Causal inference and confounding effects

Yongjin Park
University of British Columbia

13 March, 2022

Learning objectives

- Introduction to causal inference
- Causal structural model (DAG)
- Confounding factor adjustment
- Potential outcome model

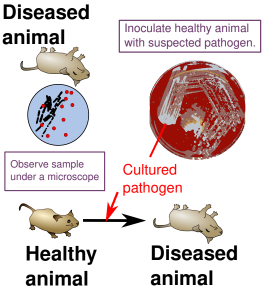
A century-old question, finding the cause of a disease

Henle-Koch Postulates

1. The parasite occurs in every case of the disease in question and under circumstances which can account for the pathological changes and clinical course of the disease.
2. It occurs in no other disease as a fortuitous and nonpathogenic parasite.
3. After being fully isolated from the body and repeatedly grown in pure culture, it can induce the disease anew.

Jakob Henle

Robert Koch



source: wikipedia

Evans, Causation and Disease (1976)

Scientists have been doing causal inference research

Henle-Koch Postulates

1. The parasite occurs in every case of the disease in question and under circumstances which can account for the pathological changes and clinical course of the disease.
2. It occurs in **no other disease** as a fortuitous and nonpathogenic parasite.
3. After being *fully isolated* from the body and *repeatedly grown* in pure culture, it can induce the disease anew.

Evans, Causation and Disease (1976)

Scientists have been doing causal inference research

Henle-Koch Postulates

1. The parasite occurs in **every case of the disease** in question and under circumstances which can account for the pathological changes and clinical course of the disease.
2. It occurs in **no other disease** as a fortuitous and nonpathogenic parasite.
3. After being *fully isolated* from the body and *repeatedly grown* in pure culture, it can induce the disease anew.

- **Invariance**,

Evans, Causation and Disease (1976)

Scientists have been doing causal inference research

Henle-Koch Postulates

1. The parasite occurs in **every case of the disease** in question and under circumstances which can account for the pathological changes and clinical course of the disease.
2. **It occurs in no other disease** as a fortuitous and nonpathogenic parasite.
3. After being *fully isolated* from the body and *repeatedly grown* in pure culture, it can induce the disease anew.

- **Invariance**, **stability** arguments in the 1st and 2nd points

Evans, Causation and Disease (1976)

Scientists have been doing causal inference research

Henle-Koch Postulates

1. The parasite occurs in **every case of the disease** in question and under circumstances which can account for the pathological changes and clinical course of the disease.
2. **It occurs in no other disease** as a fortuitous and nonpathogenic parasite.
3. After being *fully isolated* from the body and *repeatedly grown* in pure culture, **it can induce the disease anew.**

- **Invariance**, **stability** arguments in the 1st and 2nd points
- The notion of **intervention** in the 3rd point.

Evans, Causation and Disease (1976)

Scientists have been doing causal inference research

Henle-Koch Postulates

1. The parasite occurs in **every case of the disease** in question and under circumstances which can account for the pathological changes and clinical course of the disease.
2. **It occurs in no other disease** as a fortuitous and nonpathogenic parasite.
3. After being *fully isolated* from the body and *repeatedly grown* in pure culture,

- **Invariance**, **stability** arguments in the 1st and 2nd points
- The notion of **intervention** in the 3rd point.
- Just oberving a pathogen in the diseased animals isn't credible evidence of the cause of disease

Evans, Causation and Disease (1976)

Today's lecture

What is causal inference (fundamentals)

Graphical language for causal inference

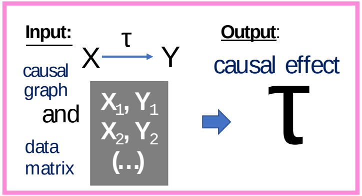
Mendelian Randomization

Potential outcome framework (what if?)

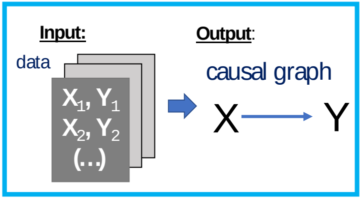
Model-based causal inference

When we don't know an underlying causal graph

Causal Effect Inference



Causal structure discovery



Causal "effect" inference in two ways: $X \to Y$

Experiments/interventions

What would be the effect of *doing* $X = x$?

$$P(Y|do(X = x))$$

Causal "effect" inference in two ways: $X \to Y$

Experiments/interventions

What would be the effect of *doing* $X = x$?

$$P(Y|do(X = x))$$

Although we are only ”seeing” $X = x$, can we still estimate the ”doing” effect?

$$P(Y|do(X = x)) \overset{???}{=} P(Y|X = x)$$

Causal "effect" inference in two ways: $X \to Y$

Experiments/interventions

What would be the effect of *doing* $X = x$?

$$P(Y|do(X = x))$$

Although we are only "seeing" $X = x$, can we still estimate the "doing" effect?

$$P(Y|do(X = x)) \overset{???}{=} P(Y|X = x)$$

Judea Pearl

Sewall Wright

R.A. Fisher

Causal "effect" inference in two ways: $X \to Y$

Experiments/interventions

What would be the effect of *doing* $X = x$?

$$P(Y|do(X = x))$$

Counterfactual reasoning

What if we had $X = 0$, not $X = 1$, or vice versa?

$$Y_i^{(X=1)} - Y_i^{(X=0)}$$

Although we are only "seeing" $X = x$, can we still estimate the "doing" effect?

$$P(Y|do(X = x)) \stackrel{???}{=} P(Y|X = x)$$


Judea Pearl


Sewall Wright


R.A. Fisher

Causal "effect" inference in two ways: $X \to Y$

Experiments/interventions

What would be the effect of *doing* $X = x$?

$$P(Y|do(X = x))$$

Although we are only "seeing" $X = x$, can we still estimate the "doing" effect?

$$P(Y|do(X = x)) \overset{???}{=} P(Y|X = x)$$


Judea Pearl


Sewall Wright


R.A. Fisher

Counterfactual reasoning

What if we had $X = 0$, not $X = 1$, or vice versa?

$$Y_i^{(X=1)} - Y_i^{(X=0)}$$

Although we only observed "factual" Y, can we estimate "counterfactual" Y?

$$\overset{\text{observed}}{Y_i} = \begin{cases} Y_i^{(1)}, & X = 1 \\ Y_i^{(0)}, & X = 0 \end{cases}$$

Causal "effect" inference in two ways: $X \to Y$

Experiments/interventions

What would be the effect of *doing* $X = x$?

$$P(Y|do(X = x))$$

Although we are only "seeing" $X = x$, can we still estimate the "doing" effect?

$$P(Y|do(X = x)) \overset{???}{=} P(Y|X = x)$$


Judea Pearl


Sewall Wright


R.A. Fisher

Counterfactual reasoning

What if we had $X = 0$, not $X = 1$, or vice versa?

