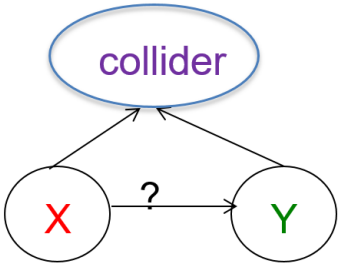
Adjust for a confounder!

```
y ~ x + c
```



Do not adjust for a collider!

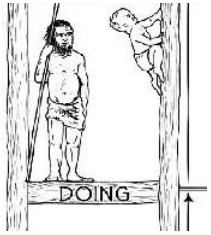
```
y ~ x
```



Can and should we try to learn about causal relationships?

If yes – what and how can we learn?

Ascending the second rung: go from "seeing" to "doing"

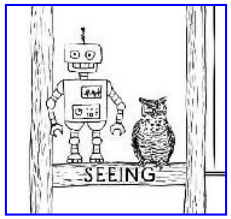


Research question:

What is the distribution of the blood pressure if people do not drink coffee?

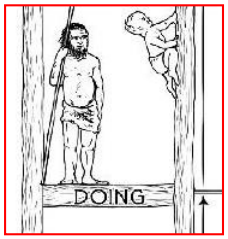
Conditioning / Seeing:
Filter - restrict on non-coffee drinker

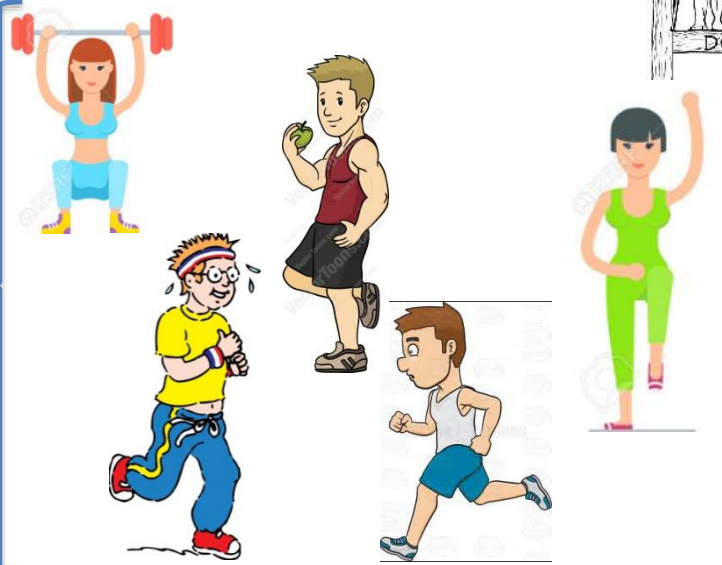
$$P(\mathrm{BP} \mid \mathrm{coffee} = 0)$$



"Do"-Operator:
Full population, after intervention that prohibits coffee consume

$$P(\mathrm{BP} \mid \mathrm{do}(\mathrm{coffee} = 0))$$

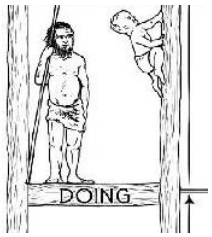


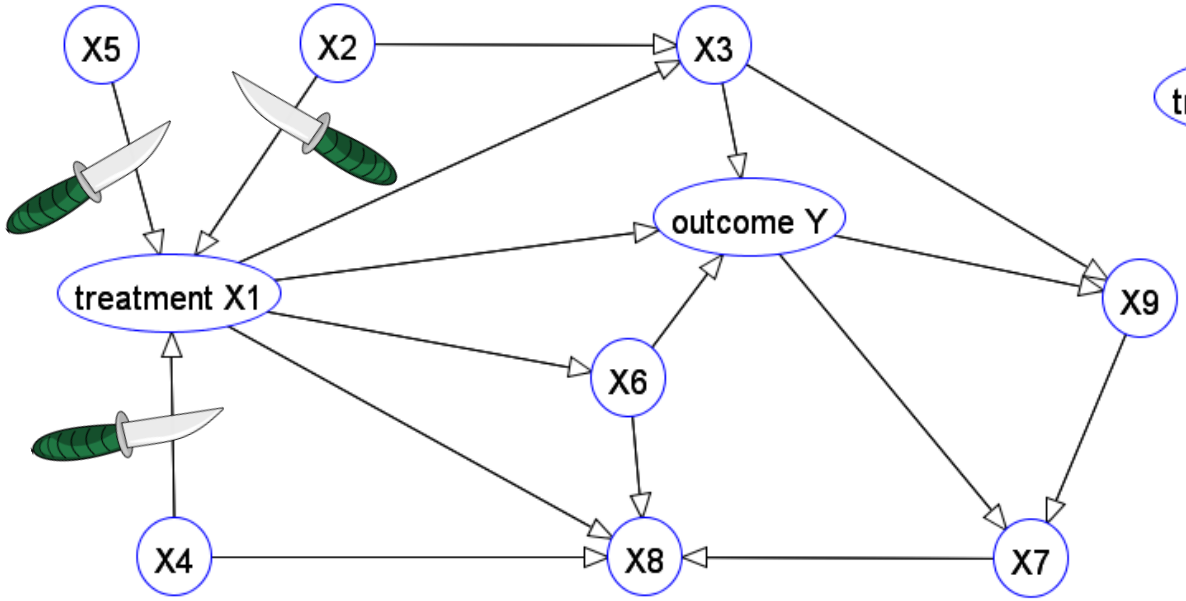


non-coffee drinker by choice

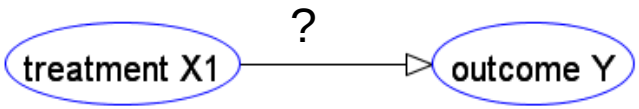
coffee drinker by choice

On the second "doing" rung of the ladder
Assessing the intervention effect by a RCT





RCT through the lens of a causal graphical model



Since the treatment is assigned randomly to both treatment groups are exchangeable. Hence observed differences of the outcome in both groups is due to the treatment.

