Week 5: A Causal Inference Perspective on Methodological Issues

Causal Inference & Structural Equation Modeling

Noémi K. Schuurman
based on slides by Oisín Ryan and Ellen Hamaker

March 2022

Causal graphs and conceptual clarity

Causal models imply statistical models - causal models contain information not contained statistical models

- Many questions which are confusing, difficult or impossible to answer in purely statistical terms become clear when we take a causal inference perspective.
- I.e., Draw the relevant causal model!

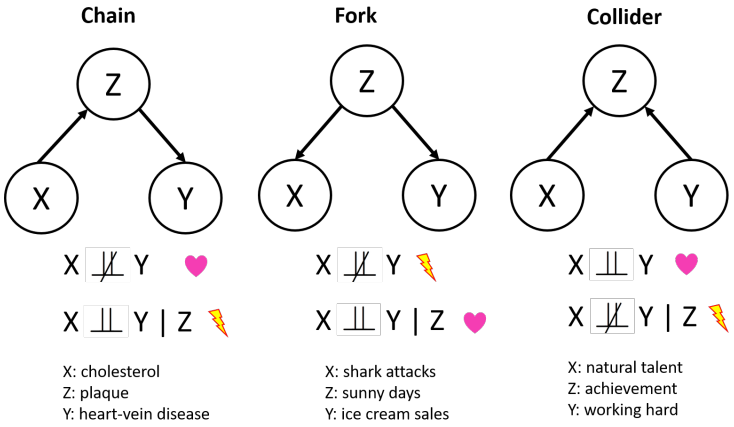
Motto of Miguel Hernan: Draw your assumptions before your conclusions!

Overview

- **DAGs & Interventions & the RCT**
- Selection Bias, Berksons Paradox, Simpsons Paradox
- Repeated Measures: Change Score vs Controlling for Pre-measure

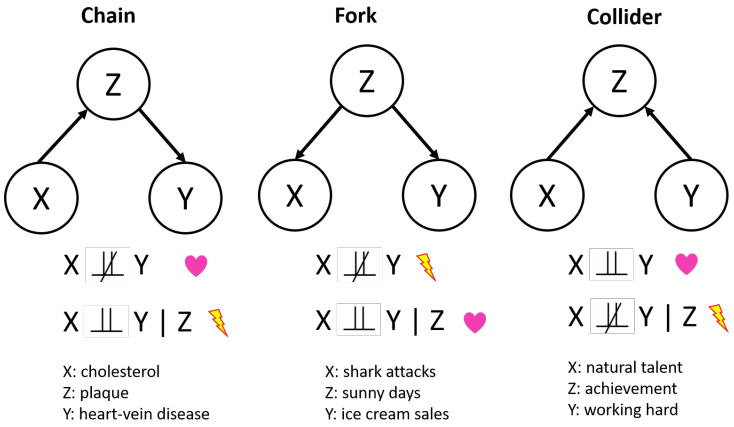
DAGs, SCMs, and interventions

Lecture 1: On which variables should we (not) condition to obtain the causal effect for observational data?



DAGs, SCMs, and interventions

Lecture 1: On which variables should we (not) condition to obtain the causal effect for observational data?



- What if we intervened on the system represented by the DAG?

DAGs, SCMs, and interventions

Remember: Conditioning in observational data, and intervening are not the same actions.

DAGs, SCMs, and interventions

Remember: Conditioning in observational data, and intervening are not the same actions.

- Conditioning on X, observational data: Let's look at people in our dataset that have X=1, and then at the people who have X=0.

DAGs, SCMs, and interventions

Remember: Conditioning in observational data, and intervening are not the same actions.

- Conditioning on X, observational data: Let's look at people in our dataset that have X=1, and then at the people who have X=0.
- Intervening on X: Let's *set* these people's X to 1 (experimental group), and *set* these people's X to 0 (control group).

DAGs, SCMs, and interventions: The "do-operator"

To represent interventions in SCMs, we use the "do-operator":

do-operator:

The do-operator $do(X = x)$ represents a "surgical intervention" to set the value of the variable X to a constant value x

DAGs, SCMs, and interventions: The "do-operator"

To represent interventions in SCMs, we use the "do-operator":

do-operator:

The do-operator $do(X = x)$ represents a "surgical intervention" to set the value of the variable X to a constant value x

"Surgical" interventions: Modularity

Assume that it is possible to intervene on a variable without fundamentally changing how it relates to other variables, e.g.:

- We can intervene on X without changing $p(Z \mid X)$
- We can intervene on one cause-effect mechanism without changing the others

Read more about such assumptions by searching the jargon 'Modularity', 'Localized Interventions' and 'Fat Hand Interventions'.

Average Causal Effect - DAG edition

We can use the DAGs, SCM and the do-operator to define and estimate any causal effect based on an intervention we want:

Often we are interested in the effect of an intervention on the *mean* of our outcome variable - the effect of the intervention on average across different people.

Average causal effect of X (0 vs 1) on Y:

$$ACE = E[Y \mid do(X = x_1)] - E[Y \mid do(X = x_0)]$$

Average Causal Effect - DAG edition

We can use the DAGs, SCM and the do-operator to define and estimate any causal effect based on an intervention we want:

Often we are interested in the effect of an intervention on the *mean* of our outcome variable - the effect of the intervention on average across different people.

Average causal effect of X (0 vs 1) on Y:

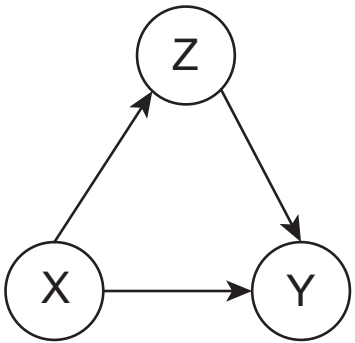
$$ACE = E[Y \mid do(X = x_1)] - E[Y \mid do(X = x_0)]$$

Note: $ACE = E[Y \mid do(X = x_1)] - E[Y \mid do(X = x_0)] = E[Y_i^1] - E[Y_i^0]$

Interventions with DAGs and SCMs: Partial Mediation Example

The observational DAG and SCM:

$$
\begin{aligned}
X &: \ bernoulli(0.5) \\
Z &:= 2X + \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$



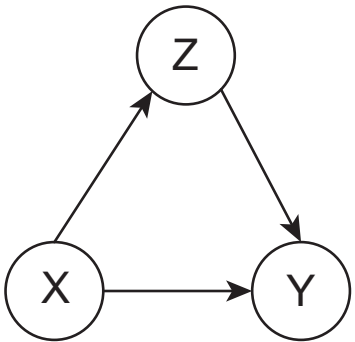
where

- X is bernoulli distributed (0 or 1) with probability 0.5, ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$

Interventions with DAGs and SCMs: Partial Mediation Example

The observational DAG and SCM:

$$
\begin{aligned}
X &: \ bernoulli(0.5) \\
Z &:= 2X + \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$



where

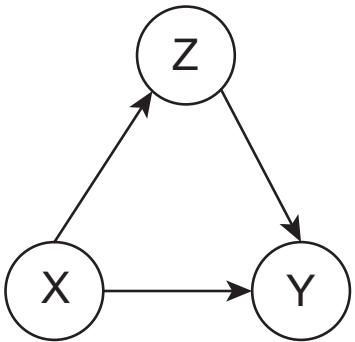
- X is bernoulli distributed (0 or 1) with probability 0.5, ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$

Prima Facie Effect: $E[Y \mid X = 1] - E[Y \mid X = 0] =$

Interventions with DAGs and SCMs: Partial Mediation Example

The observational DAG and SCM:

$$
\begin{aligned}
X &: \ bernoulli(0.5) \\
Z &:= 2X + \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$



where

- X is bernoulli distributed (0 or 1) with probability 0.5, ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$

Prima Facie Effect: $E[Y \mid X = 1] - E[Y \mid X = 0] = 5 - 0 = 5$

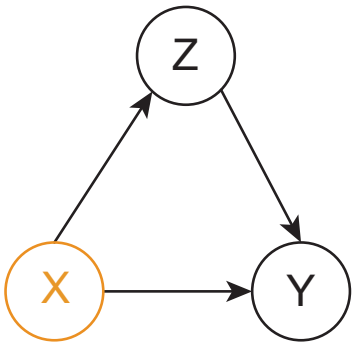
Interventions with DAGs and SCMs: Partial Mediation Example

SCMs model with intervention $do(X = 1)$.

$$
\begin{aligned}
X &:= 1 \\
Z &:= 2X + \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$

where

- ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$



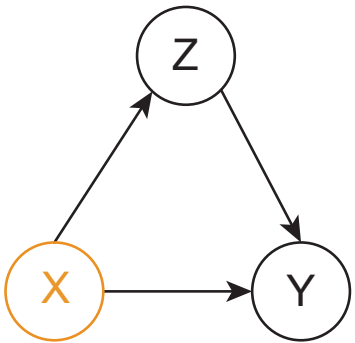
Interventions with DAGs and SCMs: Partial Mediation Example

SCMs model with intervention $do(X = 1)$.

$$
\begin{aligned}
X &:= 1 \\
Z &:= 2X + \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$

where

- ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$



Average Causal Effect: $E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] =$

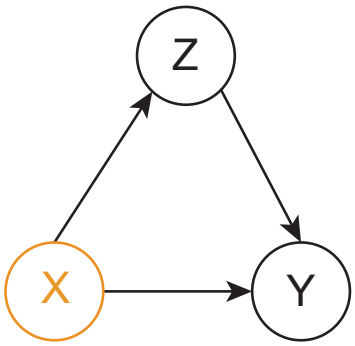
Interventions with DAGs and SCMs: Partial Mediation Example

SCMs model with intervention $do(X = 1)$.

$$
\begin{aligned}
X &:= 1 \\
Z &:= 2X + \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$

where

- ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$

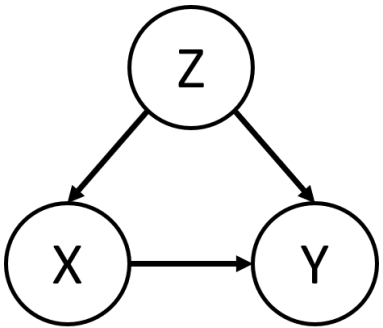


Average Causal Effect: $E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] = 5 - 0 = 5$