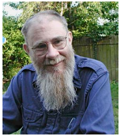
$$Y_i^{(X=1)} - Y_i^{(X=0)}$$

Although we only observed "factual" Y, can we estimate "counterfactual" Y?

$$\overset{\text{observed}}{Y_i} = \begin{cases} Y_i^{(1)}, & X = 1 \\ Y_i^{(0)}, & X = 0 \end{cases}$$


Don Rubin


Jerzy Neyman


David Lewis

Today's lecture

What is causal inference (fundamentals)

Graphical language for causal inference

Mendelian Randomization

Potential outcome framework (what if?)

Model-based causal inference

When we don't know an underlying causal graph

Causal effect inference from an experimentalist's perspective

Goal:

$$P(Y|\text{do}(X = x)) \leftarrow P(Y|X = x)$$

- What you are seeing is **not** what you are doing!
- What if we have a nearly complete picture of dependency?

Common misconception

Isn't causal inference learning a DAG from data?

[1] A student of Karl Pearson, who also argued the same.

Common misconception

[On Seawall Wright's graphical model], Henry Niles[1] disparaged, "The basic fallacy of the method appears to be **the assumption that it is possible to set up a priori a comparatively simple graphical system.**"

Pearl, *The Book of Why*, p.78

[1] A student of Karl Pearson, who also argued the same.

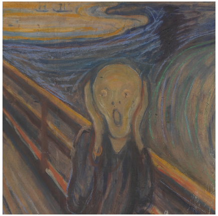
Common misconception - correlation implies causation

Henry Niles also wrote, "*To contrast 'causation' and 'correlation' is unwarranted* ***because causation is simply perfect correlation***"

Pearl, *The Book of Why*, p.78

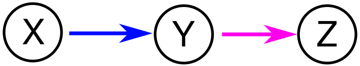
Common misconception - correlation implies causation

Henry Niles also wrote, "*To contrast 'causation' and 'correlation' is unwarranted* ***because causation is simply perfect correlation***"



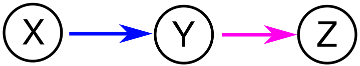
Pearl, *The Book of Why*, p.78

Causal Directed Acyclic Graph induces joint probability



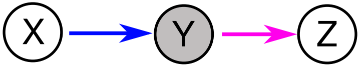
$$p(X,Y,Z) = \underbrace{p(Y|X)}_{\text{first edge}} \quad \underbrace{p(Z|Y)}_{\text{second edge}} \quad \underbrace{p(X)}_{\text{first node}}$$

Causal DAG: Causal path/trail and reachability



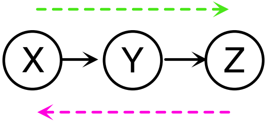
- There is a path between X and Y
- Also between Y and Z
- Also from X to Z

Causal DAG: Conditioning (and/or adjustment)



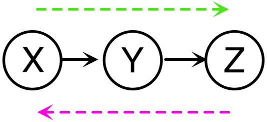
- A shaded node = conditioning like $P(Z|Y = y^{\star})$ and $P(Y = y^{\star}|X)$
- There is a path between X and Y
- Also between Y and Z
- **But** no path from X to Z

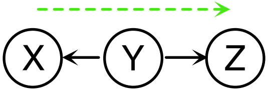
d-separation: testing conditional independence (flow vs. no flow)



Flow: The outcome of Z depends on the effect of X

d-separation: testing conditional independence (flow vs. no flow)

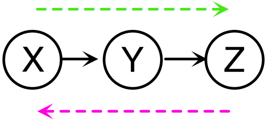


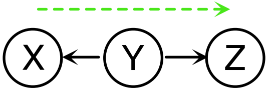


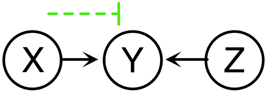
Flow: The outcome of Z depends on the effect of X

Flow: The outcome of Z depends on the effect of Y; so does the outcome of X on that of Y

d-separation: testing conditional independence (flow vs. no flow)





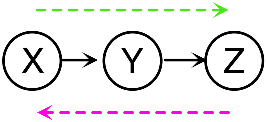


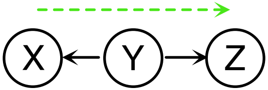
Flow: The outcome of Z depends on the effect of X

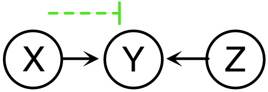
Flow: The outcome of Z depends on the effect of Y; so does the outcome of X on that of Y

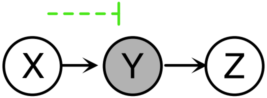
No Flow: The outcome of Z only affects Y; do does X

d-separation: testing conditional independence (flow vs. no flow)



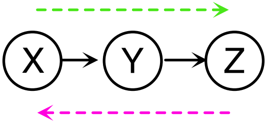


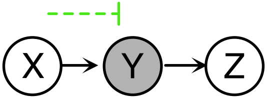




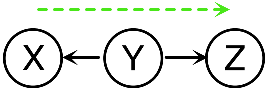
No Flow: The outcome of Z solely depends on $Y = y^*$, not X

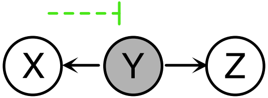
d-separation: testing conditional independence (flow vs. no flow)



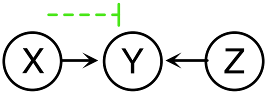


No Flow: The outcome of Z solely depends on $Y = y^*$, not X

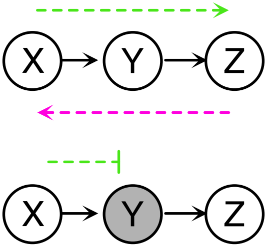




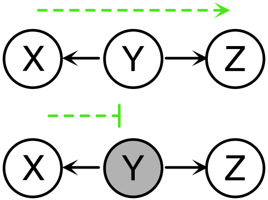
No Flow: The outcome of Z solely depends on $Y = y^*$, not X