-> Model after collecting data from a RT: $outcome \sim treatment$

Hands-on time

Which of these concepts are used in causal inference or association modeling?

Conditioning

Confounding

Intervention

Randomization

Regression

Independence

Conditioning

Spurious association

Exchangeability

Odds ratio

Interaction

adjusting

Propensity score

Experimentally fixing / holding constant

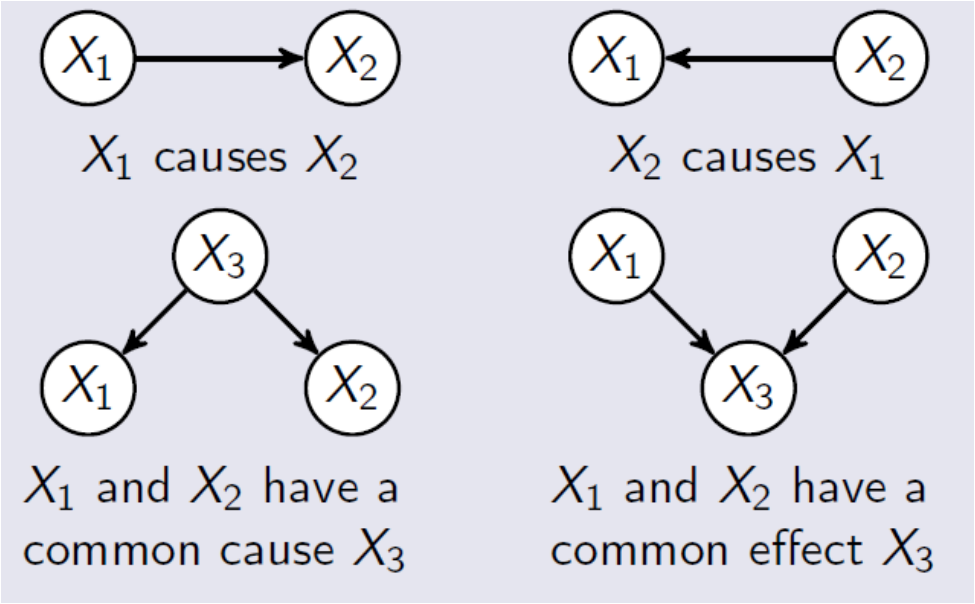
Causality

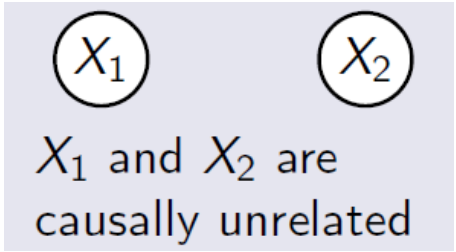
Association

Association	Causality
Conditioning	Confounding
Regression	Intervention
Independence	Randomization
Odds ratio	Spurious association
Interaction	Exchangeability
Adjusting	Experimentally fixing
Propensity score	

Causal graphical models

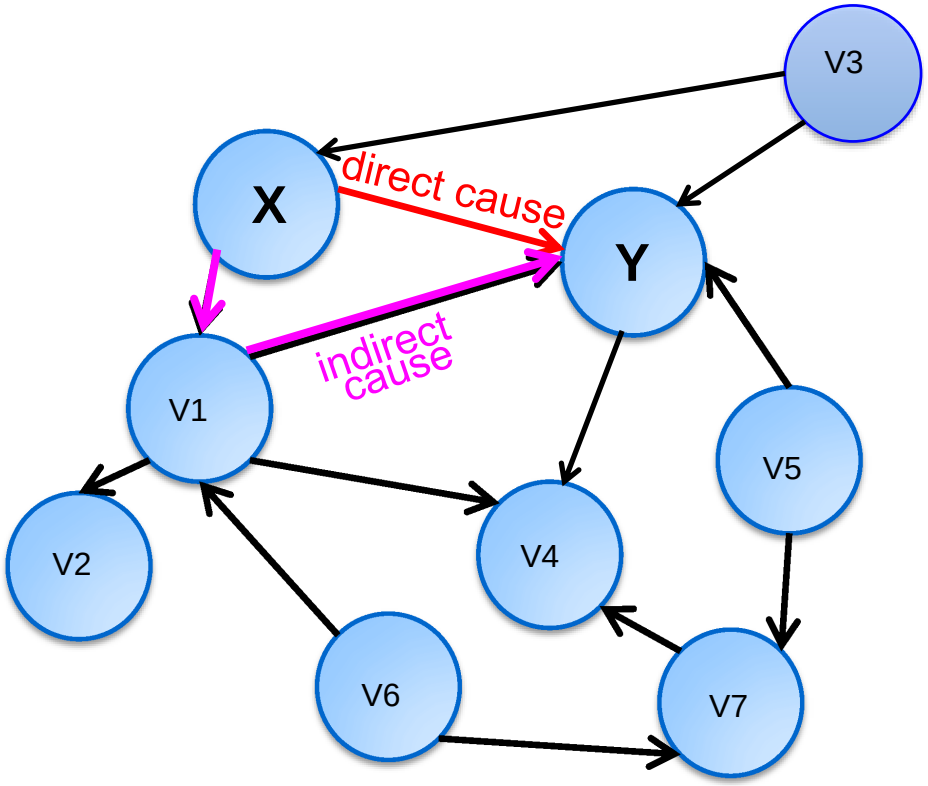
A causal graphical model represents a model about causal relationships where the nodes are variables and a directed edge represents a potential causal effect.





Causal effects can only be transported along the direction of arrows!

What is a causal path?