Interventions with DAGs and SCMs: Partial Confounding Example

Another observational DAG and SCM:

$$
\begin{aligned}
logodds(X) &:= 2Z \\
Z &:= \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$



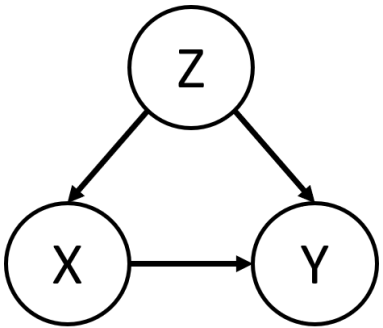
where

- X can be 0 or 1 and follows a logistic regression model, ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$

Interventions with DAGs and SCMs: Partial Confounding Example

Another observational DAG and SCM:

$$
\begin{aligned}
logodds(X) &:= 2Z \\
Z &:= \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$



where

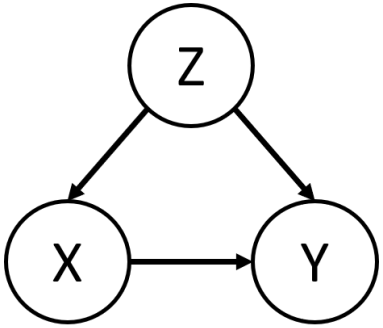
- X can be 0 or 1 and follows a logistic regression model, ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$

Prima Facie Effect: $E[Y \mid X = 1] - E[Y \mid X = 0] \sim$

Interventions with DAGs and SCMs: Partial Confounding Example

Another observational DAG and SCM:

$$
\begin{aligned}
logodds(X) &:= 2Z \\
Z &:= \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$



where

- X can be 0 or 1 and follows a logistic regression model, ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$

Prima Facie Effect: $E[Y \mid X = 1] - E[Y \mid X = 0] \sim$
$\sim 2.5 - (-1.5) = 4$

If someone proves this exactly I'll treat the class to boterkoek in our next meeting (for the approximation via simulation see rcode on bb).

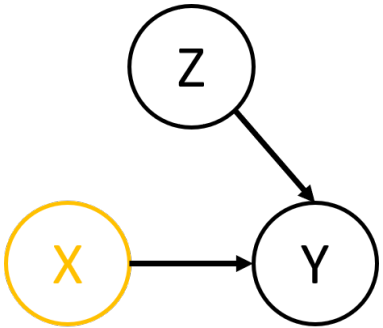
Interventions with DAGs and SCMs: Example 2

SCMs model the intervention $do(X = 1)$.
Note: X is now no longer affected by Z. We fully control the value of X.

$$
\begin{aligned}
X &:= 1 \\
Z &:= \epsilon_Z \\
Y &:= 1X + 2Z + \epsilon_Y
\end{aligned}
$$

where

- ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$



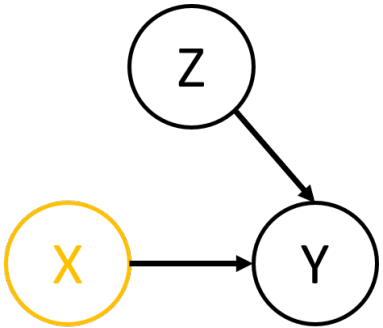
Interventions with DAGs and SCMs: Example 2

SCMs model the intervention $do(X = 1)$.
Note: X is now no longer affected by Z. We fully control the value of X.

$$X := 1$$
$$Z := \epsilon_Z$$
$$Y := 1X + 2Z + \epsilon_Y$$

where

- ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$



Average Causal Effect: $E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] =$

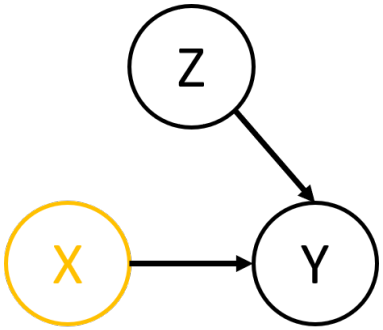
Interventions with DAGs and SCMs: Example 2

SCMs model the intervention $do(X = 1)$.
Note: X is now no longer affected by Z. We fully control the value of X.

$$X := 1$$
$$Z := \epsilon_Z$$
$$Y := 1X + 2Z + \epsilon_Y$$

where

- ϵ_Z, ϵ_Y are iid, $\sim \mathcal{N}(0, 1)$



Average Causal Effect: $E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] = 1 - 0 = 1$

Interventions vs Conditioning

Observing/Seeing $\neq$ Intervening/Doing:

$E[Y \mid A = a]$ is *not* necessarily the same as $E[Y \mid do(A = a)]$

When statistical relationship $\neq$ causal effect, we say the former is *confounded*.

Example 1 - partial mediation:

- $E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] = 5 - 0 = 5$
- $E[Y \mid X = 1] - E[Y \mid X = 0] = 5 - 0 = 5$

Interventions vs Conditioning

Observing/Seeing $\neq$ Intervening/Doing:

$E[Y \mid A = a]$ is *not* necessarily the same as $E[Y \mid do(A = a)]$

When statistical relationship $\neq$ causal effect, we say the former is *confounded*.

Example 1 - partial mediation:

- $E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] = 5 - 0 = 5$
- $E[Y \mid X = 1] - E[Y \mid X = 0] = 5 - 0 = 5$

Example 2 - partial confounding:

- $E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] = 1 - 0 = 1$
- $E[Y \mid X = 1] - E[Y \mid X = 0] \sim 2.5 - (-1.5) = 4$

Interventions vs Conditioning

Observing/Seeing $\neq$ Intervening/Doing:

$E[Y \mid A = a]$ is *not* necessarily the same as $E[Y \mid do(A = a)]$

When statistical relationship $\neq$ causal effect, we say the former is *confounded*.

Example 1 - partial mediation:

- $E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] = 5 - 0 = 5$
- $E[Y \mid X = 1] - E[Y \mid X = 0] = 5 - 0 = 5$

Example 2 - partial confounding:

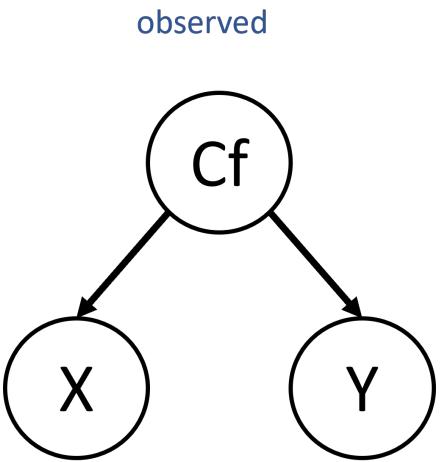
- $E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] = 1 - 0 = 1$
- $E[Y \mid X = 1] - E[Y \mid X = 0] \sim 2.5 - (-1.5) = 4$

Observationally, people with X=1 will have higher expected values for Z than people with X=0 ($\sim$ .8 vs -.8), so the mean of y will also be higher for the former!

Randomized Control Trials - Why They Work

RCTs are extremely powerful because randomization ensures no confounding.

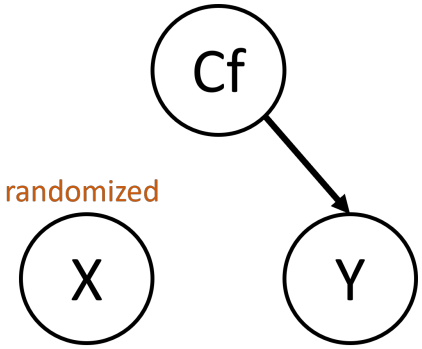
- Randomization (if succesfull) means having full control over the treatment variable.
- There can't be any backdoor paths if everyone has an equal probability of being treated or not
- But, RCTs not possible in many settings.



Randomized Control Trials - Why They Work

RCTs are extremely powerful because randomization ensures no confounding.

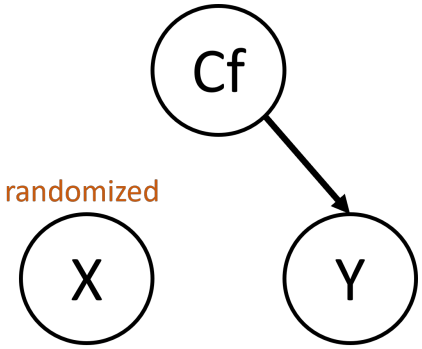
- Randomization (if succesfull) means having full control over the treatment variable.
- There can't be any backdoor paths if everyone has an equal probability of being treated or not
- But, RCTs not possible in many settings.



Randomized Control Trials - Why They Work

RCTs are extremely powerful because randomization ensures no confounding.

- Randomization (if succesfull) means having full control over the treatment variable.
- There can't be any backdoor paths if everyone has an equal probability of being treated or not
- But, RCTs not possible in many settings.



Note: From the Potential Outcomes perspective, (succesful) RCTs work because they by design adhere to all the assumptions (e.g., exchangeability, positivity, etc).

Overview

- DAGs & Interventions & the RCT
- **Selection Bias, Berksons Paradox, Simpsons Paradox**
- Repeated Measures: Change Scores vs Controlling for Pre-measure

Simpsons Paradox

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Simpsons Paradox

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Example (Pearl, Glymour & Jewell, 2016):

- 700 sick patients are given the choice to take a new drug: 350 choose to take it.
- We are interested in effects of a drug (D) on recovery (R). We also record the gender (G)
- Should we prescribe the drug?

Simpsons Paradox

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Example (Pearl, Glymour & Jewell, 2016):

- 700 sick patients are given the choice to take a new drug: 350 choose to take it.
- We are interested in effects of a drug (D) on recovery (R). We also record the gender (G)
- Should we prescribe the drug?

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Simpsons Paradox

Counter-intuitive, but not really a paradox

- A marginal dependency ($P(R \mid D)$) is not necessarily the same as a conditional dependency ($P(R \mid D, G = 0)$)
- But which piece of information should we use to make treatment decisions?
- (Who) should we treat?!

Simpsons Paradox

Counter-intuitive, but not really a paradox

- A marginal dependency ($P(R \mid D)$) is not necessarily the same as a conditional dependency ($P(R \mid D, G = 0)$)
- But which piece of information should we use to make treatment decisions?
- (Who) should we treat?!

Draw your DAG!