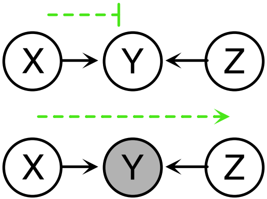
d-separation: testing conditional independence (flow vs. no flow)



No Flow: The outcome of Z solely depends on $Y = y^*$, not X

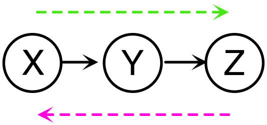


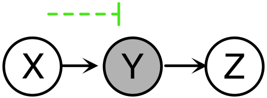
No Flow: The outcome of Z solely depends on $Y = y^*$, not X



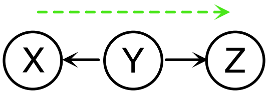
Flow: The observed outcome $Y = y^*$ could stem from X or/and Z

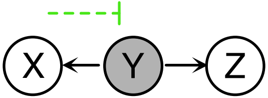
d-separation: testing conditional independence (flow vs. no flow)



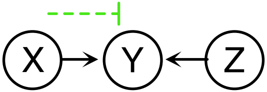


No Flow: The outcome of Z solely depends on $Y = y^*$, not X

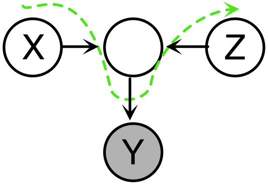




No Flow: The outcome of Z solely depends on $Y = y^*$, not X

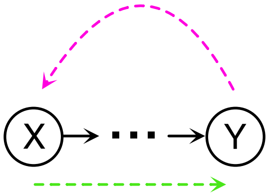


Flow: a deeper v-structure



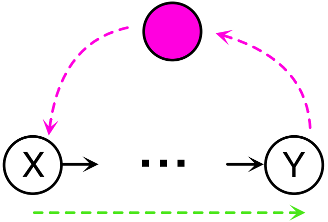
"Backdoor" exercise

Question: Which nodes should be "conditioned" and/or "adjusted" to block a reverse path from Y to X?



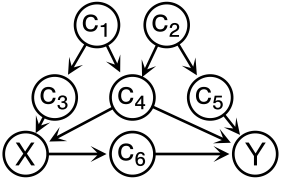
“Backdoor” exercise

Question: Which nodes should be “conditioned” and/or “adjusted” to block a reverse path from Y to X?



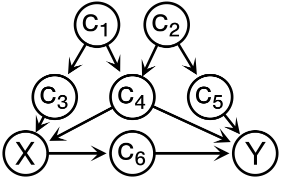
Can you identify and close "backdoor" paths?

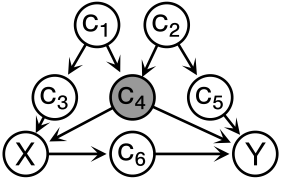
Goal: $X \to Y$ (so close other paths $Y \to X$)



Can you identify and close "backdoor" paths?

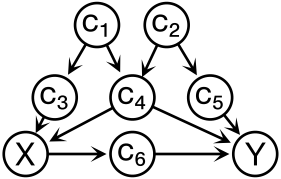
Goal: $X \rightarrow Y$ (so close other paths $Y \rightarrow X$)

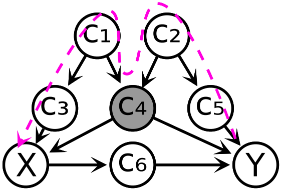




Can you identify and close "backdoor" paths?

Goal: $X \rightarrow Y$ (so close other paths $Y \rightarrow X$)

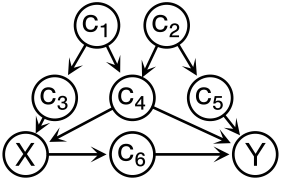


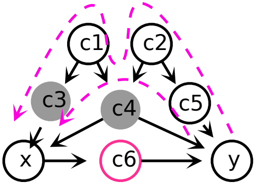


$$p(Y|do(X)) \neq \int_{C_4} p(Y|X, C_4)p(C_4)dC_4$$

Can you identify and close "backdoor" paths?

Goal: $X \rightarrow Y$ (so close other paths $Y \rightarrow X$)





$$p(Y|do(X)) = \int_{C_2,C_4} p(Y|X, C_2, C_4)p(C_2, C_4)dC_2C_4$$

Three steps of causal inference using a causal graph

1. Build a causal structural model

Three steps of causal inference using a causal graph

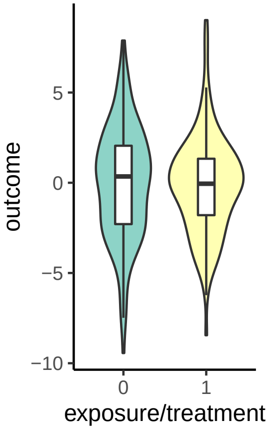
1. Build a causal structural model
2. Identify "back-door" variables