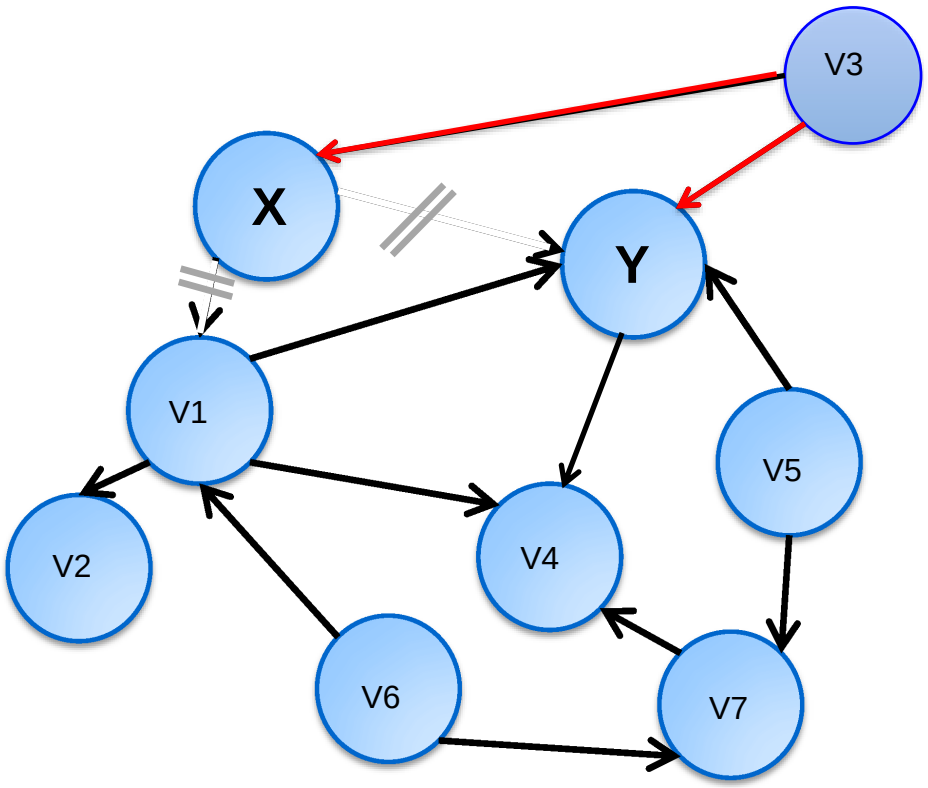


In a causal path from X to Y is a directed path from X to Y

→ if follow the arrows in a causal path we get from X to Y.

→ We have 2 causal paths transporting direct and indirect causes

What is a backdoor path from X to Y?

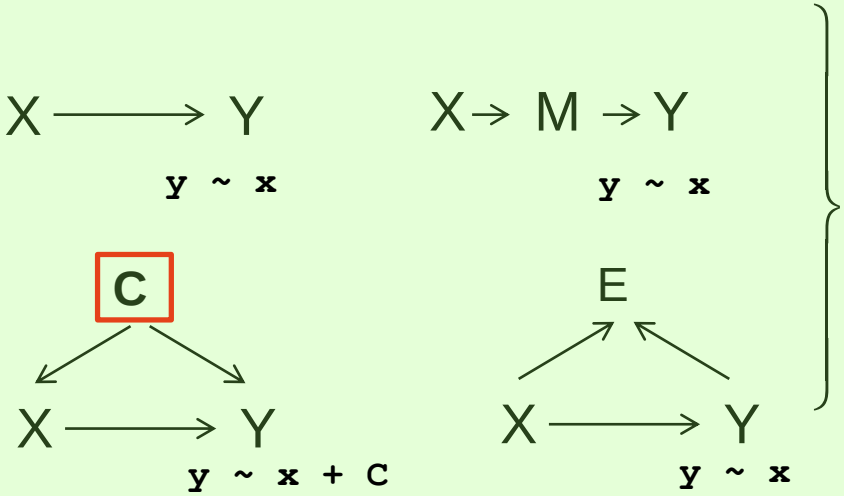


First we ignore (delete) all arrows starting from X (these are either causal paths or blocked by a collider)

A backdoor path from X to Y starts with an arrow pointing into X: $X \leftarrow \cdots Y$

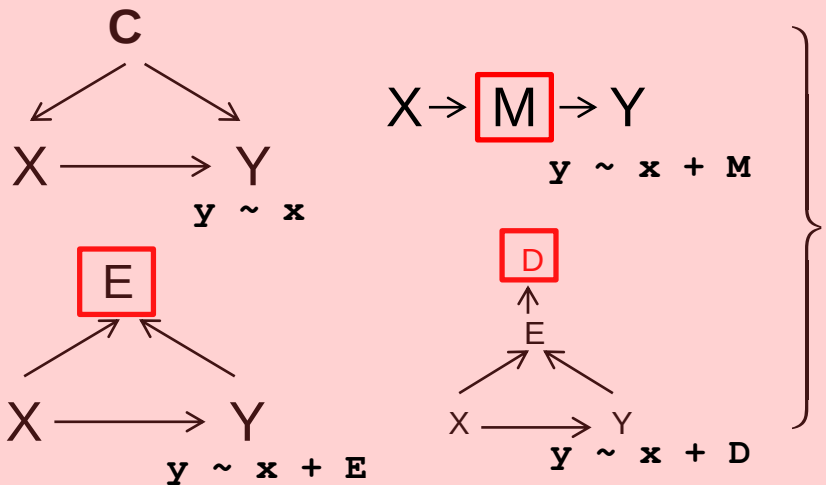
→ Any path (regardless of the arrow directions) that still connects X and Y.