Variables: Gender (sex; 'G'), Drug ('D') and Recovery ('R')

Simpsons Paradox

- Estrogen levels negatively affect recovery
- Women are more likely to take the drug than men
- We should condition on Gender - it blocks a backdoor path!

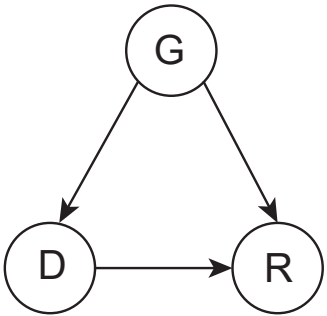


Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Simpsons Paradox

Suppose that we measured post-treatment blood pressure (B) instead of gender, next to Drug taking (D) and Recovery (R).

Draw your DAG!

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	Drug	No drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Simpsons Paradox

Suppose that we measure post-treatment blood pressure (B) instead

- Statistical information is exactly the same
- B cannot cause drug taking
- The drug works in part by decreasing blood pressure
- We should **not** condition on blood pressure

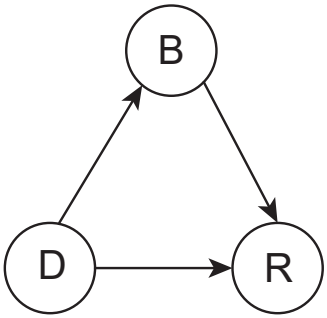
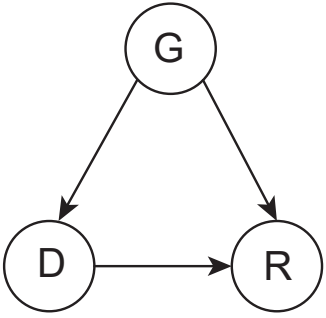


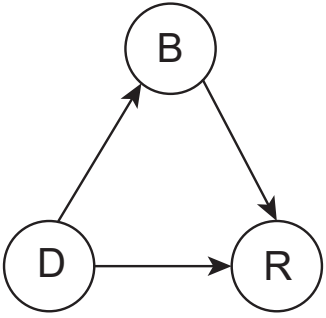
Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	Drug	No drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Simpsons Paradox

- Statistical information alone cannot provide the answer on when to treat (!)
- Two different DAGs can produce the exact same statistical dependencies in the observational setting
  - Observationally equivalent
- These DAGs imply different causal effects, and hence different models to estimate those effects from observational data.





Selection Bias

Berksons Paradox

Two phenomena which are statistically *independent* in the general population are statistically *dependent* in a sub-population that was selected.

Also know as: Selection Bias, Endogenous Selection Bias, Berskon's bias

Classic example: We are interested in the relationship between *Lung Cancer* (L) and *Diabetes* (D)

- General population, these two variables are independent.
- In a sample of *hospital patients*, there is a negative dependency - patients who don't have diabetes are *more likely* to have lung cancer.

Selection Bias

Berksons Paradox

Two phenomena which are statistically *independent* in the general population are statistically *dependent* in a sub-population that was selected.

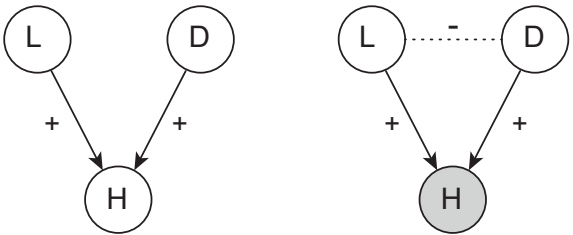
Also know as: Selection Bias, Endogenous Selection Bias, Berskon's bias

Classic example: We are interested in the relationship between *Lung Cancer* (L) and *Diabetes* (D)

- General population, these two variables are independent.
- In a sample of *hospital patients*, there is a negative dependency - patients who don't have diabetes are *more likely* to have lung cancer.

...Draw your DAG!

Selection Bias



- Lung cancer L and diabetes D cause hospitalization H
- By selecting participants from a hospital we *condition* on hospitalization ($H = 1$)
- If you are hospitalised, and you *don't* have diabetes, probably you do have lung cancer (Otherwise - why would you be in hospital?).
- We have conditioned on a *collider*!
- $P(D|L = 1, H = 1) \neq P(D|L = 1)$
- $P(D|L = 1, H = 1) \neq P(D|do(L) = 1)$

Simpsons or Berksons or?

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Berksons Paradox

Two phenomena which are statistically *independent* in the general population are statistically *dependent* in a sub-population that was selected.

What's the difference?

- In Simpsons, we find different relations when we control vs not control for a variable
- In Berksons, we find different relations as a result of 'accidental' selection via our sampling procedure.
- Either can be the result of collider bias or confounder bias or overcontrol bias (controlling for a mediator).

Simpsons or Berksons or?

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Berksons Paradox

Two phenomena which are statistically *independent* in the general population are statistically *dependent* in a sub-population that was selected.

What's the difference?

- In Simpsons, we find different relations when we control vs not control for a variable
- In Berksons, we find different relations as a result of 'accidental' selection via our sampling procedure.
- Either can be the result of collider bias or confounder bias or overcontrol bias (controlling for a mediator).

Note 1. When we have a partial mediation where (for example) the direct effect is positive and the indirect effect is negative (or other way around), so the total effect is near zero - This is called a 'supression effect'.

Simpsons or Berksons or?

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Berksons Paradox

Two phenomena which are statistically *independent* in the general population are statistically *dependent* in a sub-population that was selected.

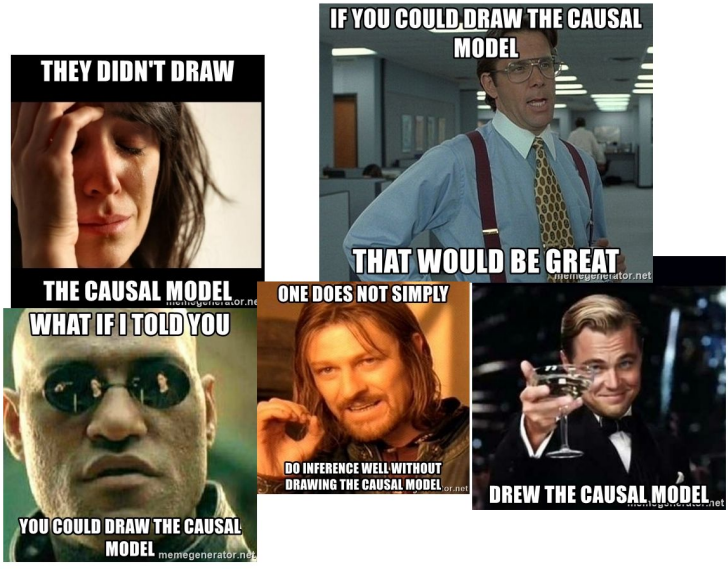
What's the difference?

- In Simpsons, we find different relations when we control vs not control for a variable
- In Berksons, we find different relations as a result of 'accidental' selection via our sampling procedure.
- Either can be the result of collider bias or confounder bias or overcontrol bias (controlling for a mediator).

Note 1. When we have a partial mediation where (for example) the direct effect is positive and the indirect effect is negative (or other way around), so the total effect is near zero - This is called a 'supression effect'.

Note 2. Some people relate Simpsons expressedly to confounding bias and Berksons to collider bias.

In any case...draw the causal model!



Recap: Where are we?

We are interested in **the effect of treatment X on outcome Y**:

- What is the effect of dieting on psychological well-being (Schafer & Kang, 2008)?
- What is the effect of out-of-home-placement on children's well-being (Berger et al., 2009)?
- What is the effect of physical punishment of children's behavioral problems (Larzelere et al., 2010)?
- What is the effect of extra schooling on social economic status?

If we have only **observational data** for this, we should:

- be concerned about **confounding**, collider bias, overcontrol bias
- **including the right covariates** to account for this
- we can **use DAGs** to see what we should control for (i.e., condition on)

But a DAG is of course only as good as our theory is...

What covariates should be included?

Steiner, Cook, Shadish and Clark (2010) used a creative design:

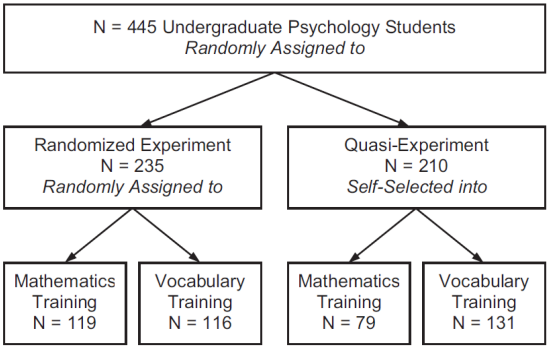


Figure 1. Overall design of the within-study comparison of the randomized experiment and the quasi-experiment.

This allows for a comparison of RCT results (true ACE) and observational results.

Results from Steiner er al. (2010)

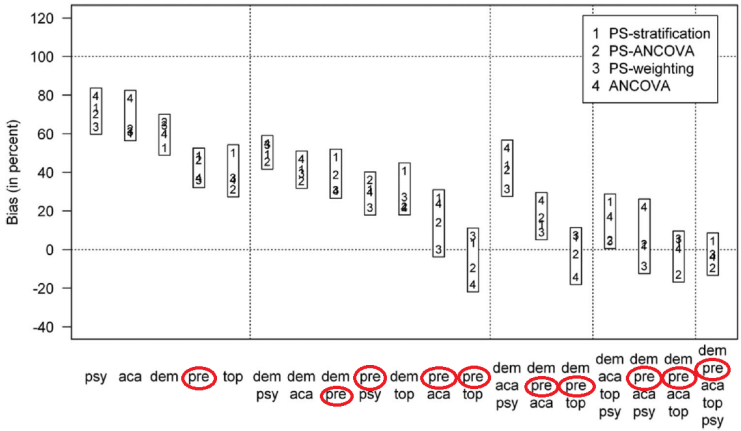


Figure 2. Remaining bias in vocabulary by construct set and analytic method (in the order of average bias). Construct sets are composed of demographics (dem), proxy-pretests (pre), prior academic achievement (aca), topic preference (top), and psychological predisposition (psy). PS = propensity score; ANCOVA = analysis of covariance.

Overall, pre-test measure seems a valuable covariate to include.