Three steps of causal inference using a causal graph

1. Build a causal structural model
2. Identify "back-door" variables
3. Adjust "back-door" variables

Estimate causal effects and sensitivity analysis

A working example: confounder adjustment in case-control study

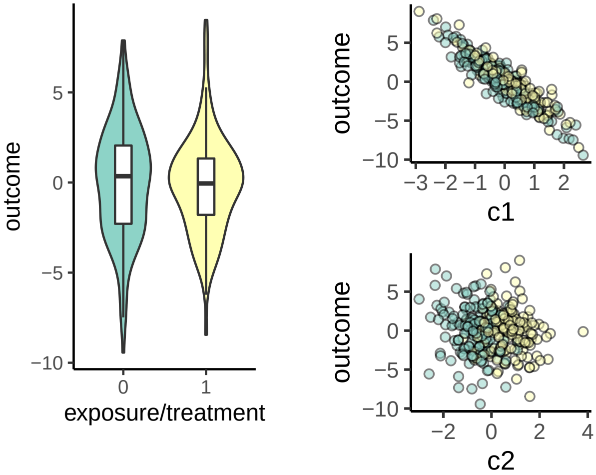


Key question: $X \overset{?}{\rightarrow} Y$

X: exposure variable

Y: outcome variable

A working example: confounder adjustment in case-control study



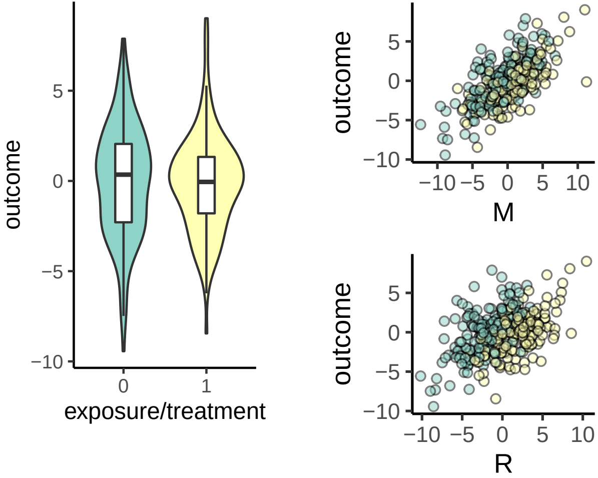
Key question: $X \overset{?}{\rightarrow} Y$

$\mathbf{X}$: exposure variable

$\mathbf{Y}$: outcome variable

$\mathbf{C_1}$ and $\mathbf{C_2}$: covariates

A working example: confounder adjustment in case-control study



Key question: $X \xrightarrow{?} Y$

$\mathbf{X}$: exposure variable

$\mathbf{Y}$: outcome variable

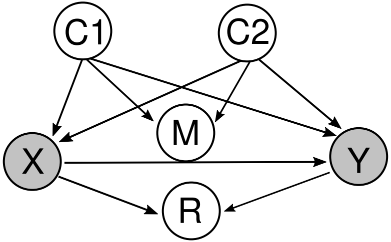
$\mathbf{C_1}$ and $\mathbf{C_2}$: covariates

$\mathbf{M}$: other covariate

$\mathbf{R}$: other covariate

Causal inference with a graphical model

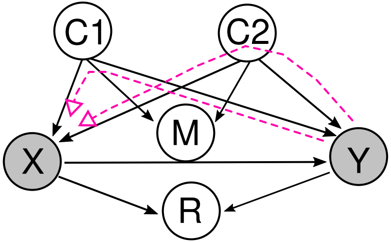
1. Build a causal structural model



What are potential backdoors?

Causal inference with a graphical model

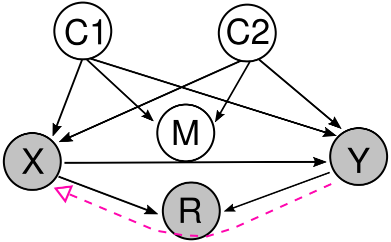
1. Build a causal structural model
2. Identify "back-door" paths/variables (*closing* $Y \rightarrow X$, *while opening* $X \rightarrow Y$)



How do we close them?

Causal inference with a graphical model

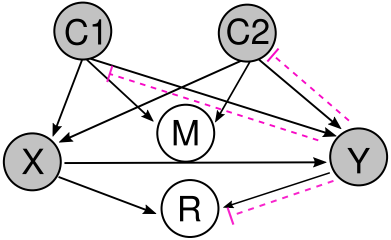
1. Build a causal structural model
2. Identify "back-door" paths/variables (*closing* $Y \rightarrow X$, *while opening* $X \rightarrow Y$)



What about this?

Causal inference with a graphical model

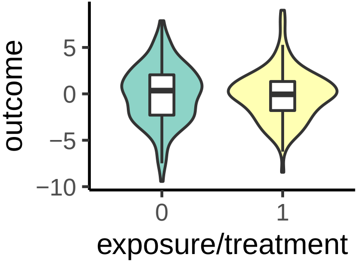
1. Build a causal structural model
2. Identify "back-door" paths/variables (*closing* $Y \to X$, *while opening* $X \to Y$)



Is this enough?

Causal inference with a graphical model

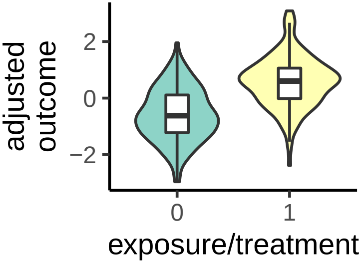
1. Build a causal structural model
2. Identify "back-door" paths/variables (*closing* $Y \to X$, *while opening* $X \to Y$)
3. Adjust "back-door" variables
4. Estimate causal effects



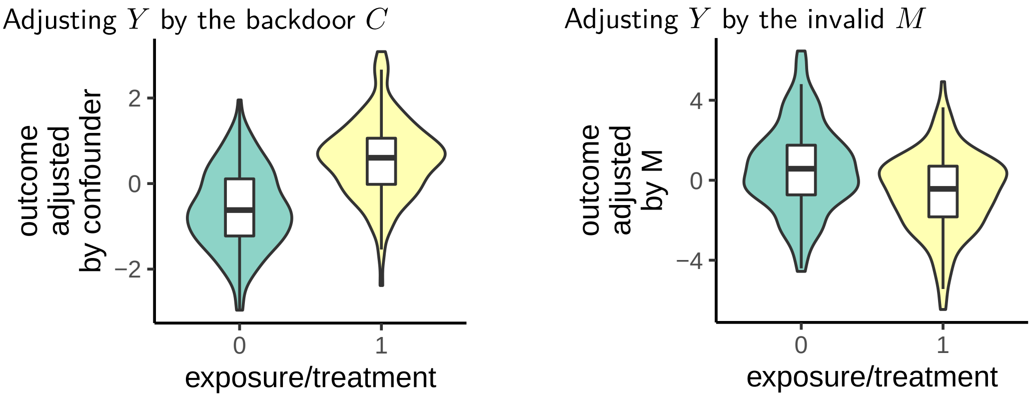
$$Y \leftarrow Y - \sum_{k=1}^{2} C_k \hat{\beta}_k$$