Building blocks of causal model



⇒ inference from assocation between X and Y ≙ causal effects

adjusted variable



⇒ inference from association between X and Y on causation will be spurious

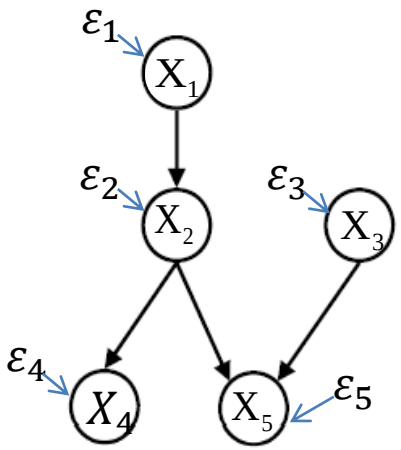
The data generating process defined by structural equations

We assume that each (endogenous) variable X_i in the model is generated as a function of its graphical parents and a (exogenous) noise variables ε_i, where the noise variables are independent. The collections of these data generation functions are called structural equation model (SEM):

$$X_i \leftarrow h_i\left(\text{pa}(X_i) , \varepsilon_i \right)$$

left-side variable is **generated** by right-side function

$$X_1 \leftarrow h_i\left(\varepsilon_1 \right)$$
$$X_3 \leftarrow h_i\left(\varepsilon_3 \right)$$
$$X_2 \leftarrow h(X_1)+\varepsilon_2$$
$$X_4 \leftarrow h_i\left(X_2 , \varepsilon_4 \right)$$
$$X_5 \leftarrow h_5\left(X_2, X_3, \varepsilon_5 \right)$$



Remark: We assume that we have measured all involved variables (no hidden or latent variables case)
h can be an arbitrary function, errors need to be independent but they have not to be additive.

Assuming a linear Wonderland
Linear structural equation models and the path method

An easy special case are linear SEM (structure equation model) with additive noise:

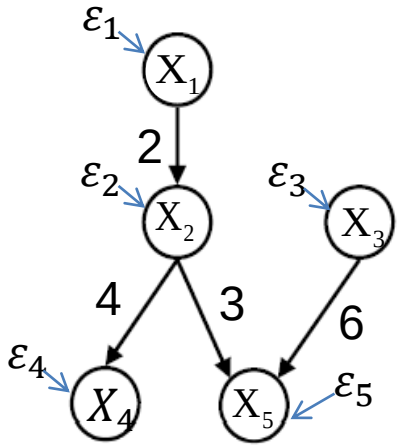
$$X_1 \leftarrow \varepsilon_1$$

$$X_3 \leftarrow \varepsilon_3$$

$$X_2 \leftarrow 2 \cdot X_1 + \varepsilon_2$$

$$X_4 \leftarrow 4 \cdot X_2 + \varepsilon_4$$

$$X_5 \leftarrow 3.X_2 + 6 \cdot X_3 + \varepsilon_5$$

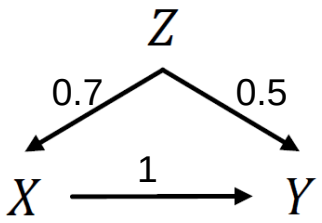


Linear Gaussian SEM: Example 1 in numbers

The linear SEM assumes independent Standard-Gaussian noise $\epsilon_Z, \epsilon_X, \epsilon_Y$.

$$Z = \varepsilon_Z$$
$$X = 0.7 * Z + \varepsilon_X$$
$$Y = 1 * X + 0.5 * Z + \varepsilon_Y$$

DAG



Data Simulation:

```
set.seed(123)
n <- 1000
z <- rnorm(n)
x <- 0.7*z + rnorm(n)
y <- 1*x + 0.5*z + rnorm(n)
```

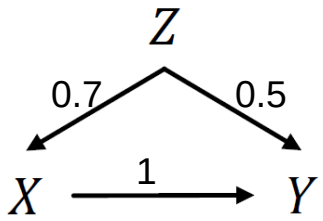
SEM tells: true causal effect of X on Y: 1
If we increase X by one unit, Y will also increase by one unit

Can we estimate the true causal effect with a linear regression ?

Example credits: M. Kalish, ETH

Estimating causal effects by adjusted regression

Goal: Find causal effect of treatment X on Y.
(we know from the SEM that the true causal effect is 1)



Adjustment set via backdoor criterion:

Adjustment set is : $S = \{Z\}$

Simple Regression: $lm(Y \sim X)$

```
> confint(lm(y~x))
                2.5 %     97.5 %
(Intercept) -0.09005941 0.03942158
x            1.19606286 1.29767266
```

✗

Multiple Regression: $lm(Y \sim X + Z)$

```
> confint(lm(y~x+z))
                2.5 %     97.5 %
(Intercept) -0.08172964 0.03986164
x            0.96709528 1.08791825
z            0.38165727 0.53685033
```

✓

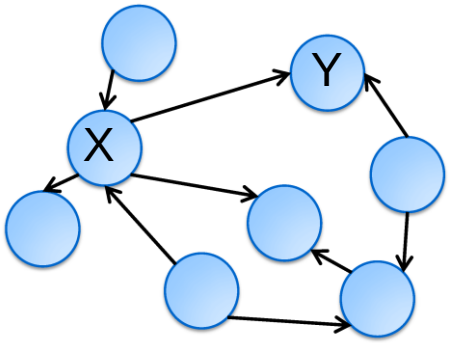
Example credits: M. Kalish, ETH

Can we do causal/intervential inference from observational data?

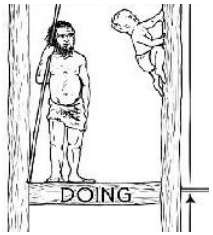
The very short answer: No!

Principle be Cartwright (1989): **No causes in – no causes out!**

observational data



Backdoor criterion
or frontdoor criterion
or 3 Rules of do-Calculus



$$P'(y \mid do(X = x_0))$$

=

Expression without do (!!)
which only uses information
from observed JPD P

Pearl's backdoor criterion to identify covariate sets

Backdoor paths can transport non-causal association and must be closed to avoid spurious association. The backdoor criterion allows to identify an appropriate set of covariates for which we should adjust to estimate the causal effect of X on Y:

- Remove all arrows starting from X
  a open paths (without a collider on the way) starting from X transports causes from X on Y
- Identify all open (active) back-door paths
  open backdoor paths may introduce association which is not due to the causal effect of X on Y
- Determine whether a set S of covariates is sufficient to block all backdoor paths
  a open backdoor paths can be blocked by a collider on the path, when you do not adjust for the collider, or by a confounder on the path, when you adjust for the confounder.