Overview

- DAGs & Interventions & the RCT
- Selection Bias, Berksons Paradox, Simpsons Paradox
- **Repeated Measures: Change Scores vs Controlling for Pre-measure**

Pre-Post Designs

Pre-post test designs: when the outcome is measured twice

- Lord's paradox
- ANCOVA vs. change score analysis
- Five scenarios
- How DAGs can help (Pearl, 2016)
- Unmeasured confounders (Kim & Steiner, 2019)

Lord's paradox

Comparing non-randomly assigned groups

Psychological Bulletin
1967, Vol. 68, No. 5, 304–305

A PARADOX IN THE INTERPRETATION OF GROUP COMPARISONS

FREDERIC M. LORD

Educational Testing Service

Attention is called to a basic source of confusion in the interpretation of certain types of group comparison data.

Comparing non-randomly assigned groups

Psychological Bulletin
1967, Vol. 68, No. 5, 304–305

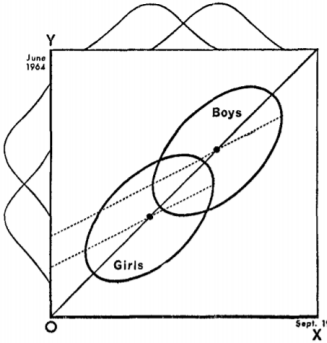
A PARADOX IN THE INTERPRETATION OF GROUP COMPARISONS

FREDERIC M. LORD

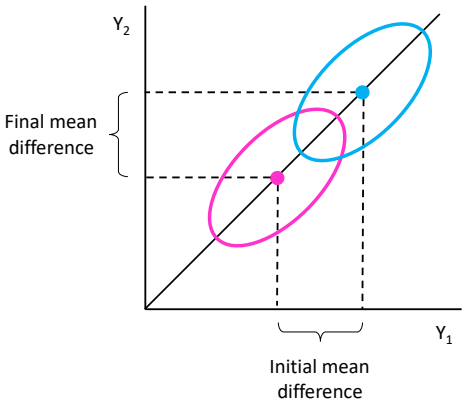
Educational Testing Service

Attention is called to a basic source of confusion in the interpretation of certain types of group comparison data.

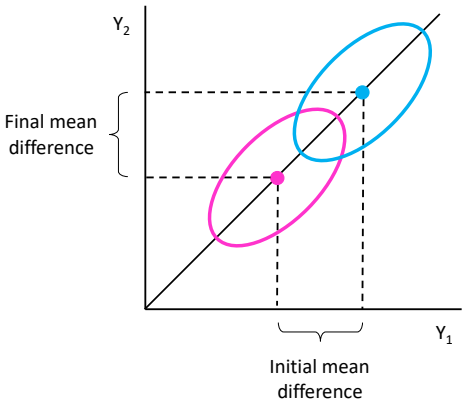
"A large university is interested in investigating the effects on the students of the diet provided in the university dining halls and any sex difference in these effects." (p.304)

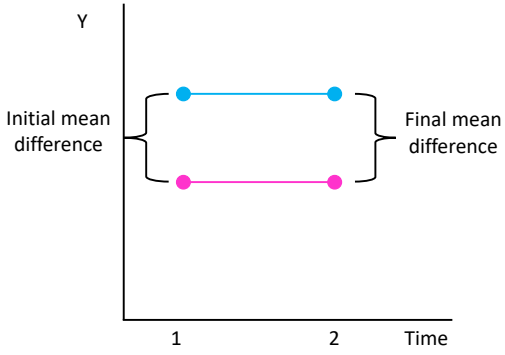


Statistician 1: Looks at the difference in the differences

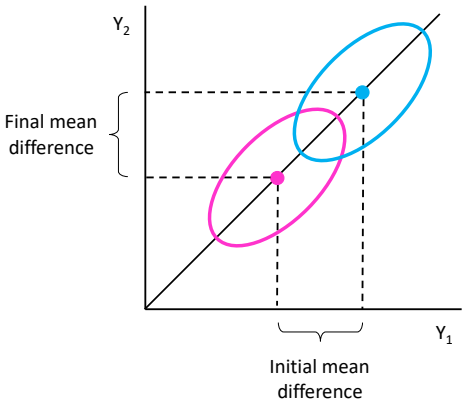


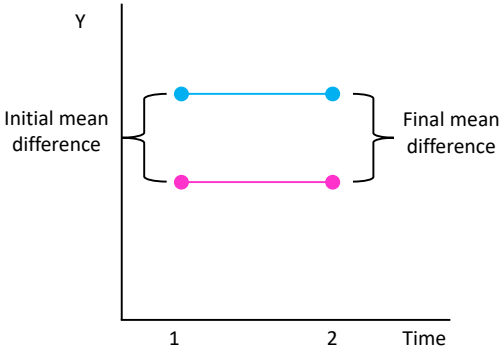
Statistician 1: Looks at the difference in the differences





Statistician 1: Looks at the difference in the differences

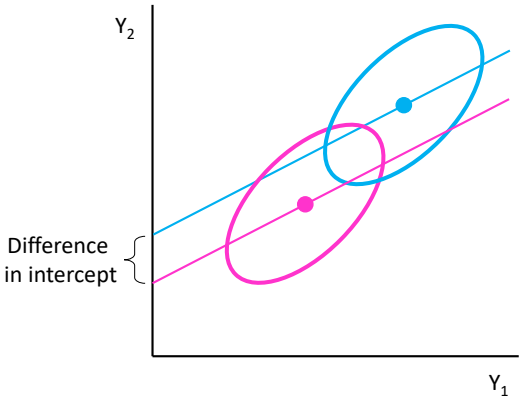


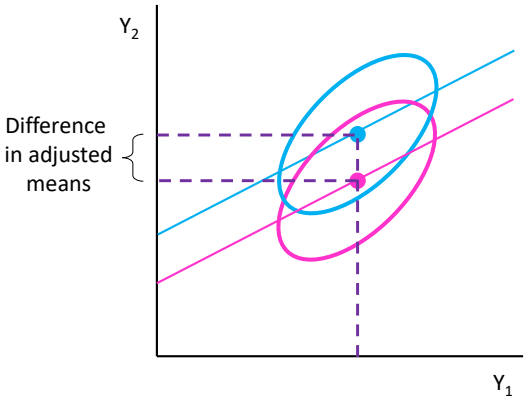


Conclusion: No change per group, so no difference in their change either.

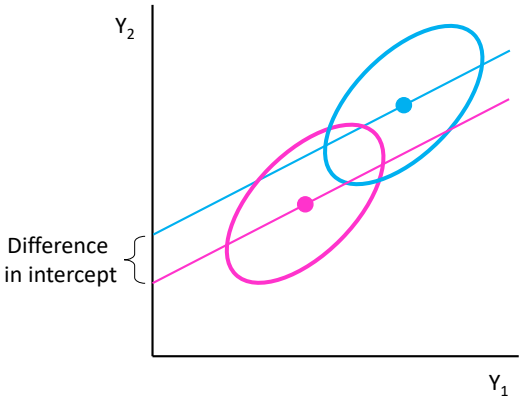
Or **more generally**: There is no difference over time in the differences between the groups.

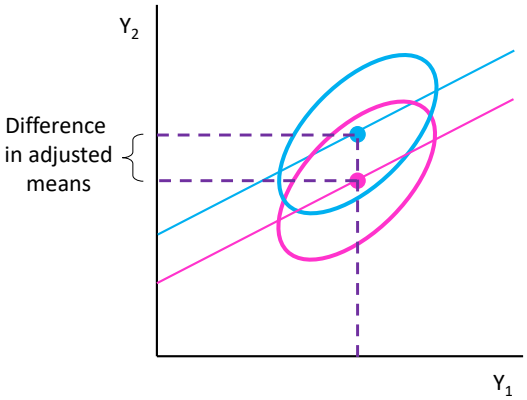
Statistician 2: Uses ANCOVA - Pre-weight as control variable





Statistician 2: Uses ANCOVA - Pre-weight as control variable





Conclusion: When comparing a boy and girl of **equal weight to begin with**, the boy **tends to weigh more afterwards** than the girl; hence, there is a difference, after correcting for initial differences.

Sex differences in effect of dining hall diet

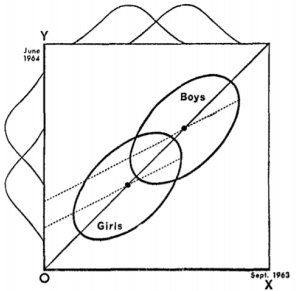


FIG. 1. Hypothetical scatterplots showing initial and final weight for boys and for girls.

Statistician 1: No difference Statistician 2: Boys gain more than girls

Lord's conclusion (p.305, 1967):

"The researcher wants to know how the groups would have compared if there had been no preexisting uncontrolled differences. The usual research study of this type is **attempting to answer a question that simply cannot be answered in any rigorous way on the basis of available data**."

Lord's paradox in empirical research

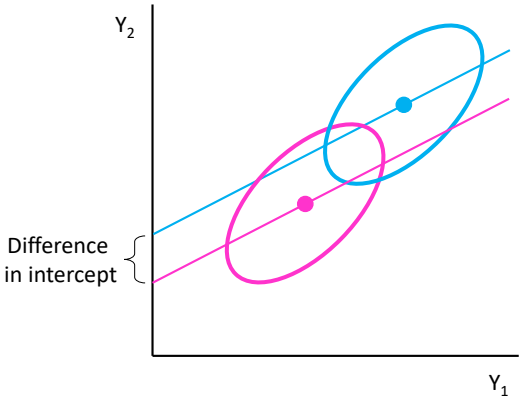
Larzelere et al. (2010) studied the effect of **corrective actions** on **problem behaviors** of 1,464 children aged 4 and 5.