which approximates

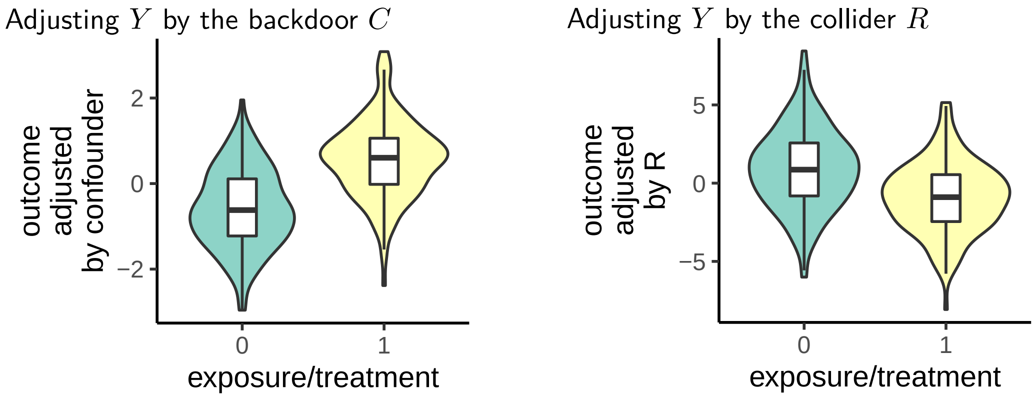
$$p(Y|X) = \int_C p(Y|X,C)p(C)dC$$



What would happen if we close a wrong "backdoor" variable?

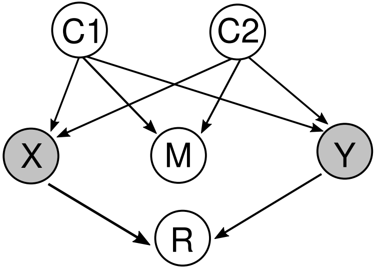


What would happen if we close a wrong "backdoor" variable?



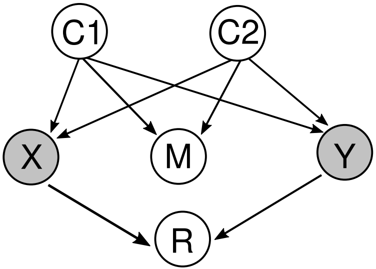
What would happen if we close a wrong “backdoor” variable?

Take another example:

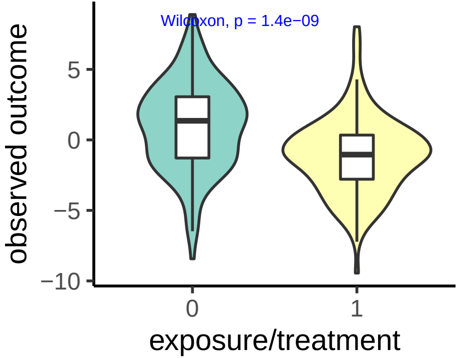


What would happen if we close a wrong "backdoor" variable?

Take another example:

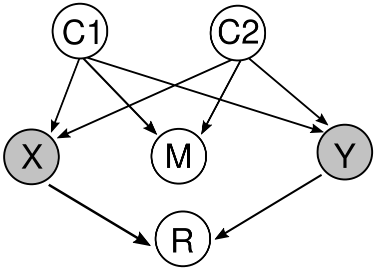


We can see the strong correlation between X and Y

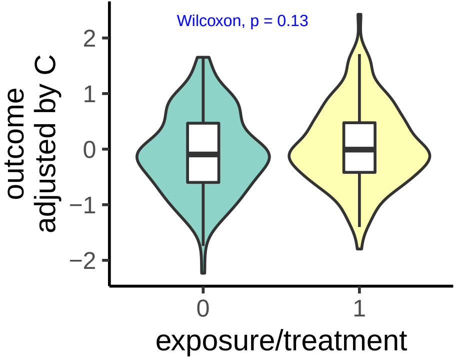


What would happen if we close a wrong “backdoor” variable?

Take another example:

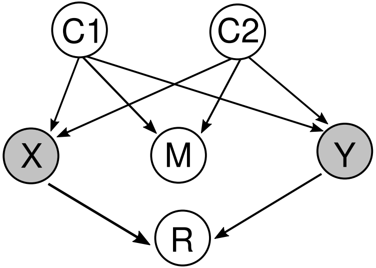


In fact, there is no causal relationship between X and Y

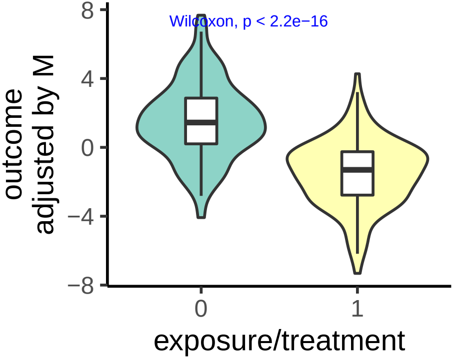


What would happen if we close a wrong “backdoor” variable?

Take another example:

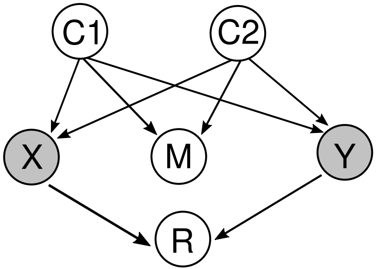


Adjusting M will not seal the backdoor... reinforce the spurious correlation

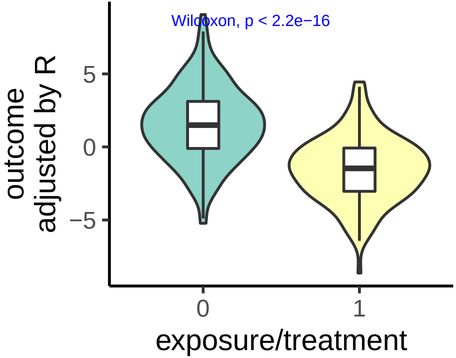


What would happen if we close a wrong “backdoor” variable?

Take another example:



Adjusting R will open another backdoor... reinforce the spurious correlation



Today's lecture

What is causal inference (fundamentals)

Graphical language for causal inference

Mendelian Randomization

Potential outcome framework (what if?)