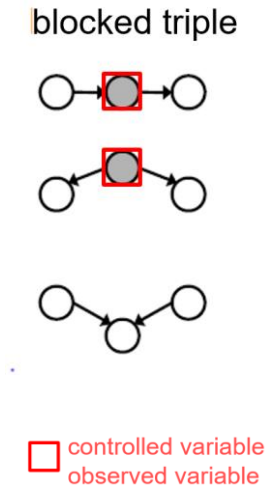
Pearl's backdoor criterion for causal graphical models

Goal: Close all backdoor paths connecting X and Y.

- Determine a set S of "de-confounder" variable closing all backdoor paths by controlling for these variables.
- S must not contain any descendent of X.
  (This ensures that we do not block a causal path from X to Y)

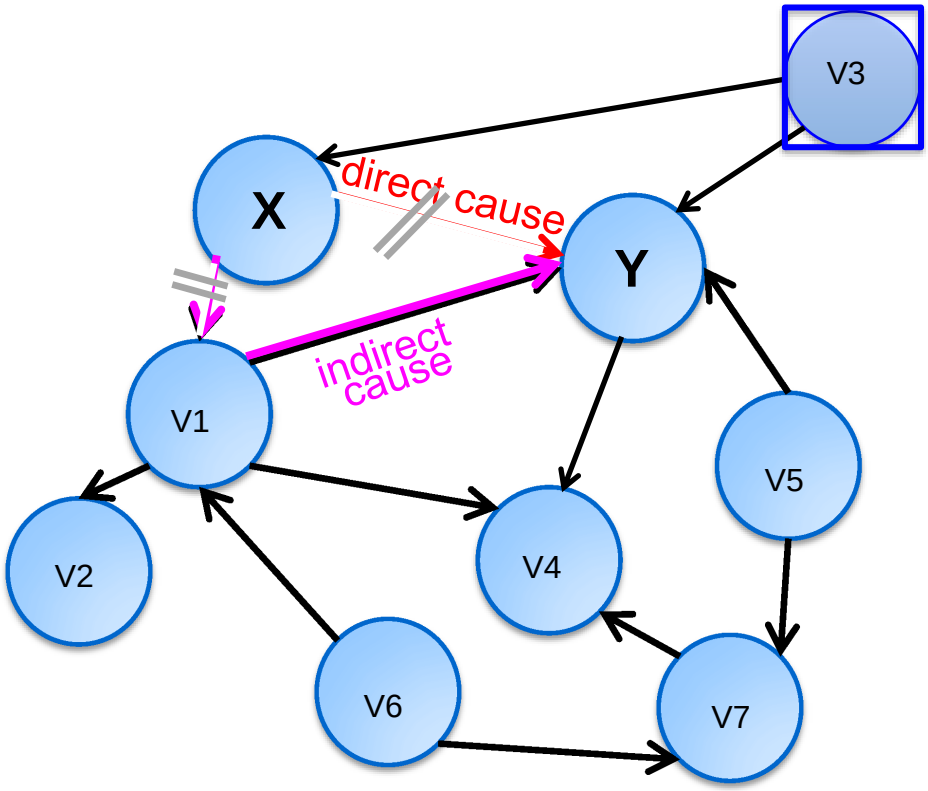
- S can be used for covariate adjustment to estimate the total causal effect of X on Y

A path is blocked if **1 single** triple-segment is blocked!



Control for a variable
=
Adjust for a variable
=
using the variable in the regression model

Has X an causal influence on Y?
Are all backdoor paths closed?



To close all backdoor paths we must adjust for this confounder.

```
y ~ x + v3
```

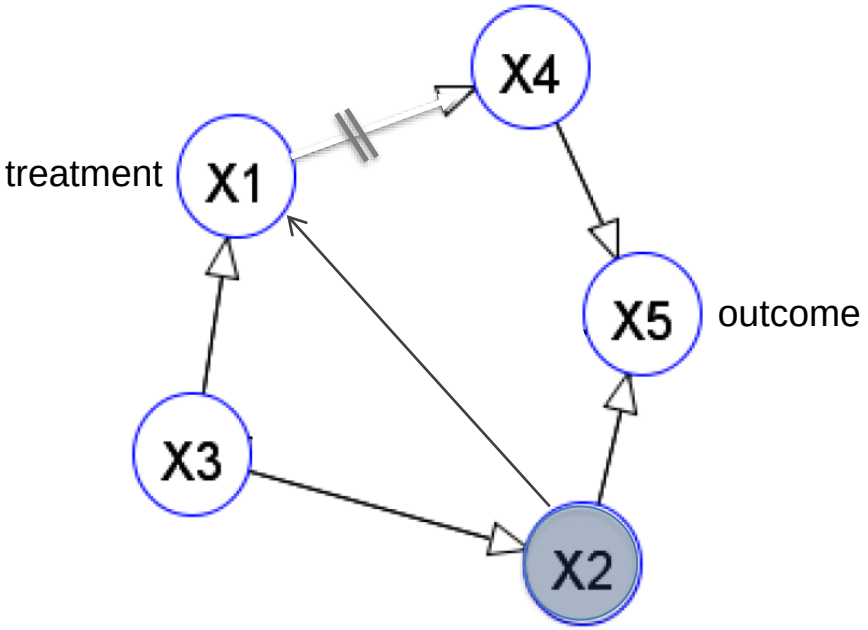
This model appropriate to
Interpret the estimated β_1 causally

Use the back door criterion to check a model

RQ: Has X_1 ("treatment") a causal effect on X_5 ("outcome")?

Is the proposed model appropriate to Interpret the estimated β_1 causally?

$$X_5 \sim X_1 + X_2$$



Are all back door paths (BDP) closed?

Yes, since all BDP go through the confounder X_2 and we control for X_2 by using it as covariable and thereby closing the BDP.

→ The estimated β_1 can be interpreted causally, given the graphical model is correct.