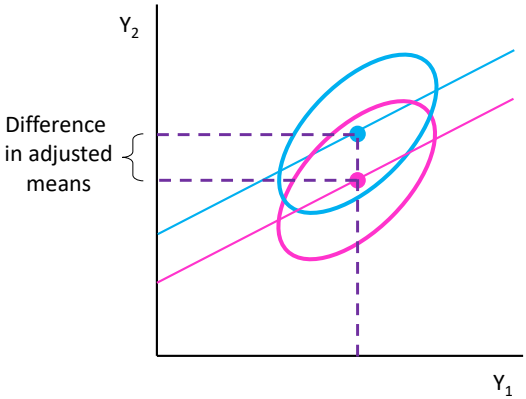
Corrective action	ANCOVA result	Change score result
	Antisocial behavior	
Professional interventions		
Psychotherapy visits	.07**	*.00*
Ritalin	.07**	*.04*
Parental disciplinary actions		
Non-physical punishment	.03	*−.08***
Physical punishment	.07**	*−.05*
Scolding/yelling	.06*	*−.08***
"Hostile/ineffective" scale	.09**	*−.15***
	Hyperactivity	
Professional interventions		
Psychotherapy visits	.03	*−.02*
Ritalin	.05*	*−.00*
Parental disciplinary actions		
Non-physical punishment	.07**	*−.01*
Physical punishment	.03	*−.01*
Scolding/yelling	.04*	*−.05*
"Hostile/ineffective" scale	.09**	*−.08***

ANCOVA vs. Change score model

ANCOVA (popular in psychology)

Analysis of covariance (ANCOVA) in a pre-posttest design is based on including the pretest as covariate.





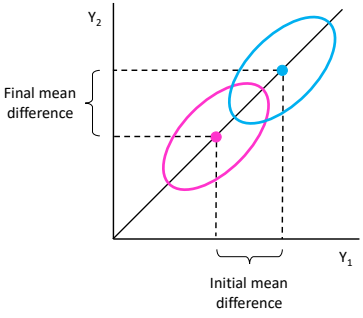
ACE based on ANCOVA model

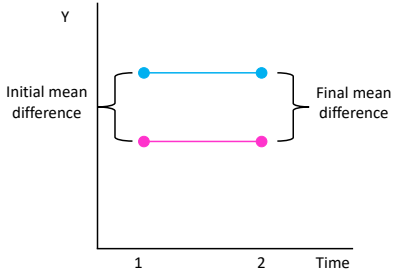
The **causal effect** β_1 in the ANCOVA model is the **difference between treated and untreated** persons **with identical values on covariate** (i.e. the pretest). We can write this as the **expected difference in potential outcomes**:

$$
\begin{aligned}
ACE_{ANCOVA} &= E\big[Y_2^1\big] - E\big[Y_2^0\big] \\
&= E\big[Y_2^1 | X = 1, Y_1\big] - E\big[Y_2^0 | X = 0, Y_1\big] && \text{No unobserved confounding} \\
&= E\big[Y_2 | X = 1, Y_1\big] - E\big[Y_2 | X = 0, Y_1\big] && \text{Consistency} \\
&= \big(\beta_0 + \beta_1 + \beta_2 Y_1\big) - \big(\beta_0 + \beta_2 Y_1\big) && \text{Correct model} \\
&= \beta_1
\end{aligned}
$$

Change score model (popular in econometrics)

The change score model (or gain score model) is based on investigating difference-in-differences.





Change score model:

$$Y_{2i} - Y_{1i} = \gamma_0 + \gamma_1 X_i + \epsilon_{2i}$$

where γ_1 is interpreted as the causal effect of X.

ACE based on Change score model

Let $G_i = Y_{2i} - Y_{1i}$ represent a person's **gain score** (aka change or difference score).

Then the **average causal effect** can be expressed as the expected difference in potential outcomes of G_i:

$$
\begin{aligned}
ACE_{CS} &= E\big[G^1\big] - E\big[G^0\big] && \\
&= E\big[G^1|X=1\big] - E\big[G^0|X=0\big] && \text{No unobserved confounding} \\
&= E\big[G|X=1\big] - E\big[G|X=0\big] && \text{Consistency} \\
&= E\big[\gamma_0 + \gamma_1\big] - E\big[\gamma_0\big] && \text{Correct model} \\
&= \gamma_1 &&
\end{aligned}
$$

Alternative expression of ACE_{CS}

Instead of expressing the **ACE** of the changes score model in terms of the regression parameters, we can also express it in terms of the **pre- and post-test means**, that is:

$$
\begin{aligned}
ACE_{CS} \quad &= E\big[G^1\big] - E\big[G^0\big] \\
&= E\big[G^1|X=1\big] - E\big[G^0|X=0\big] && \text{No unobserved confounding} \\
&= E\big[G|X=1\big] - E\big[G|X=0\big] && \text{Consistency} \\
&= E\big[\{Y_2 - Y_1\}|X=1\big] - E\big[\{Y_2 - Y_1\}|X=0\big] \\
&= \big(E\big[Y_2|X=1\big] - E\big[Y_1|X=1\big]\big) - \big(E\big[Y_2|X=0\big] - E\big[Y_1|X=0\big]\big)
\end{aligned}
$$

ACE_{CS} as difference-in-differences

Thus we have

$$ACE_{CS} = \big(E\big[Y_2|X=1\big] - E\big[Y_1|X=1\big]\big) - \big(E\big[Y_2|X=0\big] - E\big[Y_1|X=0\big]\big)$$

that is, the **ACE is equal to the difference between groups in their gain scores**.

Alternatively, we can write

$$ACE_{CS} = \big(E\big[Y_2|X=1\big] - \big(E\big[Y_2|X=0\big]\big) - \big(E\big[Y_1|X=1\big] - E\big[Y_1|X=0\big]\big)$$

that is, the ACE is equal to the difference over time in the difference between the groups (**difference-in-differences**).

Conclusion so far

With the **ANCOVA model** we answer the question: If the groups had been equal on the pre-test, would we observe a difference between them on the post-test?

$Y_{2i} = \beta_0 + \beta_1 X_i + \beta_2 Y_{1i} + e_{2i}$ If so (i.e., $\beta_1 \neq 0$), we conclude **treatment has an effect**.

With the **CS model**, we answer the question: Is the change over time different for the two groups?

$Y_{2i} - Y_{1i} = \gamma_0 + \gamma_1 X_i + \epsilon_i$ If so (i.e., $\gamma_1 \neq 0$), we conclude **treatment has an effect**.

Conclusion so far

With the **ANCOVA model** we answer the question: If the groups had been equal on the pre-test, would we observe a difference between them on the post-test?

$Y_{2i} = \beta_0 + \beta_1 X_i + \beta_2 Y_{1i} + e_{2i}$ If so (i.e., $\beta_1 \neq 0$), we conclude **treatment has an effect**.

With the **CS model**, we answer the question: Is the change over time different for the two groups?

$Y_{2i} - Y_{1i} = \gamma_0 + \gamma_1 X_i + \epsilon_i$ If so (i.e., $\gamma_1 \neq 0$), we conclude **treatment has an effect**.

Yet, these questions can lead to **different answers**.

Question: Would it help if we could decide whether we are interested in Y_{2i} or $G_i = Y_{21} - Y_{1i}$ as the outcome?

DIY

We have **two models**:

- ANCOVA: $Y_{2i} = \beta_0 + \beta_1 X_i + \beta_2 Y_{1i} + e_{2i}$
- CSM: $Y_{2i} - Y_{1i} = \gamma_0 + \gamma_1 X_i + \epsilon_i$

1) **Rewrite** the ANCOVA model as a changes score model (i.e., with G_i as the outcome); what does this tell you?

2) **Rewrite** CSM as ANCOVA model (i.e., with Y_{2i} as the outcome); what does this tell you?

DIY

We have **two models**:

- ANCOVA: $Y_{2i} = \beta_0 + \beta_1 X_i + \beta_2 Y_{1i} + e_{2i}$
- CSM: $Y_{2i} - Y_{1i} = \gamma_0 + \gamma_1 X_i + \epsilon_i$

1) **Rewrite** the ANCOVA model as a changes score model (i.e., with G_i as the outcome); what does this tell you?

2) **Rewrite** CSM as ANCOVA model (i.e., with Y_{2i} as the outcome); what does this tell you?

1) ANCOVA as CSM: $Y_{2i} - Y_{1i} = \beta_0 + \beta_1 X_i + (\beta_2 - 1) Y_{1i} + e_{2i}$

DIY

We have **two models**:

- ANCOVA: $Y_{2i} = \beta_0 + \beta_1 X_i + \beta_2 Y_{1i} + e_{2i}$
- CSM: $Y_{2i} - Y_{1i} = \gamma_0 + \gamma_1 X_i + \epsilon_i$

1) **Rewrite** the ANCOVA model as a changes score model (i.e., with G_i as the outcome); what does this tell you?

2) **Rewrite** CSM as ANCOVA model (i.e., with Y_{2i} as the outcome); what does this tell you?

1) ANCOVA as CSM: $Y_{2i} - Y_{1i} = \beta_0 + \beta_1 X_i + (\beta_2 - 1) Y_{1i} + e_{2i}$

It shows that doing an ANCOVA controlling for pre-test with either Y_{2i} or $G_i = Y_{2i} - Y_{1i}$ as the outcome leads to the same effect of X_i.

DIY

We have **two models**:

- ANCOVA: $Y_{2i} = \beta_0 + \beta_1 X_i + \beta_2 Y_{1i} + e_{2i}$
- CSM: $Y_{2i} - Y_{1i} = \gamma_0 + \gamma_1 X_i + \epsilon_i$

1) **Rewrite** the ANCOVA model as a changes score model (i.e., with G_i as the outcome); what does this tell you?

2) **Rewrite** CSM as ANCOVA model (i.e., with Y_{2i} as the outcome); what does this tell you?

1) ANCOVA as CSM: $Y_{2i} - Y_{1i} = \beta_0 + \beta_1 X_i + (\beta_2 - 1) Y_{1i} + e_{2i}$

It shows that doing an ANCOVA controlling for pre-test with either Y_{2i} or $G_i = Y_{2i} - Y_{1i}$ as the outcome leads to the same effect of X_i.

2) CSM as ANCOVA: $Y_{2i} = \gamma_0 + \gamma_1 X_i + Y_{1i} + \epsilon_i$

DIY

We have **two models**:

- ANCOVA: $Y_{2i} = \beta_0 + \beta_1 X_i + \beta_2 Y_{1i} + e_{2i}$
- CSM: $Y_{2i} - Y_{1i} = \gamma_0 + \gamma_1 X_i + \epsilon_i$

1) **Rewrite** the ANCOVA model as a changes score model (i.e., with G_i as the outcome); what does this tell you?

2) **Rewrite** CSM as ANCOVA model (i.e., with Y_{2i} as the outcome); what does this tell you?

