Causality:

Estimating total effects with covariate adjustment

Marloes Maathuis, ETH Zurich

Risk Day 2019

Simpson's paradox

	Treatment	Placebo
Male	50/100	150/500
Female	50/500	0/100
Total	100/600	150/600

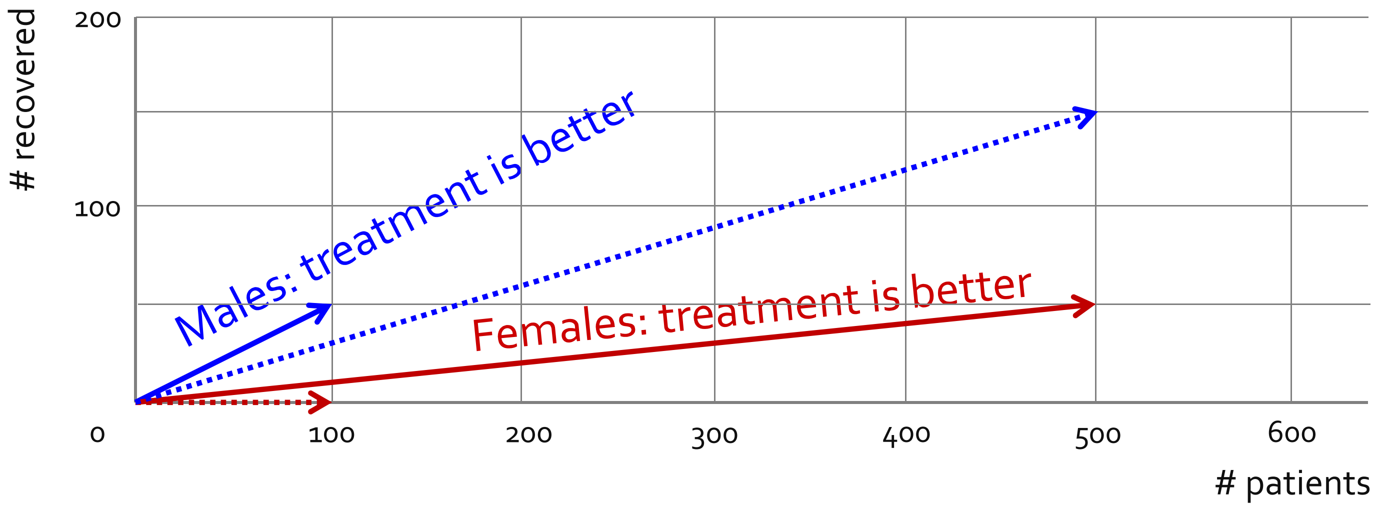
Hypothetical recovery rates, separated by gender

- Among males, treatment is better
- Among females, treatment is better
- Overall, placebo is better

References: Yule (1903), Biometrika; Simpson (1951), JRSS-B; Hernan, Clayton and Keiding (2011), Int J. Epidemiology; Pearl (2014), The American Statistician 68, 8-13.

Visual representation

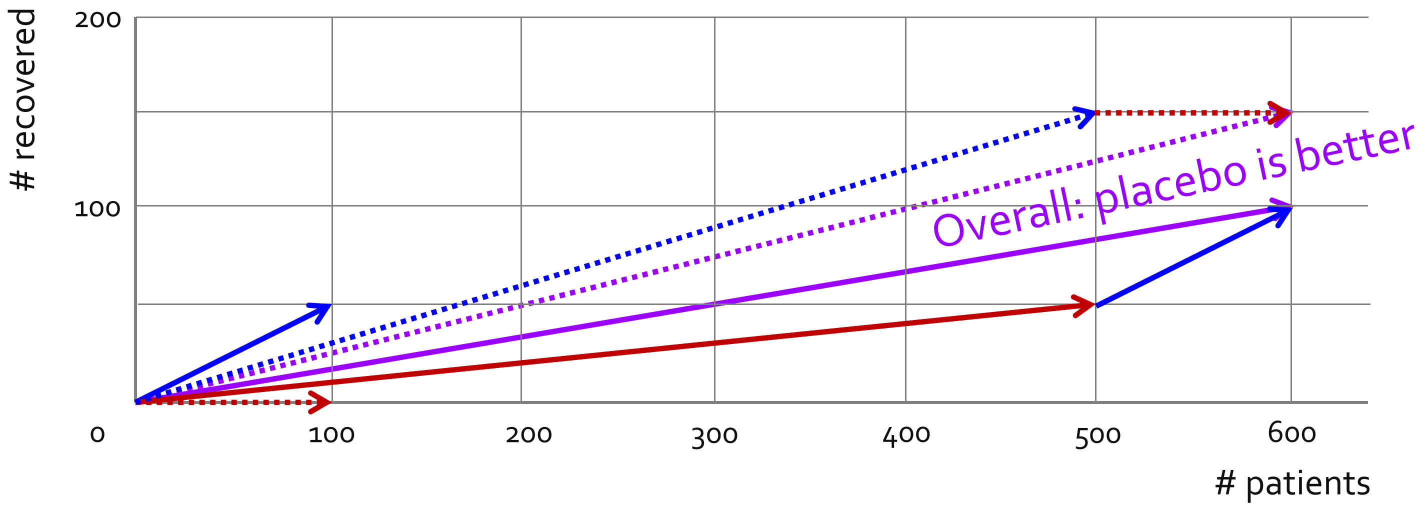
	Treatment	Placebo
Male	50/100	150/500
Female	50/500	0/100
Total	100/600	150/600
	(solid lines)	(dotted lines)



Vector representation: slope is proportion recovered

Visual representation

	Treatment	Placebo
Male	50/100	150/500
Female	50/500	0/100
Total	100/600	150/600
	(solid lines)	(dotted lines)



Vector representation: slope is proportion recovered

Simpson's paradox

	Treatment	Placebo
Male	50/100	150/500
Female	50/500	0/100
Total	100/600	150/600

Control for gender;
use the treatment

Simpson (1951), in an example similar to this one:
"*The treatment can hardly be rejected as valueless to the race when it is beneficial when applied to males and to females.*"