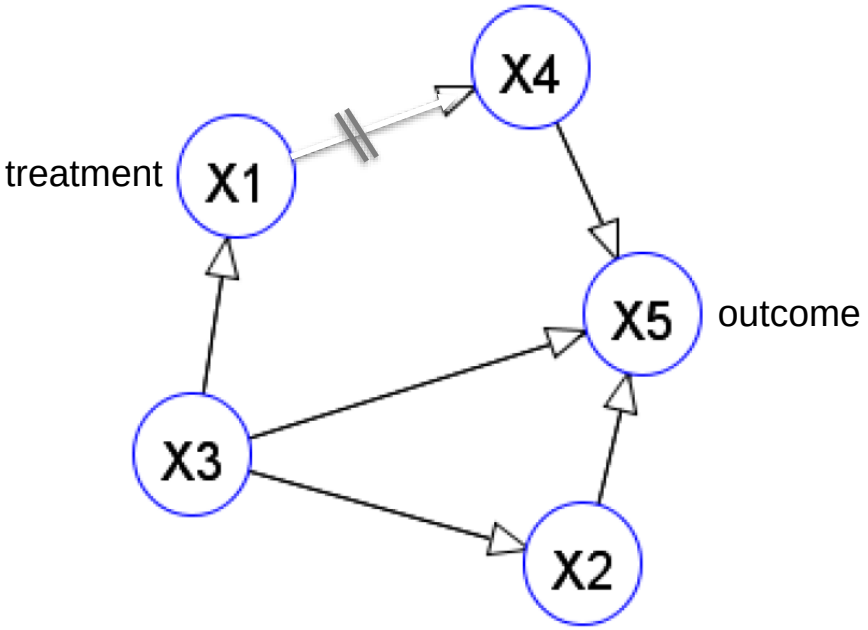
DIYS time ☺

Use the back door criterion to check a model

RQ: Has X_1 ("treatment") a causal effect on X_5 ("outcome")?

Is the proposed model appropriate to Interpret the estimated β_1 causally?

$$X_5 \sim X_1$$



Are all back door paths (BDP) closed?

No, since the BDP X1-X3-X5 goes through an uncontrolled confounder X_3 and is therefor open.

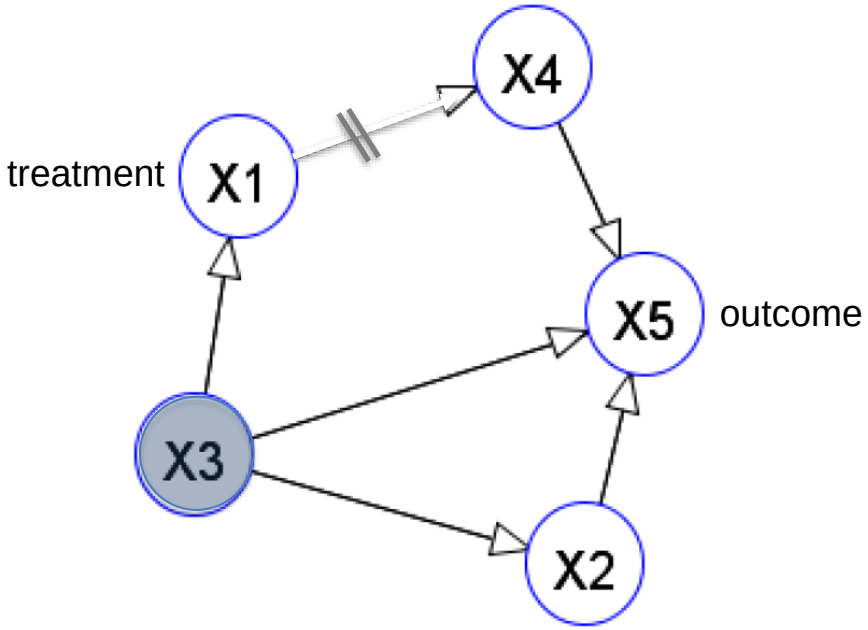
→ The estimated β_1 must not be interpreted causally, given the graphical model is correct.

Use the back door criterion to check a model

RQ: Has X_1 ("treatment") a causal effect on X_5 ("outcome")?

Is the proposed model appropriate to Interpret the estimated β_1 causally?

$$X_5 \sim X_1 + X_3$$ ✓



Are all back door paths (BDP) closed?

Yes, since all BDP go through the confounder X_3 and we control for X_3 by using it as covariable and thereby closing the BDP.

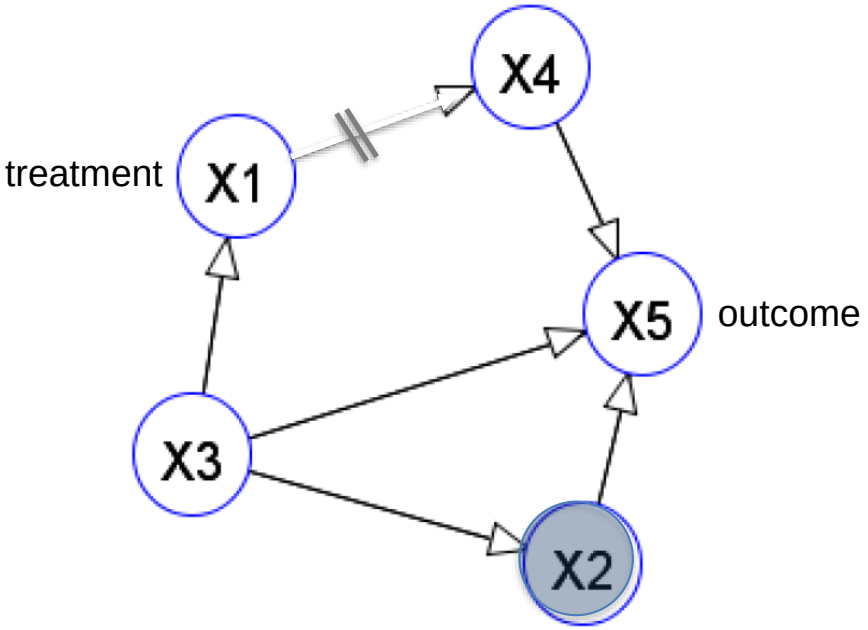
→ The estimated β_1 can be interpreted causally, given the graphical model is correct.