Model-based causal inference

When we don't know an underlying causal graph

Mendelian + **Randomization**

Randomized experiments - RA Fisher's work at Rothamsted

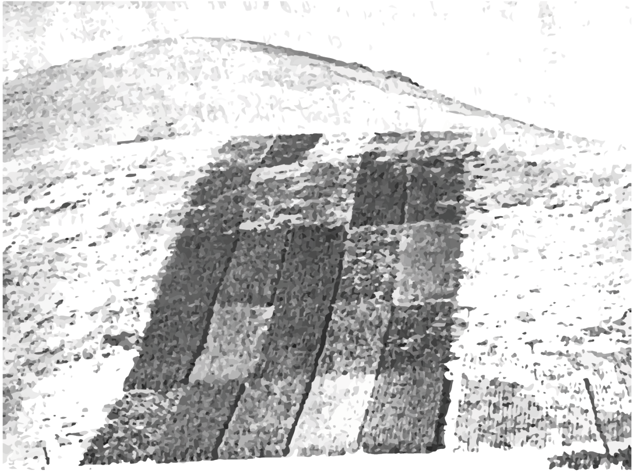
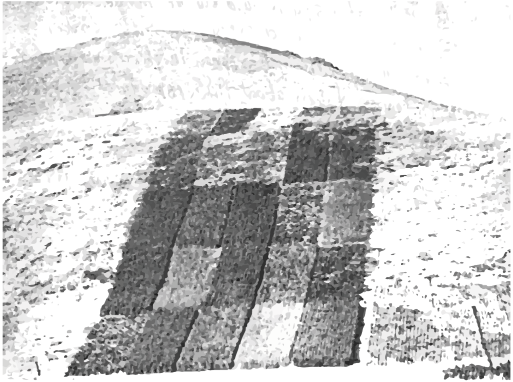


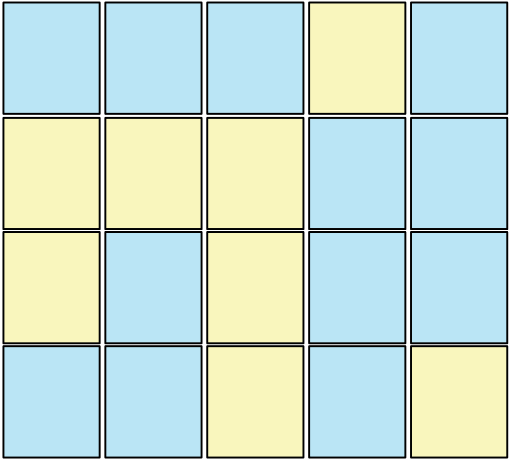
Image source: 'www.adelaide.edu.au'