Replace gender by
blood pressure (BP)

	Treatment	Placebo
High BP	50/100	150/500
Low BP	50/500	0/100
Total	100/600	150/600

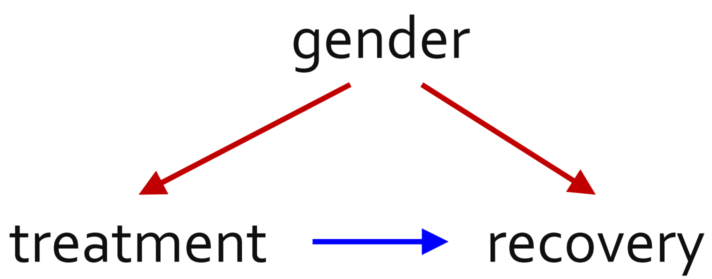
Don't control for BP;
don't use the treatment

Simpson (1951), in an example similar to this one:
"…, *yet it is the combined table which provides what we would call the sensible answer…*"

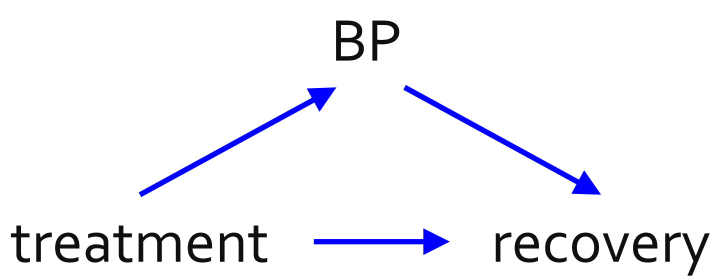
Simpson's paradox and causal diagrams

- Same numbers, different conclusions....
  - We must use additional information:
    story behind the data, causal assumptions

- We want to know the causal effect of treatment on recovery.
  Possible scenarios:



gender is a confounder;
control for gender



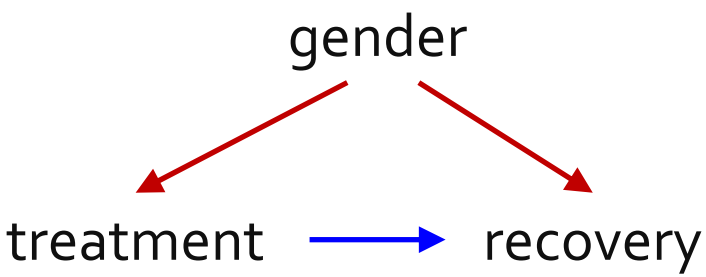
BP is an intermediate variable;
don't control for BP

Or......

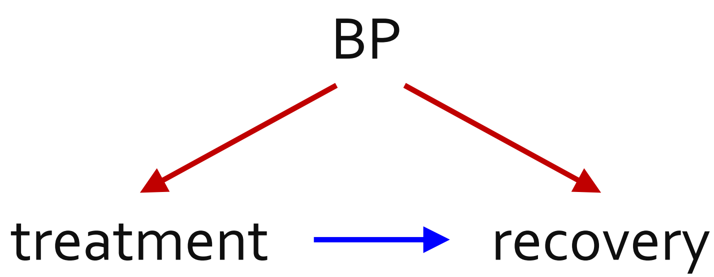
Simpson's paradox and causal diagrams

- Same numbers, different conclusions....
  - We must use additional information:
    story behind the data, causal assumptions

- We want to know the causal effect of treatment on recovery.
  Possible scenarios:

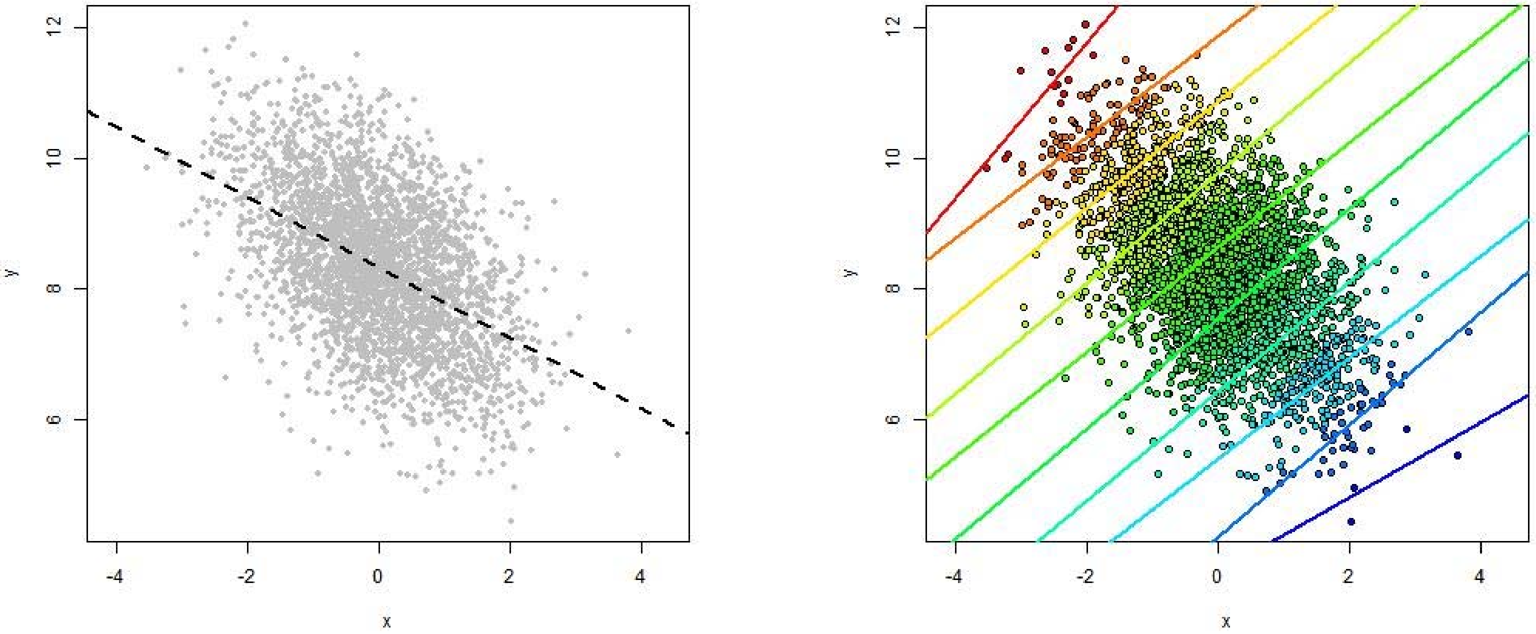


gender is a confounder;
control for gender



BP is an confounder;
control for BP

Simpson's paradox in regression



Color encodes third variable Z

Source: http://www.r-bloggers.com/fun-with-simpsons-paradox-simulating-confounders/

Regression

- Different variables in model can lead to different conclusions:
  - β_j reflects the partial association between X_j and Y when all other variables in the model are held constant
  - Simpson's paradox is an extreme case where the sign flips
  - Often little guidance about the choice of variables in the model, apart from standard model selection techniques

- We can make this more precise by using causal reasoning

Outline for the remainder of the talk

- Terminology:
  - Causal versus non-causal questions
  - Experimental versus observational data
  - Total causal effects
- Adjustment sets for total causal effects:
  - What are valid and invalid sets?
  - Which valid set provides the most efficient estimator?
  - Examples
- What to do when the causal structure is unknown?
- Summary and conclusion

Causal versus non-causal questions

- Non-causal questions are about predictions in the same system:
  - Predicting life expectancy of smokers
  - Predicting the recidivism rate of prisoners based on their participation in a rehabilitation program and other covariates
  - ....

- Causal questions are about the mechanism behind the data or about predictions after some outside intervention
  - Does smoking cause lung cancer?
  - Does the rehabilitation program for prisoners lower the recidivism rate?
  - How much money is saved by a health insurance company by assigning case managers to patients with complex diseases?
  - ...

Estimating causal effects from observational data

- Causal questions are ideally answered by randomized controlled experiments. Examples:
  - agricultural experiments
  - clinical trials to test new drugs

  If possible: do such experiments!

- Sometimes such experiments are impossible, as they may be:
  - unethical (smoking)
  - infeasible (global warming)
  - expensive / time consuming (gene knock-outs)

- How to estimate causal effects from observational data?

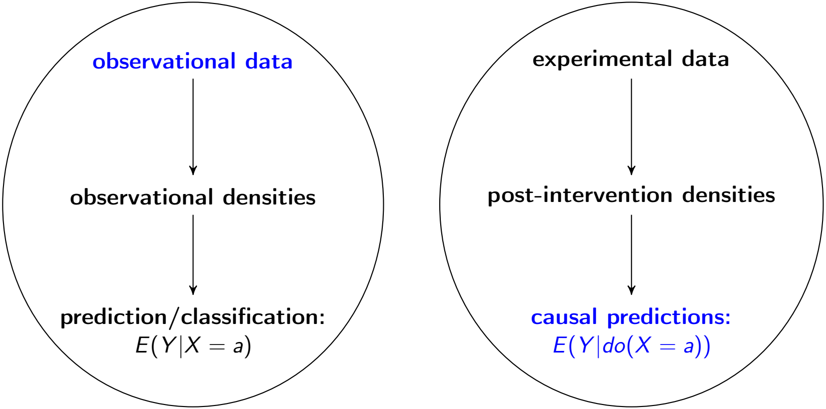
Definition of total causal effect

- Interventional notion of causal effect: If I set X to different values by an outside intervention, do I see a change in Y?
- do-operator (Pearl): $\text{do}(X = a)$: mathematical notation for setting the variable X to the value a by an outside intervention
  - Telmed example:
    - Let Y be health care costs
    - Let $X = 1$ if a patient uses a Telmed model; $X = 0$ otherwise
    - $E(Y|do(X = 1))$ versus $E(Y|X = 1)$

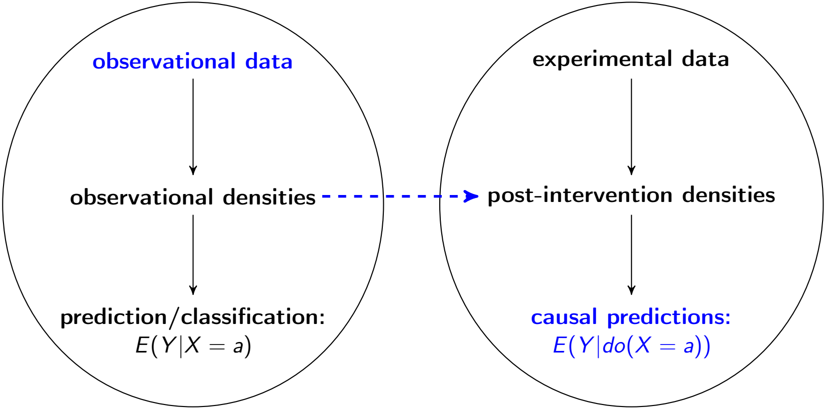
Definition of total causal effect

- Interventional notion of causal effect: If I set X to different values by an outside intervention, do I see a change in Y?
- do-operator (Pearl): $\text{do}(X = a)$: mathematical notation for setting the variable X to the value a by an outside intervention
  - Telmed example:
    - Let Y be health care costs
    - Let $X = 1$ if a patient uses a Telmed model; $X = 0$ otherwise
    - $E(Y|do(X = 1))$ versus $E(Y|X = 1)$
- Total causal effect of X on Y: $\frac{\partial}{\partial a} E(Y|do(X = a))$
  - Telmed example:
    - $E(Y|do(X = 1)) - E(Y|do(X = 0))$ versus $E(Y|X = 1) - E(Y|X = 0)$

How can we estimate causal effects from observational data?