1) ANCOVA as CSM: $Y_{2i} - Y_{1i} = \beta_0 + \beta_1 X_i + (\beta_2 - 1) Y_{1i} + e_{2i}$

It shows that doing an ANCOVA controlling for pre-test with either Y_{2i} or $G_i = Y_{2i} - Y_{1i}$ as the outcome leads to the same effect of X_i.

2) CSM as ANCOVA: $Y_{2i} = \gamma_0 + \gamma_1 X_i + Y_{1i} + \epsilon_i$ This shows that the CSM can be considered a special case of ANCOVA (with $\beta_2 = 1$).

DIY: When are the ACEs the same?

When will the ANCOVA model and the change score model give the same ACE?

$$ACE_{CS} = \big(E\big[Y_2|X=1\big] - \big(E\big[Y_2|X=0\big]\big) - \big(E\big[Y_1|X=1\big] - E\big[Y_1|X=0\big]\big)$$

$ACE_{ANCOVA} = \beta_1$
$ACE_{CS} = \beta_1 + (\beta_2 - 1)\big(E\big[Y_1|X=1\big] - E\big[Y_1|X=0\big]\big)$

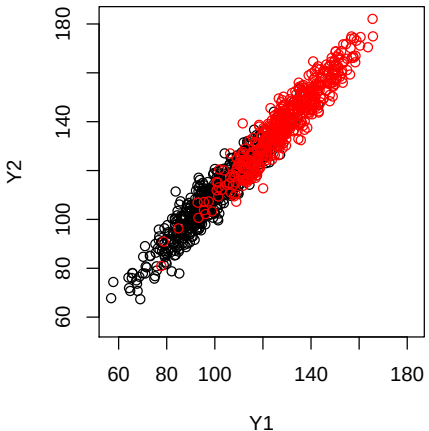
ANSWER: These are **identical when** $(\beta_2 - 1)\big(E\big[Y_1|X=1\big] - E\big[Y_1|X=0\big]\big)$

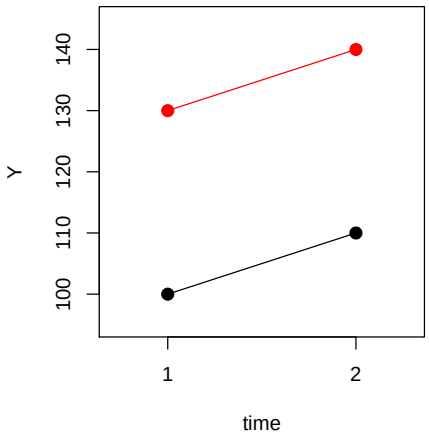
This is the case when either:

- $\beta_2 = 1$: the **effect of the pretest on posttest within groups is 1** (as you had already found when rewriting a CSM as an ANCOVA model); or
- $\big(E\big[Y_1|X=1\big] - E\big[Y_1|X=0\big]\big)$: there are **no initial group differences** (as in an RCT!)

Five scenarios

Scenario 1: No causal effect

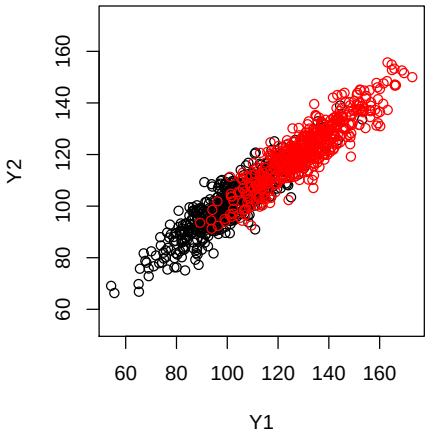


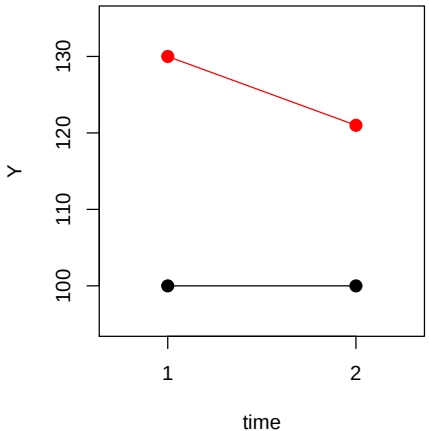


ANCOVA: $\beta_1 = 0$ so no causal effect

CSM: $\gamma_1 = 0$ so no causal effect

Scenario 2: CS model negative causal effect

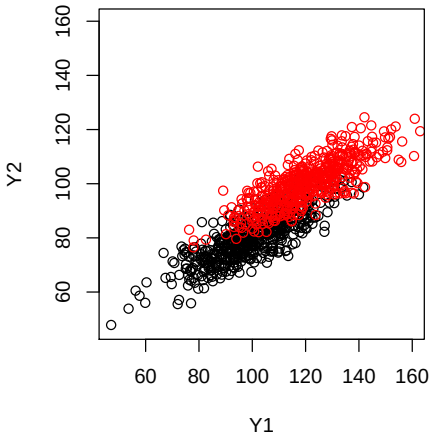


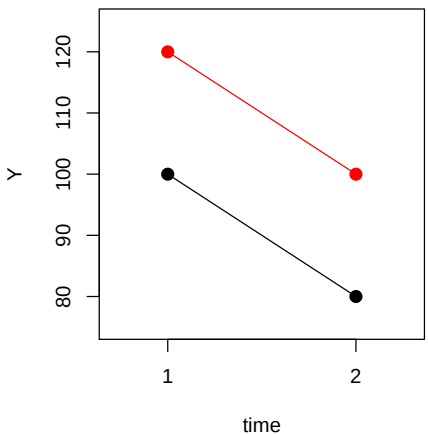


ANCOVA: $\beta_1 = 0$ so no causal effect

CSM: $\gamma_1 < 0$ so a (negative) causal effect

Scenario 3: ANCOVA model positive causal effect

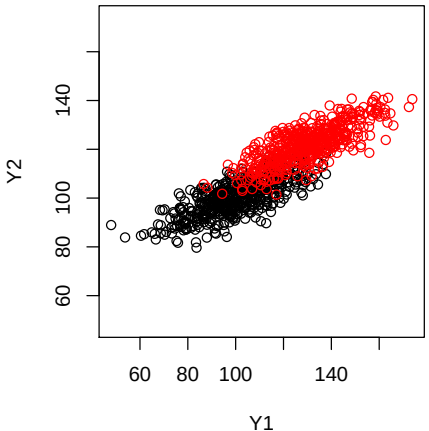


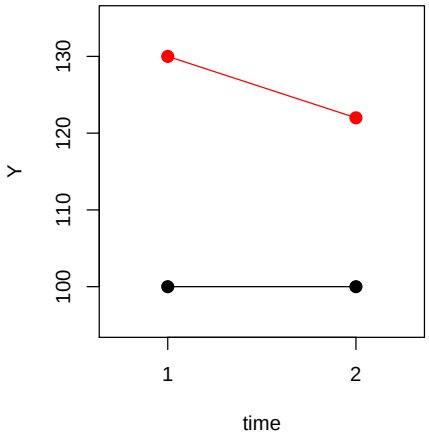


ANCOVA: $\beta_1 > 0$ so a (positive) causal effect

CSM: $\gamma_1 = 0$ so no causal effect

Scenario 4: Opposite conclusions regarding direction!

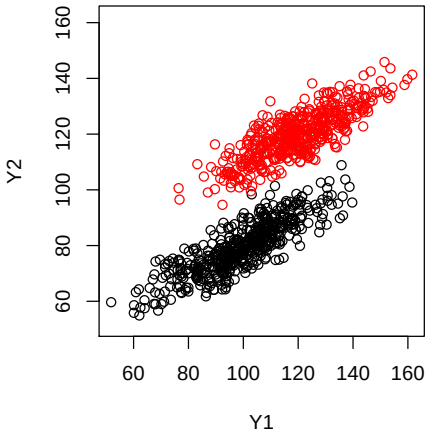


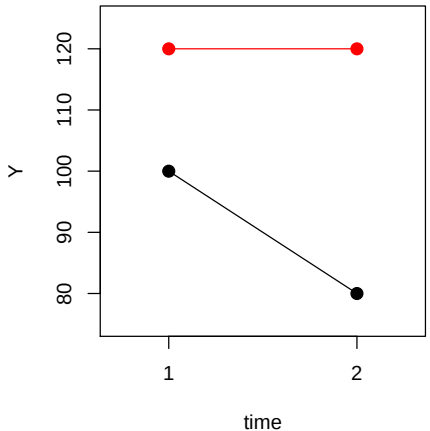


ANCOVA: $\beta_1 > 0$ so a (positive) causal effect

CSM: $\gamma_1 < 0$ so a (negative) causal effect

Scenario 5: Some agreement





ANCOVA: $\beta_1 > 0$ so a (positive) causal effect

CSM: $\gamma_1 > 0$ so a (positive) causal effect

So what now?

It can be shown that only when $\beta_2 = 1$ and/or $\mu_{1|1} = \mu_{1|0}$, are β_1 (ACE_{ANCOVA}) and γ_1 (ACE_{CSM}) identical.

Allison (p.109, 1990):

"It is unrealistic to expect either model to be best in all situations; [...] the choice will rarely be obvious, and there will almost always be some residual uncertainty. One should also consider the possibility that neither of these models is appropriate [...]."

Allison (p.100, 1990):

"A problem with much of the work comparing change score and regressor variable methods is that the conclusions are rarely based on an explicit model for generation of the data."

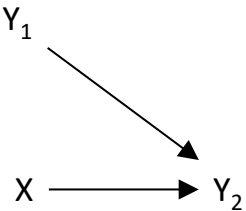
Draw your DAG!