Use the back door criterion to check a model

RQ: Has X_1 ("treatment") a causal effect on X_5 ("outcome")?

Is the proposed model appropriate to Interpret the estimated β_1 causally?

$$X_5 \sim X_1 + X_2$$



Are all back door paths (BDP) closed?

No, since the BDP X_1-X_3-X_5 goes through an uncontrolled confounder and is therefor open.

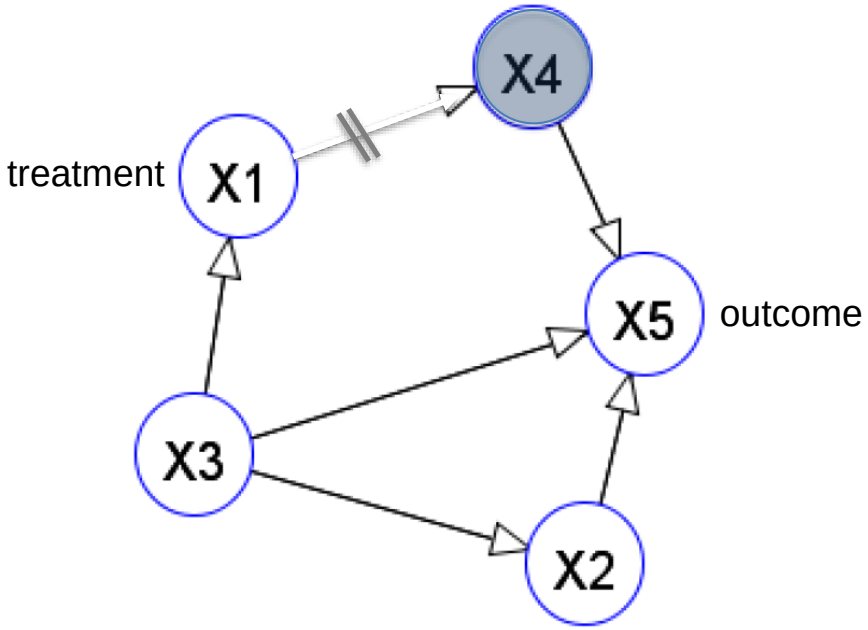
→ The estimated β_1 must not be interpreted causally, given the graphical model is correct.

Use the back door criterion to check a model

RQ: Has X_1 ("treatment") a causal effect on X_5 ("outcome")?

Is the proposed model appropriate to
Interpret the estimated β_1 causally?

$$X_5 \sim X_1 + X_4$$

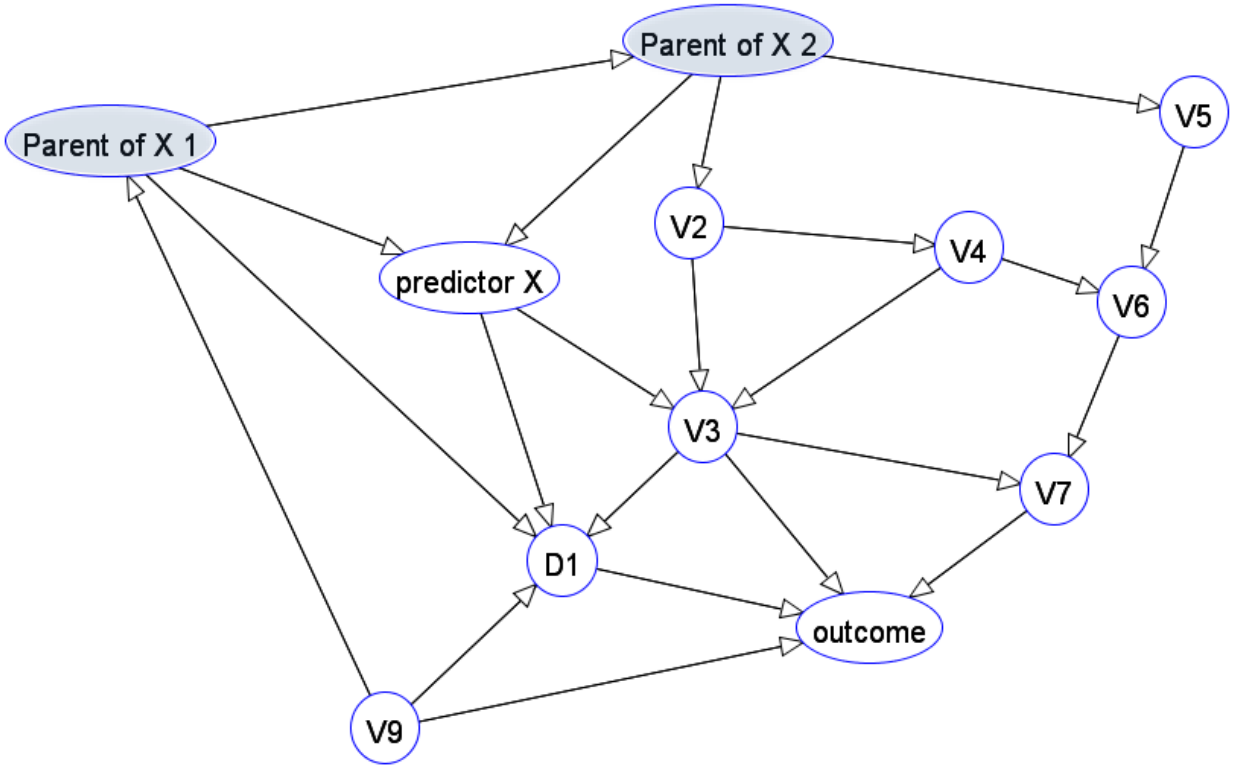


Are all back door paths (BDP) closed?