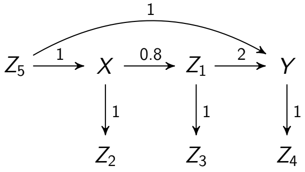
How can we estimate causal effects from observational data?



Common assumption:
Data are generated from a known directed acyclic graph (DAG)

Example: linear structural equation model

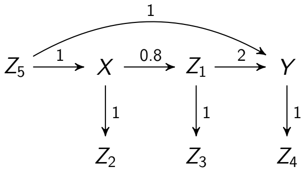
Directed acyclic graph (DAG) with weighted edges:



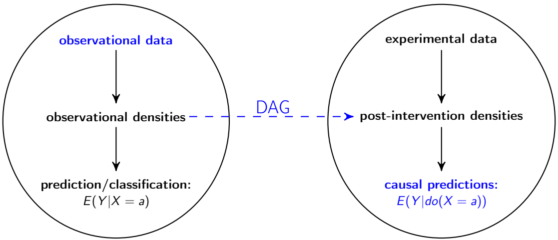
Each variable is generated as a linear function of its parents:

```
n   <- 100000                         # sample size
eps <- matrix(rnorm(7*n,0,1), ncol=7)   # random noise
Z5  <- eps[,1]
X   <- Z5 + eps[,2]
Z1  <- 0.8*X + eps[,3]
Y   <- 2*Z1 + Z5 + eps[,4]
Z2  <- X  + eps[,5]
Z3  <- Z1 + eps[,6]
Z4  <- Y  + eps[,7]
```

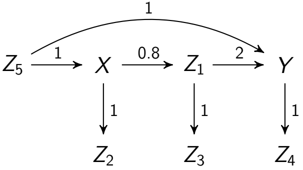
Example: linear structural equation model



Key assumption is autonomy: each structural equation is invariant to changes in the other structural equations.



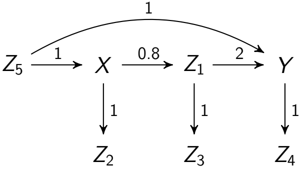
Example: linear structural equation model



We can easily simulate from this system under $do(X = a)$:

```
n   <- 100000
eps <- matrix(rnorm(7*n,0,1), ncol=7)
Z5  <- eps[,1]
X   <- rep(a,n)              # before: X <- Z5 + eps[,2]
Z1  <- 0.8*X + eps[,3]
Y   <- 2*Z1 + Z5 + eps[,4]
Z2  <- X  + eps[,5]
Z3  <- Z1 + eps[,6]
Z4  <- Y  + eps[,7]
```

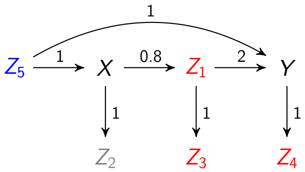
Example: linear structural equation model



The total effect of X on Y in this example is 1.6.
Now suppose we know the DAG, but not the edge weights:

- can we compute the total effect of X on Y via regression?
- if so, what variables to adjust for / not to adjust for?

Example: linear structural equation model




```
> lm(Y~X)$coeff[2]
2.0975                              # wrong
> lm(Y~X+Z5)$coeff[2]
1.6038                              # OK!
> lm(Y~X+Z5+Z2)$coeff[2]
1.6025                              # OK!
> lm(Y~X+Z5+Z3)$coeff[2]
0.8082                              # wrong
> lm(Y~X+Z1+Z2+Z3+Z4+Z5)$coeff[2]
0.000185                            # wrong
```

Identifying total causal effects

Set-up:

- Given: i.i.d. observational data and causal DAG
- Goal: identify $f(y|do(x))$ via covariate adjustment

Identifying total causal effects

- A probability density f is compatible with a causal DAG $\mathcal{D}$ if for any $\mathbf{X} \subseteq \mathbf{V}$ we have $f(\mathbf{v}\setminus\mathbf{x} \,|\, do(\mathbf{X} = \mathbf{a})) = \prod_{V_j \in \mathbf{V}\setminus\mathbf{X}} f(v_j | pa(v_j, \mathcal{D}))\Big|_{\mathbf{X}=\mathbf{a}}$

Identifying total causal effects

- A probability density f is compatible with a causal DAG $\mathcal{D}$ if for any $\mathbf{X} \subseteq \mathbf{V}$ we have
  $f(\mathbf{v} \setminus \mathbf{x} \mid do(\mathbf{X} = \mathbf{a})) = \prod_{V_j \in \mathbf{V} \setminus \mathbf{X}} f(v_j | pa(v_j, \mathcal{D}))\Big|_{\mathbf{X}=\mathbf{a}}$
- $\mathbf{S}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in a causal DAG $\mathcal{D}$ if for any f compatible with $\mathcal{D}$:

$$f(\mathbf{y}|do(\mathbf{x})) = \begin{cases} f(\mathbf{y}|\mathbf{x}) & \text{if } \mathbf{S} = \emptyset, \\ \int_{\mathbf{S}} f(\mathbf{y}|\mathbf{x}, \mathbf{s}) f(\mathbf{s}) d\mathbf{s} = E_{\mathbf{S}}\{f(\mathbf{y}|\mathbf{x}, \mathbf{s})\} & \text{otherwise} \end{cases}$$

Identifying total causal effects

- A probability density f is compatible with a causal DAG $\mathcal{D}$ if for any $\mathbf{X} \subseteq \mathbf{V}$ we have $f(\mathbf{v} \backslash \mathbf{x} \mid do(\mathbf{X} = \mathbf{a})) = \prod_{V_j \in \mathbf{V} \backslash \mathbf{X}} f(v_j | pa(v_j, \mathcal{D})) \Big|_{\mathbf{X}=\mathbf{a}}$
- $\mathbf{S}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in a causal DAG $\mathcal{D}$ if for any f compatible with $\mathcal{D}$:

$$f(\mathbf{y}|do(\mathbf{x})) = \begin{cases} f(\mathbf{y}|\mathbf{x}) & \text{if } \mathbf{S} = \emptyset, \\ \int_{\mathbf{S}} f(\mathbf{y}|\mathbf{x}, \mathbf{s}) f(\mathbf{s}) d\mathbf{s} = E_{\mathbf{S}}\{f(\mathbf{y}|\mathbf{x}, \mathbf{s})\} & \text{otherwise} \end{cases}$$