Randomized experiments - RA Fisher's work at Rothamsted

Crops grown in the plots

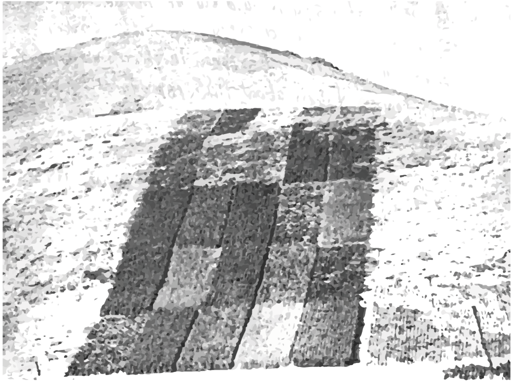


Experimental design

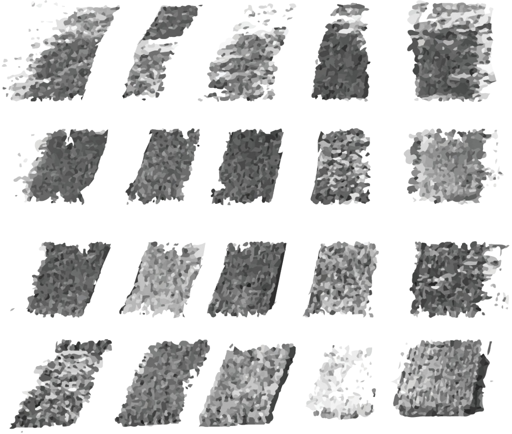


Randomized experiments - RA Fisher's work at Rothamsted

Crops grown in the plots

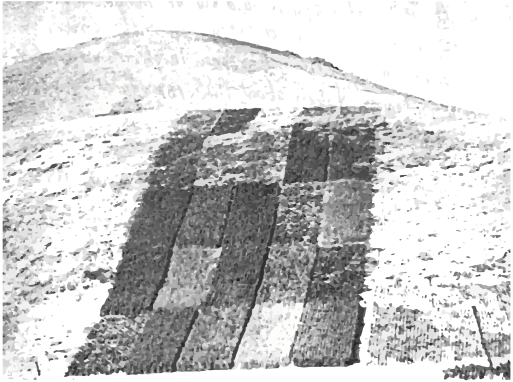


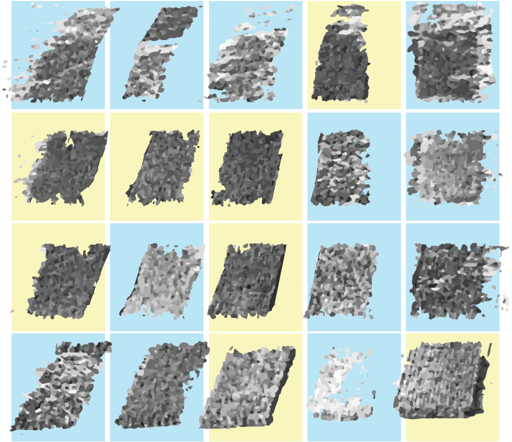
Random assignment of fertilizer



Randomized experiments - RA Fisher's work at Rothamsted

Crops grown in the plots **Compare the yields between the treated vs non-trea**

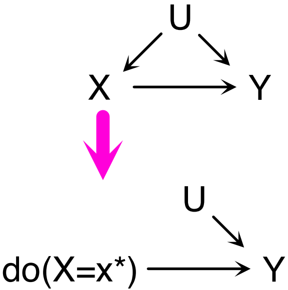


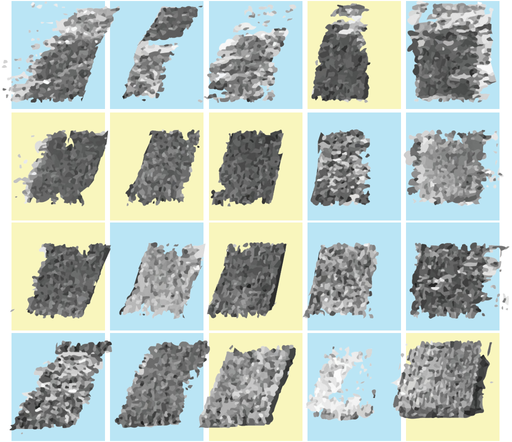


Randomized experiments - RA Fisher's work at Rothamsted

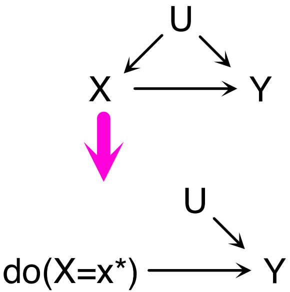
RCT $\Longrightarrow$ **"doing"**

Compare the yields between the treated vs non-trea

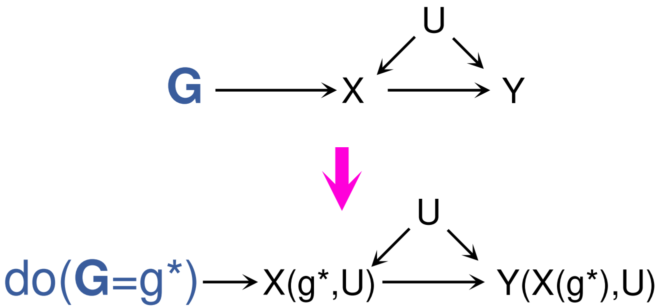




Randomization

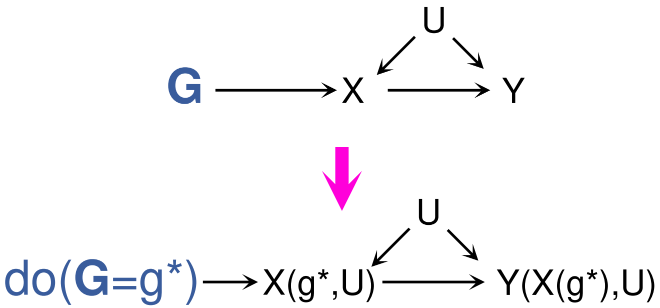


Mendelian Randomization



We will use a genetic variable G to mimic RCT

Mendelian Randomization



If a genetic variable G is a valid instrumental variable...

Mendelian Randomization in modern epidemiology studies

$$G \to X \to Y$$

- G: genotype
- X: APOE
- Y: cancer

APOLIPOPROTEIN E ISOFORMS, SERUM CHOLESTEROL, AND CANCER

SIR,—It is unclear whether the relation between low serum cholesterol levels and cancer[1] is causal. In many studies occult tumour may have depressed cholesterol levels though in others the relation was found when serum cholesterol had been measured many years before the cancer was diagnosed. The relation is probably not explained by diet, because in the Seven Countries Study cohorts with widely different diets and corresponding differences in mean cholesterol levels experienced similar mean cancer rates.[2,3] On the other hand, within each region cancer incidence was higher in men with a serum cholesterol in the lowest part of the cholesterol distribution for that country.[3] Thus, naturally low cholesterol levels are sometimes associated with increased cancer risk.[1,3]

Differences in the aminoacid sequence of apolipoprotein E (apo

Katan, *Lancet*, (1986)

Mendelian Randomization in modern epidemiology studies

$$G \to X \to Y$$

- G: genotype
- X: APOE
- Y: cancer

APOLIPOPROTEIN E ISOFORMS, SERUM CHOLESTEROL, AND CANCER

SIR,—It is unclear whether the relation between low serum cholesterol levels and cancer[1] is causal. In many studies occult tumour may have depressed cholesterol levels though in others the relation was found when serum cholesterol had been measured many years before the cancer was diagnosed. The relation is probably not explained by diet, because in the Seven Countries Study cohorts with widely different diets and corresponding differences in mean cholesterol levels experienced similar mean cancer rates.[2,3] On the other hand, within each region cancer incidence was higher in men with a serum cholesterol in the lowest part of the cholesterol distribution for that country.[3] Thus, naturally low cholesterol levels are sometimes associated with increased cancer risk.[1,3]

Differences in the aminoacid sequence of apolipoprotein E (apo

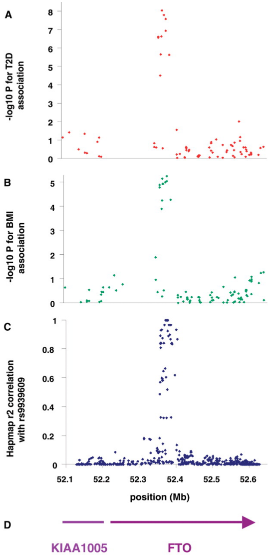
Katan, *Lancet*, (1986)

30TH THOMAS FRANCIS JR MEMORIAL LECTURE

'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*

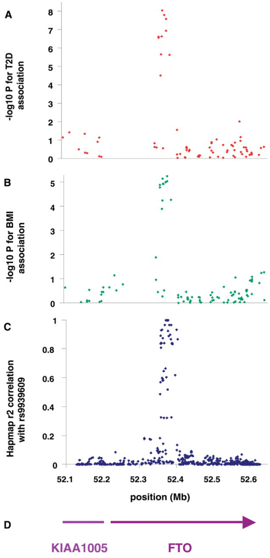
MR in action: *FTO* → fat mass → obesity, diabetes

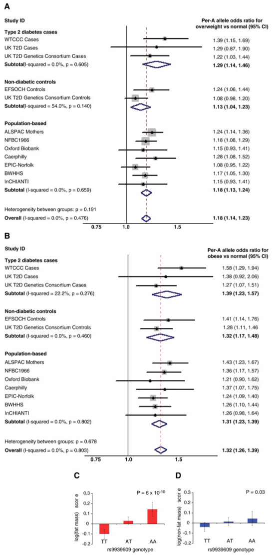
- Genotype in *FTO* locus → T2D



Frayling .. McCarthy, *Science* (2007)