How DAGs can help

Should we control for pretest? - Y2 as outcome

The key issue is whether **assignment is**:

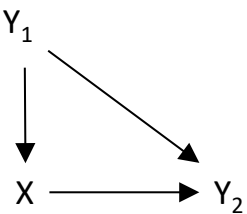
Random



ANCOVA is preferred,

it has more power

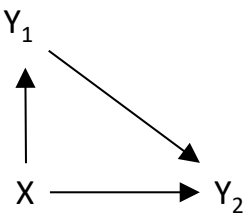
Based on pretest



ANCOVA is correct;

pretest is a **confounder**

Existing groups



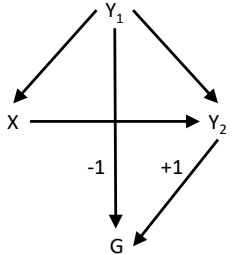
ANCOVA gives direct effect

pretest is a **mediator**

Controlling for Pre-Treatment - change score

Change/Gain score: $G_i = Y_{2i} - Y_{1i}$

Pretest as confounder

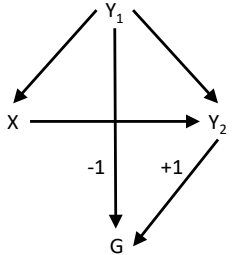


Need to block backdoor paths:
$X \leftarrow Y1 \rightarrow G$
$X \leftarrow Y1 \rightarrow Y2 \rightarrow G$
Control for pre-treatment (use ANCOVA) to get
$X \rightarrow Y2 \rightarrow G$

Controlling for Pre-Treatment - change score

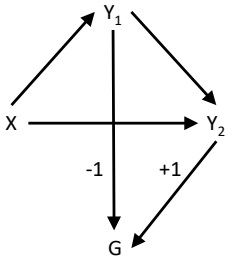
Change/Gain score: $G_i = Y_{2i} - Y_{1i}$

Pretest as confounder



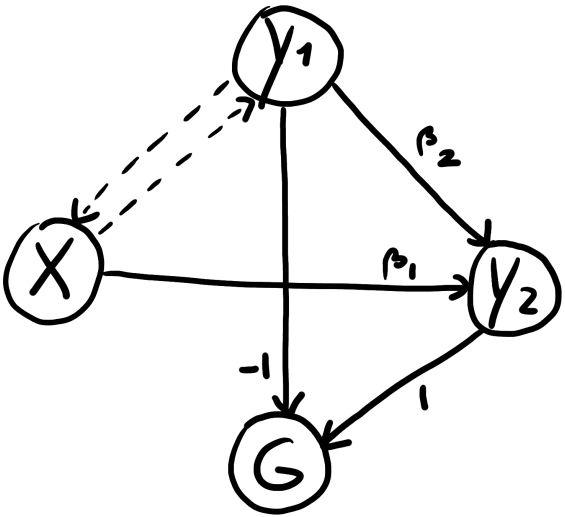
Need to block backdoor paths:
$X \leftarrow Y1 \rightarrow G$
$X \leftarrow Y1 \rightarrow Y2 \rightarrow G$
Control for pre-treatment (use ANCOVA) to get
$X \rightarrow Y2 \rightarrow G$

Pretest as mediator



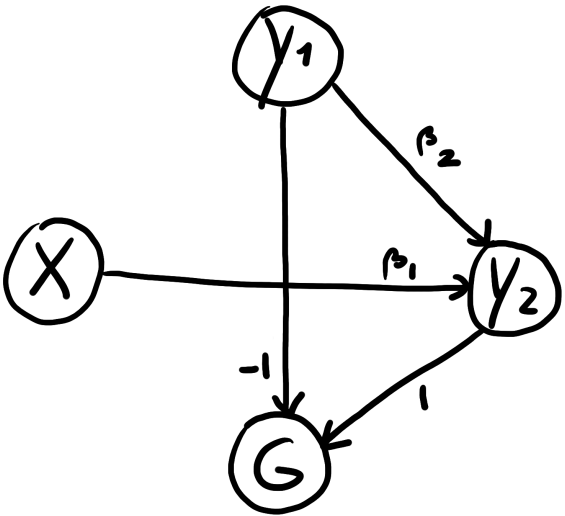
Total effect consists of:
$X \rightarrow Y2 \rightarrow G$
$X \rightarrow Y1 \rightarrow G$
$X \rightarrow Y1 \rightarrow Y2 \rightarrow G$
Do not control for pre-treatment (classical change score model) to get total effect; control for pre-treatment (ANCOVA) to get direct effect.

Revisit: When classical Change Score and ANCOVA model have the same results

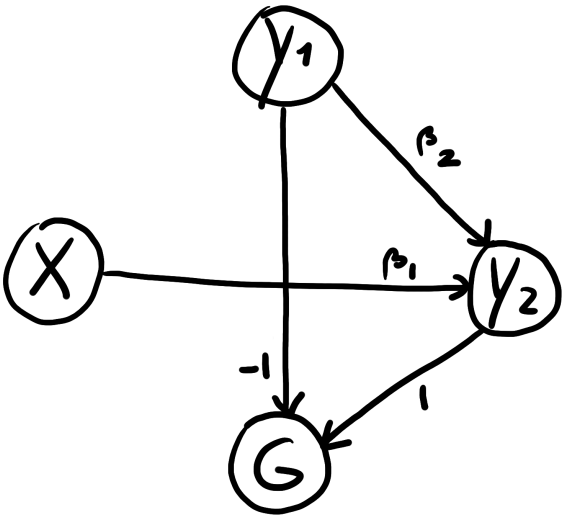


- When the expected values for Y1 are the same in each group (no group differences in Y1)
- When the effect of Y1 on Y2 is equal to 1

Revisit: When classical Change Score and ANCOVA model have the same results: EVs Y1 the same in each group

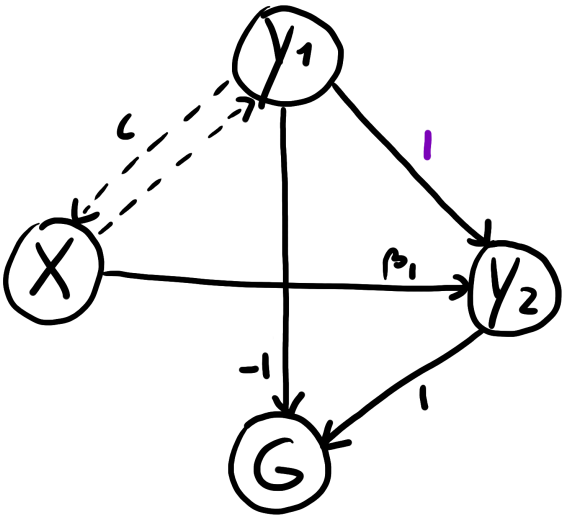


Revisit: When classical Change Score and ANCOVA model have the same results: EVs Y1 the same in each group

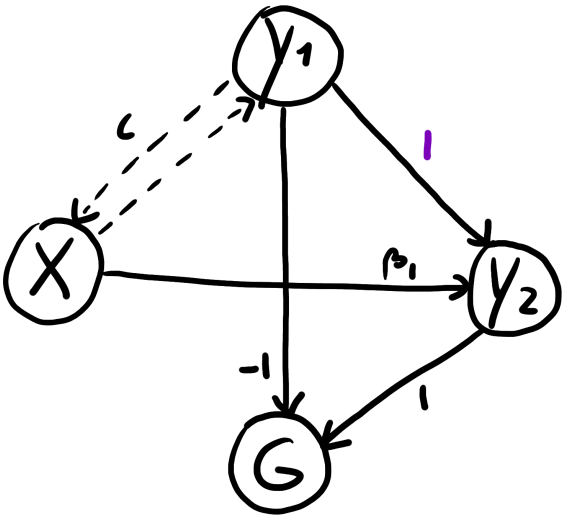


CS (do not control for Y1)	ANCOVA Y2 (control for Y1)
G ~ X	Y2 ~ X + Y1
paths X - G	paths X - Y2
$\beta_1 \times 1$	β_1

Revisit: When classical Change Score and ANCOVA model have the same results: Effect of Y1 on Y2 equal to 1

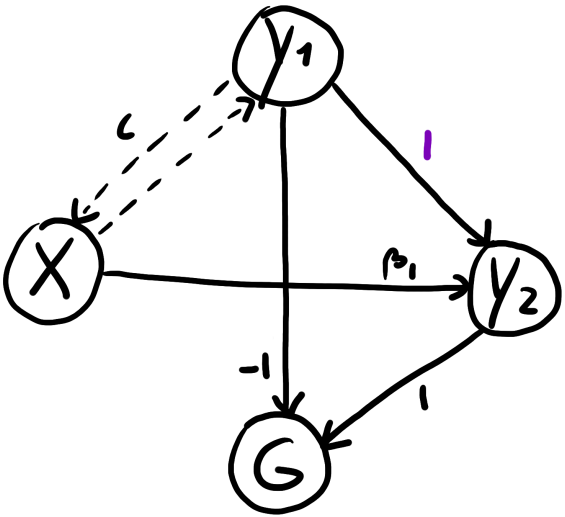


Revisit: When classical Change Score and ANCOVA model have the same results: Effect of Y1 on Y2 equal to 1



CS (do not control for y_1)	ANCOVA y_2 (control for y_1)
$G \sim X$	$y_2 \sim X + y_1$
paths X - G	paths x - y_2
$X \leftrightarrow y_1 \xrightarrow{1} y_2 \xrightarrow{1} G$	$X \leftrightarrow y_1 \xrightarrow{1} y_2$
$X \leftrightarrow y_1 \xrightarrow{-1} G$	$X \leftrightarrow y_1 \xrightarrow{-1} G \xleftarrow{1} y_2$
$X \xrightarrow{\beta_1} y_2 \xrightarrow{1} G$	$X \xrightarrow{\beta_1} y_2$

Revisit: When classical Change Score and ANCOVA model have the same results: Effect of Y1 on Y2 equal to 1



CS (do not control for y_1)	ANCOVA y_2 (control for y_1)
$G \sim X$	$y_2 \sim X + y_1$
paths X – G	paths X – y_2
$X \leftrightarrow y_1 \xrightarrow{1} y_2 \xrightarrow{1} G$	blocked; $X \leftrightarrow y_1 \not\rightarrow y_2$
$X \leftrightarrow y_1 \xrightarrow{1} G$	$X \leftrightarrow y_1 \not\rightarrow G \not\leftarrow y_2$
$X \xrightarrow{\beta_1} y_2 \xrightarrow{1} G$ =	$X \xrightarrow{\beta_1} y_2$
	blocked

Timing is critical

The DAGs show that the **causal relation** between treatment X and pre-test Y1 is critical; it is about whether X or Y1 came first (i.e., their temporal order).

In Rubin's causal framework (Week 2), the timing of treatment, outcome and covariates is also considered critical.