- In a linear system, the total effect of X on Y is then the coefficient of X in the linear regression $Y \sim X + \mathbf{S}$

Graphical criteria for adjustment sets

- Sufficient:
  Backdoor criterion (Pearl '93)

- Necessary and sufficient:
  Adjustment criterion (Shpitser et al '10; Perković et al '15, '17, '18)

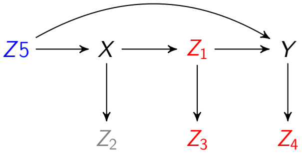
Adjustment criterion

S satisfies the adjustment criterion relative to (X, Y) in DAG $\mathcal{D}$ if:

- no node in **S** is a descendant of any $W \notin X$ that lies on a causal path from X to Y in $\mathcal{D}$; and
- **S** blocks all non-causal paths from X to Y in $\mathcal{D}$

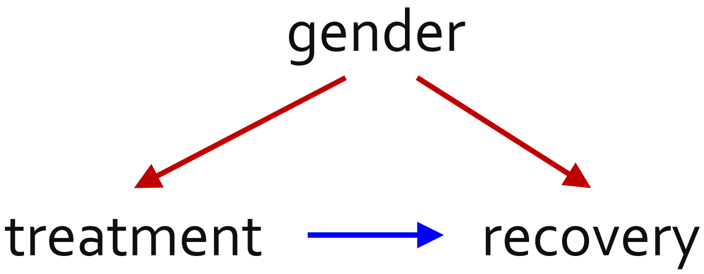
S is an adjustment set relative to (X, Y) in a causal DAG $\mathcal{D}$ $\Leftrightarrow$
S satisfies these graphical criteria relative to (X, Y) in $\mathcal{G}$

Back to the linear structural equation model

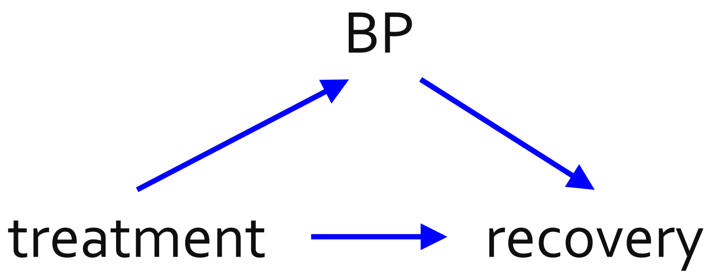


- Z_5 is required to block the non-causal path $X \leftarrow Z_5 \rightarrow Y$
- Z_1, Z_3 and Z_4 are forbidden, since they are descendants of nodes other than X on a causal path
- Z_2 is optional
- Hence, the only valid adjustment sets are $\{Z_5\}$ and $\{Z_2, Z_5\}$

Back to Simpson's paradox



control for Z=gender;
(Z is before treatment)



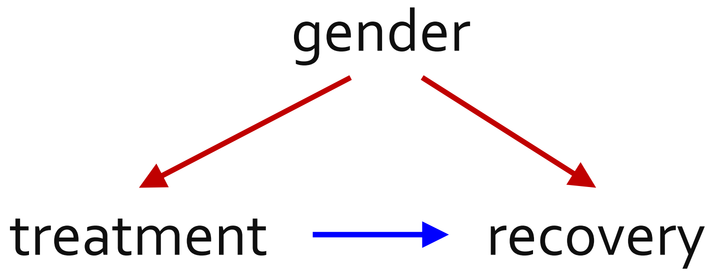
don't control for Z=BP;
(Z is after treatment)

Rule: control for pre-treatment covariates, and not for post-treatment covariates?

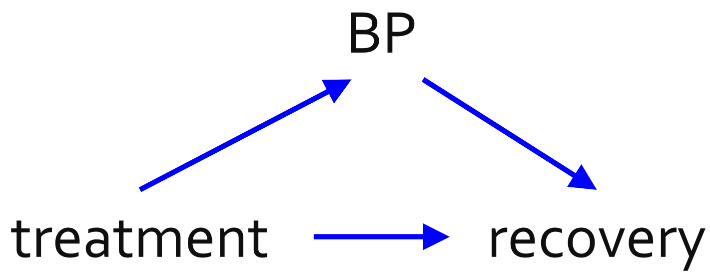
No....

Reference: Pearl (2014). The American Statistician 68, 8-13.

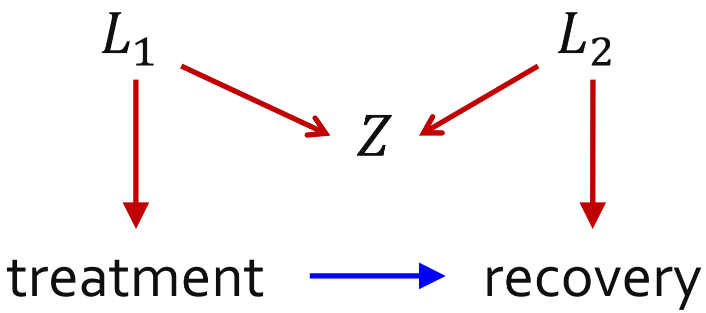
Back to Simpson's paradox



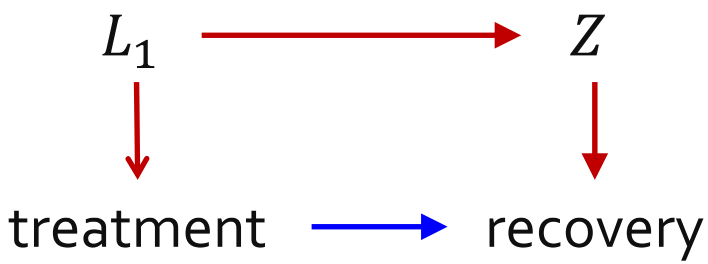
control for Z=gender;
(Z is before treatment)



don't control for Z=BP;
(Z is after treatment)



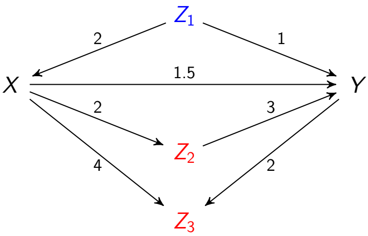
don't control for Z;
(Z can be before/after treatment)



control for Z
(Z can be before/after treatment)

Reference: Pearl (2014). The American Statistician 68, 8-13.