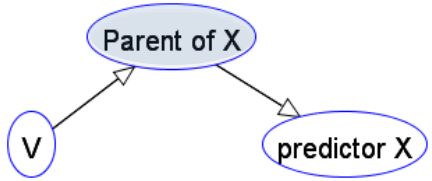
X_4 is a descendent of X_1
(mediator on causal path)
You must not use X_4 as covariable!!!

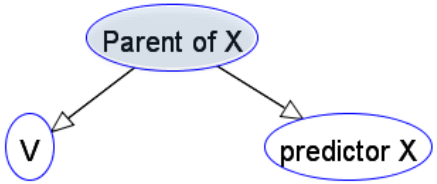
Special case of the backdoor criterion: intervention parents

All backdoor paths are closed if we control for the parents of the intervention variable X!



A controlled parent blocks the backdoor path either as controlled mediator or controlled confounder.

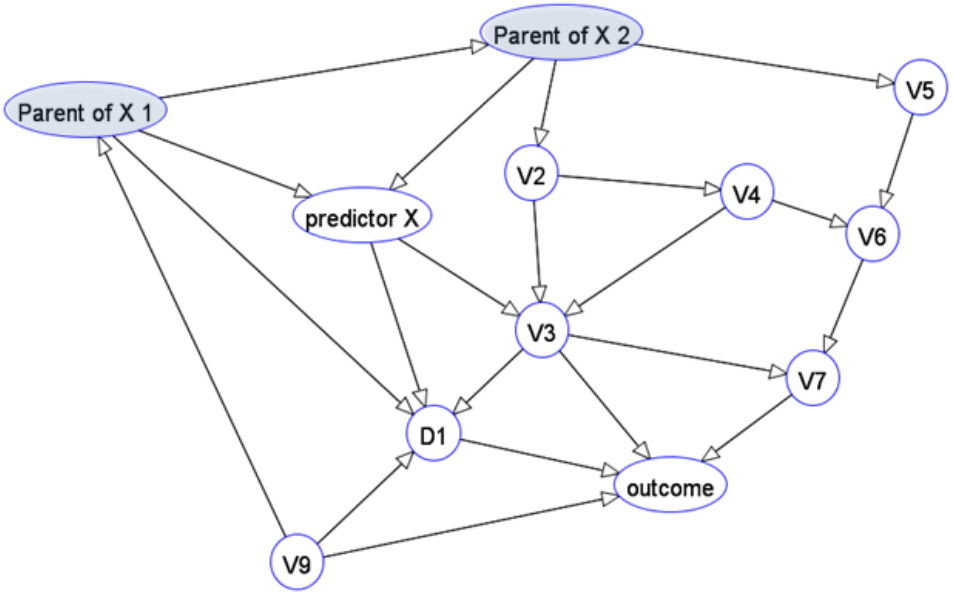




$$\text{outcome} \sim \text{predictor} + \sum \text{parents(predictor)}$$

Causal effects can be derived from observational data when using Pearl's backdoor criterion is fulfilled

When can the regression coefficient of the predictor of interest in a regression model be interpreted as causal effect?



We need to adjust with an appropriate **set S_B** of covariates V_i (e.g. all parents of X) which would be sufficient to **close all backdoor paths** from intervention X to the outcome Y

$$\text{outcome} \sim \text{predictor} + \sum_{V_i \in S_B} V_i \qquad \text{outcome} \sim \text{predictor} + \sum \text{parents(predictor)}$$

Are smoking mothers for underweighted newborns beneficial?

Since 1960 data on newborns showed consistently that low-birth-weight babies of smoking mothers had a better survival rate than those of nonsmokers.

This paradox was discussed for 40 years!

An article by Tyler VanderWeele in the 2014 issue of the *International Journal of Epidemiology* nails the explanation perfectly and contains a causal diagram:

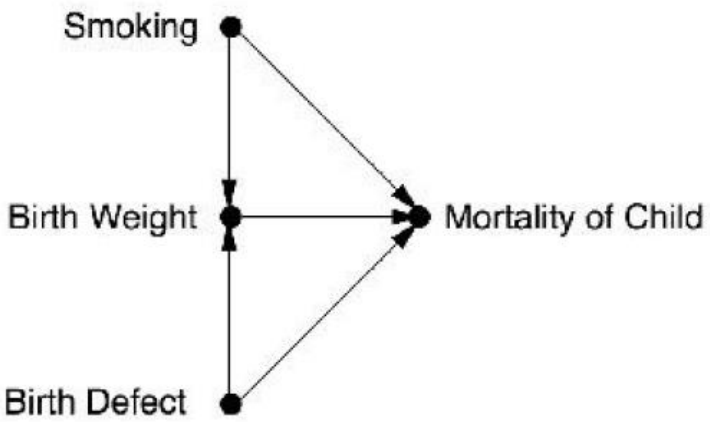


Image credits: “The Book of Why”

Outlook: How to do causal inference in R?

More Causal Inference with Graphical Models in R Package pcalg

Markus Kalisch
ETH Zurich

Martin Mächler
ETH Zurich

Diego Colombo
ETH Zurich

Alain Hauser
University of Bern

Marloes H. Maathuis
ETH Zurich

Peter Bühlmann
ETH Zurich

Abstract

The **pcalg** package for R (R Development Core Team 2014) can be used for the following two purposes: Causal structure learning and estimation of causal effects from observational and/or interventional data. In this document, we give a brief overview of the methodology, and demonstrate the package's functionality in both toy examples and applications.

This vignette is an updated and extended (FCI, RFCI, etc) version of Kalisch *et al.* (2012) which was for **pcalg** 1.1-4.

Keywords: IDA, PC, RFCI, FCI, GES, GIES, do-calculus, causality, graphical model, R.

Summary

- **Every causal inference with observational data rests on assumptions**

- **Causal inference algorithms for observational data may be useful**
  - to guide the design of interventional studies

- **Causal graphs allow to**
  - visualize (assumed or fitted) underlying causal relationships
  - represent the joint probability distribution
  - investigate if a causal effect can be identified
  - identify sets of covariates for which one should adjust in regression models
  - formulate testable hypothesis which could be used to falsify the assumptions