Holland (1986):

- exposure to a cause (i.e., treatment) occurs at a **specific time point or time interval**
- variables are thus divided into **pre-exposure** and **post-exposure**
- "The role of a response variable Y is to measure the effect of the cause, and thus **response variables** must fall into the **post-exposure class**." (p.946, Holland, 1986)
- **covariates** should come from the **pre-exposure phase**; then they cannot be affected by the treatment.

Critique of using change scores in DAGs

Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

Causal diagrams and change variables

Eyal Shahar MD MPH[1] and Doron J. Shahar[2]

Abstract

Background The true change in the value of a variable between two time points is often assumed to be a cause or an effect of interest. To our knowledge, this assumption is based on intuition, rather than on any formal theoretical justification.

Methods We used causal directed acyclic graphs to explore the causal properties of a change variable, and critically examined competing structures.

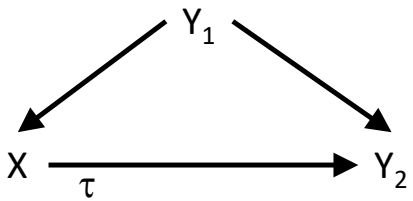
Results Based on the proposed causal structure, a change variable (true change) is no more than a derived variable. It does not cause anything and is not of causal interest.

Conclusions A true change is not a variable in the physical world. Therefore, modelling the change between two time points is justified only in a few situations.

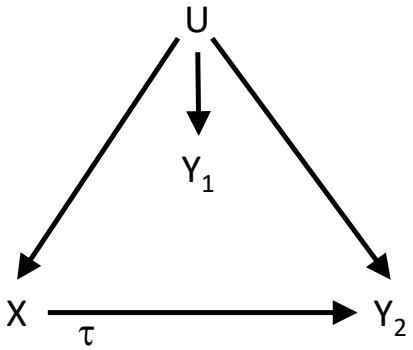
Unmeasured confounding in the pre-post test design

Pretest as a proxy for confounder

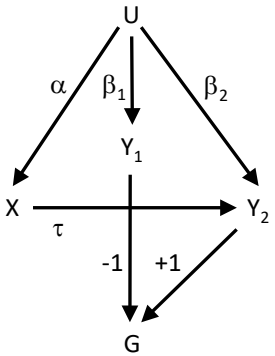
When Y_1 was **measured prior to treatment**, it could be a **confounder**; you need to control for it then (e.g., use ANCOVA model).



But Y_1 may also be a **proxy of an unobserved confounder**; controlling for Y_1 will only **partly remove bias** due to $X \leftarrow U \rightarrow Y_2$.



How can classical change score analysis help?



The interest is the causal path $X \to Y2 \to G$. This is the causal effect from X to G ($\tau \times 1$), but also the causal effect of X on Y2 (just τ).

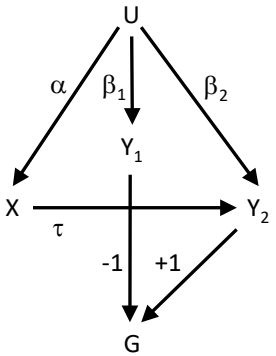
Backdoor paths between X and G:

- $X \leftarrow U \to Y1 \to G$
- $X \leftarrow U \to Y2 \to G$
- $X \to Y2 \leftarrow U \to Y1 \to G$

When we have **linear relations** (and the variance of U is equal to 1), we get:

- First path: $-\alpha\beta_1$
- Second path: $\alpha\beta_2$
- Third path: 0 (blocked because it contains the **collider** $Y2$)

How can classical change score analysis help?



The interest is the causal path $X \to Y2 \to G$. This is the causal effect from X to G ($\tau \times 1$), but also the causal effect of X on Y2 (just τ).

Backdoor paths between X and G:

- $X \leftarrow U \to Y1 \to G$
- $X \leftarrow U \to Y2 \to G$
- $X \to Y2 \leftarrow U \to Y1 \to G$

When we have **linear relations** (and the variance of U is equal to 1), we get:

- First path: $-\alpha\beta_1$
- Second path: $\alpha\beta_2$
- Third path: 0 (blocked because it contains the **collider** $Y2$)

If $\beta_1 = \beta_2$, the first and second path backdoor **cancel each other out** (cf. Kim & Steiner, 2019)!

Important to realize (i.e., conclusion so far)

We have written the change score model as a **special case** of the ANCOVA model.

This may suggest we should just test **which model fits better**.

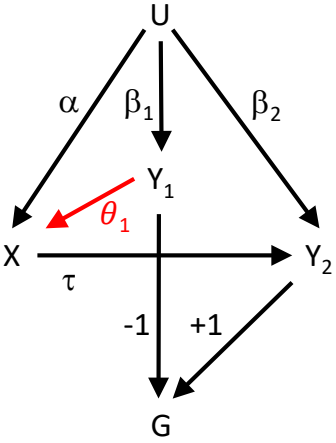
However, the point is **NOT** that we want to determine which of these two models generated the data!

The goal is to **estimate the treatment effect without bias**.

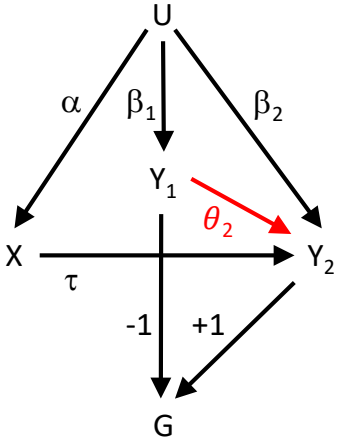
The **change score model** is:

- typically NOT considered a reasonable model as a data generating mechanism
- but a very useful model for estimating the causal effect (under specific circumstances)

What if...

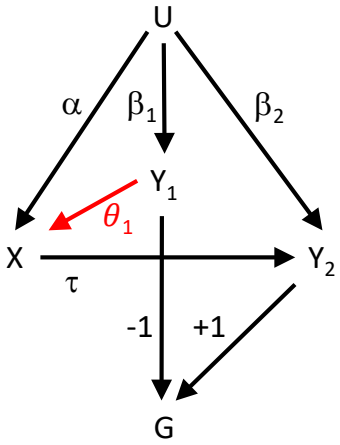


Pretest affects treatment

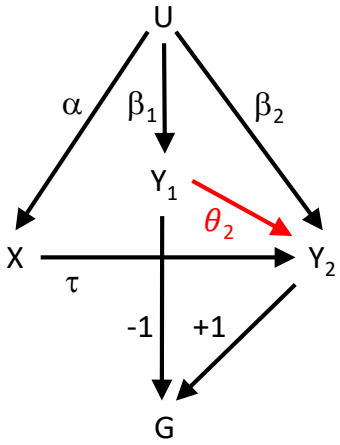


Pretest affects outcome

What if...



Pretest affects treatment

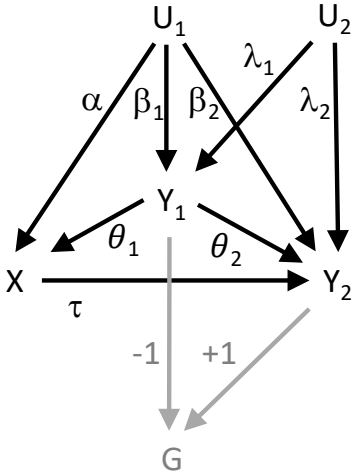


Pretest affects outcome

In these scenarios neither the Change score model nor the ANCOVA model give an unbiased estimate of the causal effect of X on Y2.

Also important to realize (wrt timing)

When pretest is from the **pretreatment phase**, it does **NOT** mean it can **only** be a confounder.



The **pretest Y1** is:

- a **confounder**: $X \leftarrow Y1 \rightarrow Y2$
- a **collider**: $X \leftarrow U1 \rightarrow Y1 \leftarrow U2 \rightarrow Y2$

Summary

ANCOVA (regress Y2 or G=Y2-Y1 on X and Y1):

- when Y1 is confounder of X and Y2
- or to get direct effect when Y1 is mediator

Marginal model (regress Y2 on X):

- when Y1 is mediator and interest is in total effect of X on Y2

Change score model (regress G=Y2-Y1 on X):

- when there is time-invariant unobserved confounding with stable effect
- (or when Y1 is mediator and the interest is in total effect of X on Y2-Y1)

Summary

ANCOVA (regress Y2 or G=Y2-Y1 on X and Y1):

- when Y1 is confounder of X and Y2
- or to get direct effect when Y1 is mediator

Marginal model (regress Y2 on X):

- when Y1 is mediator and interest is in total effect of X on Y2

Change score model (regress G=Y2-Y1 on X):

- when there is time-invariant unobserved confounding with stable effect
- (or when Y1 is mediator and the interest is in total effect of X on Y2-Y1)

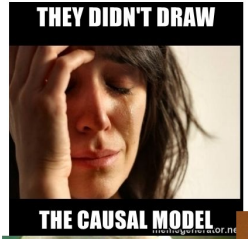
There is a lot more to say and study about causality and time

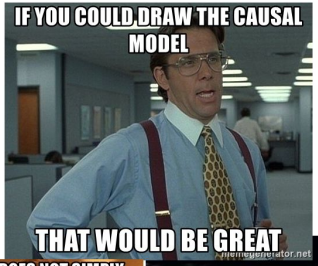
- Consider time-varying treatments, outcomes, and covariates
- Variables affecting themselves and each other continuously through time, effects may change over time, ...
- In (these) more complicated scenarios we may have variables that are simultaneously confounders/mediators/colliders.

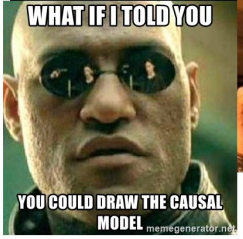
In any case...

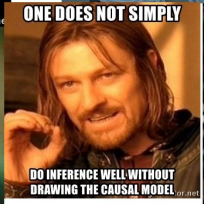
Do causal inference in a principled way!

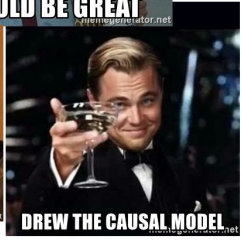
- Be explicit and clear about your causal interests/questions
- Specify your ideas (causal theory) in some causal graph and/or in equations (and SCM)..
- make assumptions explicit
- Choose a causal analysis best tailored to your particular problem.
- Replicate, triangulate, critique, etc!











In any case...

Finally, remember these science key three...

- Measurement
- Theory formation
- Causal Inference