Example: mutual adjustment or "Table 2 fallacy"



- $\{Z1\}$ is a valid adj. set for the total effect of X on Y.
- $\{X\}$ is not a valid adj. set for the total effect of $Z1$ on Y.
- There is no "mutual adjustment".

```
> lm(Y~X+Z1)$coef
Intercept          X        Z1
0.004809      7.4961    1.0072
```

Example: mutual adjustment or "Table 2 fallacy"

Table 2. Cause-specific Hazard Ratios for Use of Psychotropics in Association With Work-related Violence and Covariates, Denmark, 1996–2008

	Antidepressants		Antidepressants and Anxiolytics		Anxiolytics		Hypnotics Only	
	HR[a]	95% CI	HR[a]	95% CI	HR[a]	95% CI	HR[a]	95% CI
Work-related violence (yes vs. no)	1.38	1.09, 1.75	1.74	1.13, 2.70	1.05	0.76, 1.45	1.05	0.75, 1.46
Women vs. men	1.41	1.18, 1.68	1.56	1.09, 2.22	1.73	1.39, 2.16	1.45	1.17, 1.79
Age per 5-year increase	1.01	0.98, 1.06	1.09	1.00, 1.17	1.16	1.10, 1.22	1.21	1.15, 1.27
Cohabitation (yes vs. no)	0.81	0.67, 0.97	0.89	0.61, 1.29	0.86	0.68, 1.09	0.82	0.64, 1.03
Education per SD increase	0.88	0.81, 0.96	0.86	0.73, 1.02	0.97	0.87, 1.07	1.04	0.94, 1.15
Income per quartile increase	0.91	0.83, 0.99	0.78	0.65, 0.93	0.97	0.87, 1.09	1.06	0.94, 1.18
Social support from colleagues per unit increase	0.95	0.87, 1.05	0.93	0.78, 1.12	0.97	0.87, 1.09	1.00	0.89, 1.12
Social support from supervisor per unit increase	0.95	0.87, 1.03	1.00	0.85, 1.18	1.05	0.95, 1.16	0.97	0.87, 1.07
Influence per unit increase	0.93	0.86, 1.02	0.95	0.81, 1.12	0.96	0.87, 1.05	0.94	0.85, 1.04
Quantitative demands per SD increase	1.02	0.94, 1.11	1.04	0.88, 1.22	0.97	0.88, 1.08	0.99	0.89, 1.10

Abbreviations: CI, confidence interval; HR, hazard ratio; SD, standard deviation.
[a] Statistical model includes the following: work-related violence, gender, age, cohabitation, education, income, social support from colleagues, social support from supervisor, influence at work, and quantitative demands at work.

From: Am. J. of Epidemiology

Take home message so far

Many common ideas about adjustment are wrong:

- adjusting for more variables is always better. No!
- one should adjust for all variables correlated to X and Y. No!
- adjusting for pre-treatment variables is always safe. No!
- adjusting for descendants of X is always bad. No!
- mutual adjustment works. No!
- ...

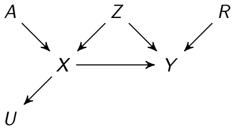
Take home message so far

Many common ideas about adjustment are wrong:

- adjusting for more variables is always better. No!
- one should adjust for all variables correlated to X and Y. No!
- adjusting for pre-treatment variables is always safe. No!
- adjusting for descendants of X is always bad. No!
- mutual adjustment works. No!
- ...

Valid approach: use graphical criteria for adjustment sets

Efficiency: among the valid sets, which one should we use?

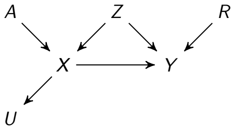


Asymptotic variances for 7 random parameter settings:

Valid adj. sets	1	2	3	4	5	6	7
$\{Z, A\}$	4.59	1.64	0.59	0.60	4.59	1.76	2.16
$\{Z, A, R\}$	4.21	1.18	0.56	0.19	3.08	1.60	0.11
$\{Z\}$	4.58	0.97	0.57	0.34	3.92	1.40	2.16
$\{Z, R\}$	4.19	0.70	0.54	0.11	2.63	1.27	0.11
$\{Z, A, U\}$	4.71	19.49	0.59	0.90	5.23	2.73	5.92
$\{Z, A, R, U\}$	4.32	14.04	0.56	0.29	3.51	2.48	0.31
$\{Z, U\}$	4.70	18.82	0.57	0.64	4.56	2.37	5.92
$\{Z, R, U\}$	4.30	13.56	0.54	0.21	3.06	2.15	0.31

$\{Z, A, R\}$ is always better than $\{Z, A\}$

Efficiency: among the valid sets, which one should we use?

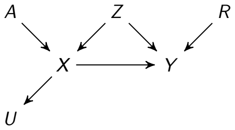


Asymptotic variances for 7 random parameter settings:

Valid adj. sets	1	2	3	4	5	6	7
$\{Z, A\}$	4.59	1.64	0.59	0.60	4.59	1.76	2.16
$\{Z, A, R\}$	4.21	1.18	0.56	0.19	3.08	1.60	0.11
$\{Z\}$	4.58	0.97	0.57	0.34	3.92	1.40	2.16
$\{Z, R\}$	4.19	0.70	0.54	0.11	2.63	1.27	0.11
$\{Z, A, U\}$	4.71	19.49	0.59	0.90	5.23	2.73	5.92
$\{Z, A, R, U\}$	4.32	14.04	0.56	0.29	3.51	2.48	0.31
$\{Z, U\}$	4.70	18.82	0.57	0.64	4.56	2.37	5.92
$\{Z, R, U\}$	4.30	13.56	0.54	0.21	3.06	2.15	0.31

$\{Z, A, R\}$ and $\{Z\}$ cannot be compared

Efficiency: among the valid sets, which one should we use?



Asymptotic variances for 7 random parameter settings:

Valid adj. sets	1	2	3	4	5	6	7
$\{Z, A\}$	4.59	1.64	0.59	0.60	4.59	1.76	2.16
$\{Z, A, R\}$	4.21	1.18	0.56	0.19	3.08	1.60	0.11
$\{Z\}$	4.58	0.97	0.57	0.34	3.92	1.40	2.16
$\{Z, R\}$	4.19	0.70	0.54	0.11	2.63	1.27	0.11
$\{Z, A, U\}$	4.71	19.49	0.59	0.90	5.23	2.73	5.92
$\{Z, A, R, U\}$	4.32	14.04	0.56	0.29	3.51	2.48	0.31
$\{Z, U\}$	4.70	18.82	0.57	0.64	4.56	2.37	5.92
$\{Z, R, U\}$	4.30	13.56	0.54	0.21	3.06	2.15	0.31

$\{Z, R\}$ yields optimal asymptotic variance, regardless of parameters

Efficiency

- We have graphical criteria for the optimal adjustment set in terms of asymptotic variance
  (Henckel et al, 2019; Witte et al, 2019)
- Intuition: Try to explain as much variance of Y as possible while avoiding unnecessary correlation with X

Complication: the DAG may be unknown

- Approach 1: Hypothesize possible DAGs
  - Drawing DAGs formalizes the causal assumptions
  - Each hypothesized DAG can be used to determine valid adjustment sets and the corresponding total effect
  - Allows sensitivity analysis and informed discussion

Complication: the DAG may be unknown

- Approach 2: Try to learn the DAG from data
- Under some assumptions, one can learn an equivalence class of DAGs that could have generated the data:
  - Assuming no latent variables:
    PC, GES, MMHC, ARGES $\Rightarrow$ CPDAG
    (Spirtes et al '00, Chickering '02, Tsarmardinos et al '06, Nandy et al '17)
  - Allowing arbitrarily many latent variables:
    FCI, RFCI, FCI+ $\Rightarrow$ PAG
    (Spirtes et al '00, Colombo et al '12, Claassen et al '13)
- Then use adjustment again

Summary

- The variables that are included in a model matter
- Carefully define research question. If interested in causal effects:
  - specify the type of causal effect:
    total effect, direct effect, indirect effect, ...
  - state causal assumptions (e.g., draw DAG)
  - use causal methods (e.g., graphical criteria for covariate adjustment)

Summary

- The variables that are included in a model matter
- Carefully define research question. If interested in causal effects:
  - specify the type of causal effect:
    total effect, direct effect, indirect effect, ...
  - state causal assumptions (e.g., draw DAG)
  - use causal methods (e.g., graphical criteria for covariate adjustment)
- This does not replace randomized controlled trials, but:
  - it uses observational data in a principled way
  - it allows formal discussion
  - it allows sensitivity analysis wrt different causal assumptions
  - if possible, follow-up with validation experiments

Open source software

R packages pcalg and dagitty:

- pc, ges, fci, rfci
- gac($\mathcal{G}$, $\mathbf{X}, \mathbf{Y}, \mathbf{S}$, graph.type)
- adjustment($\mathcal{G}$, graph.type, $\mathbf{X}, \mathbf{Y}$, type="all")
- optAdjSet($\mathcal{G}, \mathbf{X}, \mathbf{Y}$)


THANK YOU

References

Books:

- Spirtes, Glymour and Scheines (2000). *Causation, Prediction, and Search.* MIT Press, Cambridge.
- Pearl (2009). *Causality.* Cambridge University Press, New York.

Covariate adjustment:

- Shpitser, Vanderweele and Robins (2010). On the validity of covariate adjustment for estimating causal effects. *UAI 2010*.
- Maathuis and Colombo (2015). A generalized back-door criterion. *Ann. Stat.*
- Perković, Textor, Kalisch and Maathuis (2015). A complete generalized adjustment criterion. *UAI 2015*.
- Perković, Kalisch, Maathuis (2017). Interpreting and using CPDAGs with background knowledge. *UAI 2017*.
- Perković, Textor, Kalisch and Maathuis (2018). The generalized adjustment criterion. *JMLR*.
- Henckel, Perkovic and Maathuis (2019). Graphical criteria for efficient total effect estimation via adjustment in causal linear models. arXiv:1907.02435.
- Witte, Henckel, Maathuis and Didelez (2019). On efficient adjustment in causal graphs. Working paper

References

Causal structure learning without latent variables:

- Tsamardinos, Brown and Aliferis (2006). The max-min hill- climbing Bayesian network structure learning algorithm. *JMLR*.
- Chickering (2002). Learning equivalence classes of Bayesian-network structures. *JMLR*.
- Colombo and Maathuis (2015). Order-independent constraint-based causal structure learning. *JMLR*.
- Nandy, Hauser and Maathuis (2018). High-dimensional consistency in score-based and hybrid structure learning. *Ann. Stat.*

Causal structure learning with latent variables:

- Colombo, Maathuis, Kalisch and Richardson (2012). Learning high-dimensional directed acyclic graphs with latent and selection variables. *Ann. Stat.*
- Claassen, Mooij and Heskes (2013). Learning sparse causal models is not NP-hard. *UAI 2013*.
- Frot, Nandy and Maathuis (2019). Learning directed acyclic graphs with hidden variables via latent Gaussian graphical model selection. *JRSS-B*.

References

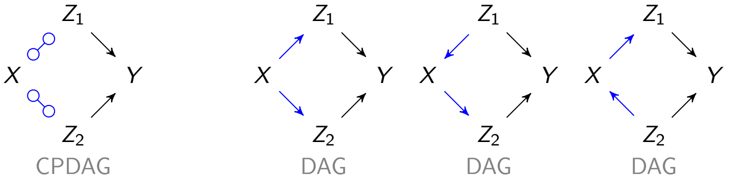
Estimating bounds on total causal effects from a CPDAG:

- Maathuis, Kalisch and Bühlmann (2009). Estimating high-dimensional intervention effects from observational data. *Ann. Stat.*
- Maathuis, Colombo, Kalisch and Bühlmann (2010). Predicting causal effects in large-scale systems from observational data. *Nature Methods*.
- Malinsky and Spirtes (2017). Estimating bounds on causal effects in high-dimensional and possibly confounded systems. *Int. J. Appr. Reason.*
- Nandy, Maathuis and Richardson (2017). Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *Ann. Statist.*

R-packages:

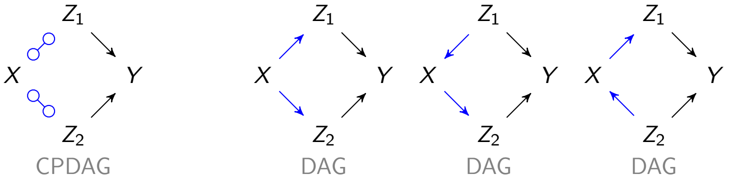
- pcalg
- dagitty

Examples

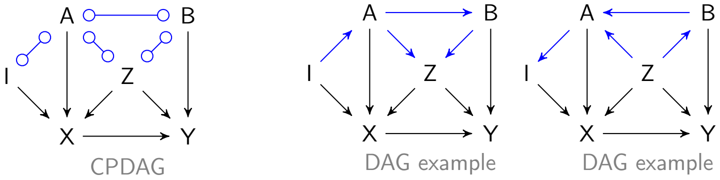


There is no adjustment set for the effect of X on Y.

Examples



There is no adjustment set for the effect of X on Y.



The set $\{A, Z\}$ is an adjustment set for the effect of X on Y.

Overview of graphical criteria

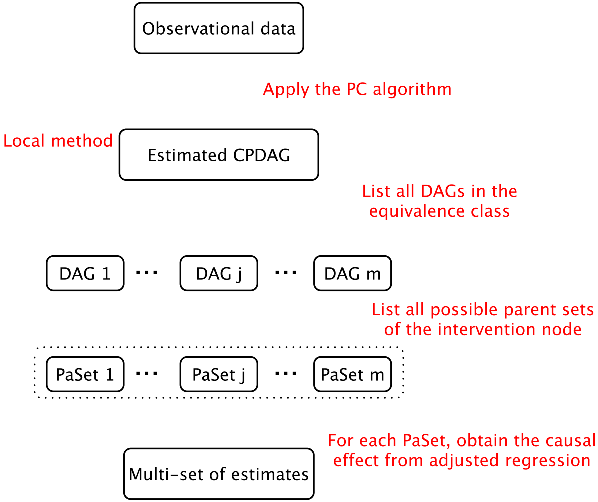
	DAG	MAG	CPDAG	PAG
Backdoor criterion Pearl '93	✓			
Adjustment criterion Shpitser et al '12, Perković et al '17a	✓			
Adjustment criterion Van der Zander et al '14	✓	✓		
Generalized backdoor criterion MM & Colombo '15	✓	✓	✓	✓
Generalized adjustment criterion Perković et al '15, '17a	✓	✓	✓*	✓

✓: sufficient for adjustment

✓: necessary and sufficient for adjustment

*: CPDAGs with background knowledge (Perković et al '17b)

Going away from identifiability: IDA



IDA

- Assuming no latent variables:
  - IDA (MM et al '09,'10)
- Allowing arbitrarily many latent variables:
  - LV-IDA (Malinsky & Spirtes '17)
- Allowing some latent variables:
  - LGES-IDA (Frot et al '17)

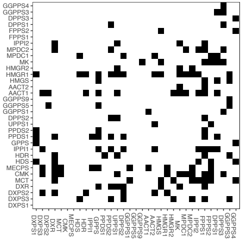
Application

Gene expression data of *Arabidopsis thaliana*:

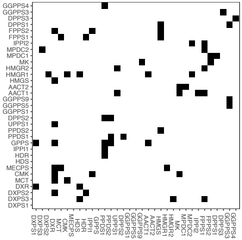
- Data: $n = 188$, $p = 33$ (Wille et al '04)
- Three groups of genes:
  MVA pathway, MEP pathway, mitochondrial genes
  (we do not use this information)

Results

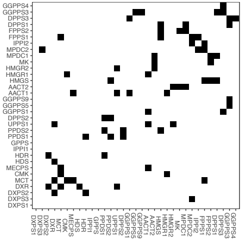
GES



LGES

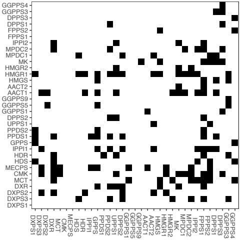


RFCI

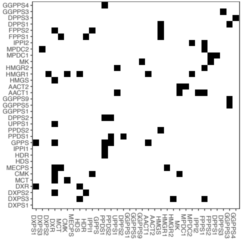


Results

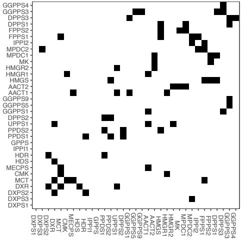
GES



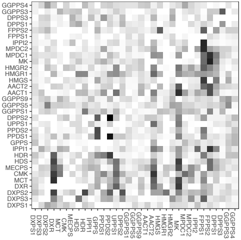
LGES



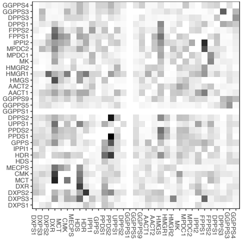
RFCI



GES-IDA



LGES-IDA



LV-IDA