MR in action: *FTO* → fat mass → obesity, diabetes

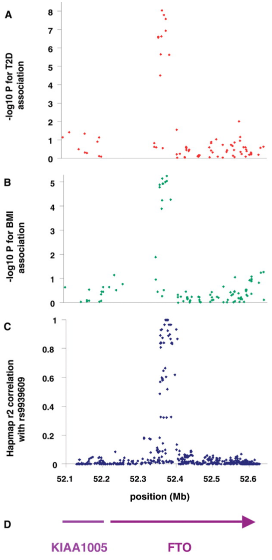


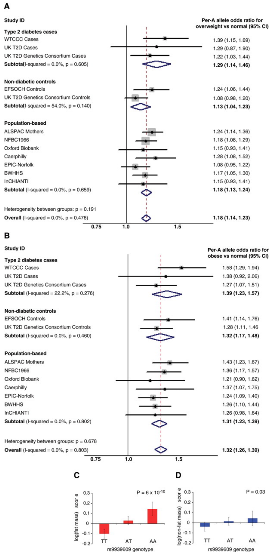


- Genotype in *FTO* locus → T2D
- *FTO* locus → fat mass

Frayling .. McCarthy, *Science* (2007)

MR in action: *FTO* $\rightarrow$ fat mass $\rightarrow$ obesity, diabetes





Frayling .. McCarthy, *Science* (2007)

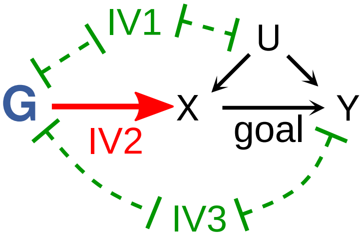
- Genotype in *FTO* locus $\rightarrow$ T2D
- *FTO* locus $\rightarrow$ fat mass
- Using *FTO* as "instrumental variable", we can ask other MR questions

Why doing Mendelian Randomization?

1. We do not have enough knowledge of U
2. We do not have a way to make interventions on $do(X = x)$

"Genotypes are beautifully randomized" - Fisher (1951)

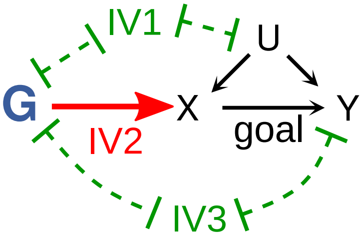
Definition of instrumental variable in MR study



- IV1: The genetic variant G is independent of the potential confounder variable U

Bowden *et al.* (2015)

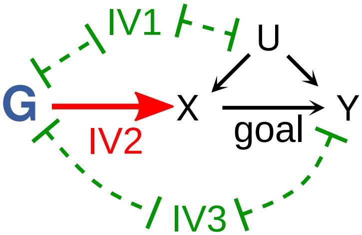
Definition of instrumental variable in MR study



- IV1: The genetic variant G is independent of the potential confounder variable U
- IV2: The genetic variant is associated with the exposure X

Bowden *et al.* (2015)

Definition of instrumental variable in MR study



- IV1: The genetic variant G is independent of the potential confounder variable U
- IV2: The genetic variant is associated with the exposure X
- IV3: The genetic variant is independent of the outcome Y conditioning on X

Bowden *et al.* (2015)

MR measures mediation effects of X to Y

Given that G is a valid instrumental variable...

Example:

- X: gene expression
- Y: disease phenotype

Suppose we have estimated

$$\begin{aligned} G &\xrightarrow{\alpha} X \\ X &= G\hat{\alpha} + \epsilon_X \end{aligned}$$

MR measures mediation effects of X to Y

Given that G is a valid instrumental variable...

Example:

- X: gene expression
- Y: disease phenotype

Suppose we have estimated

$$\begin{aligned} G &\xrightarrow{\alpha} X \\ X &= G\hat{\alpha} + \epsilon_X \end{aligned}$$

and

$$\begin{aligned} G &\xrightarrow{\gamma} Y \\ Y &= G\hat{\gamma} + \epsilon_Y \end{aligned}$$

MR measures mediation effects of X to Y

Given that G is a valid instrumental variable...

Example:

- X: gene expression
- Y: disease phenotype

Suppose we have estimated

$$\begin{aligned} G &\overset{\alpha}{\rightarrow} X \\ X &= G\hat{\alpha} + \epsilon_X \end{aligned}$$

and

$$\begin{aligned} G &\overset{\gamma}{\rightarrow} Y \\ Y &= G\hat{\gamma} + \epsilon_Y \end{aligned}$$

Goal: What is the causal effect β in $X \overset{\beta}{\rightarrow} Y$?

$$Y = X\beta + \epsilon'$$

MR measures mediation effects of X to Y

Given that G is a valid instrumental variable...

Example:

- X: gene expression
- Y: disease phenotype

Suppose we have estimated

$$
\begin{aligned}
G &\xrightarrow{\alpha} X \\
X &= G\hat{\alpha} + \epsilon_X
\end{aligned}
$$

and

$$
\begin{aligned}
G &\xrightarrow{\gamma} Y \\
Y &= G\hat{\gamma} + \epsilon_Y
\end{aligned}
$$

Goal: What is the causal effect β in $X \xrightarrow{\beta} Y$?

$$
\begin{aligned}
Y &= X\beta + \epsilon' \\
G\hat{\gamma} + \epsilon_Y &= (G\hat{\alpha} + \epsilon_X)\beta + \epsilon'
\end{aligned}
$$

MR measures mediation effects of X to Y

Given that G is a valid instrumental variable...

Example:

- X: gene expression
- Y: disease phenotype

Suppose we have estimated

$$\begin{aligned} G &\xrightarrow{\alpha} X \\ X &= G\hat{\alpha} + \epsilon_X \end{aligned}$$

and

$$\begin{aligned} G &\xrightarrow{\gamma} Y \\ Y &= G\hat{\gamma} + \epsilon_Y \end{aligned}$$

Goal: What is the causal effect β in $X \xrightarrow{\beta} Y$?

$$\begin{aligned} Y &= X\beta + \epsilon' \\ G\hat{\gamma} + \epsilon_Y &= (G\hat{\alpha} + \epsilon_X)\beta + \epsilon' \\ G\hat{\gamma} &= G\hat{\alpha}\beta + \cdots \end{aligned}$$

MR measures mediation effects of X to Y

Given that G is a valid instrumental variable...

Example:

- X: gene expression
- Y: disease phenotype

Suppose we have estimated

$$\begin{aligned} G &\xrightarrow{\alpha} X \\ X &= G\hat{\alpha} + \epsilon_X \end{aligned}$$

and

$$\begin{aligned} G &\xrightarrow{\gamma} Y \\ Y &= G\hat{\gamma} + \epsilon_Y \end{aligned}$$

Goal: What is the causal effect β in $X \xrightarrow{\beta} Y$?

$$\begin{aligned} Y &= X\beta + \epsilon' \\ G\hat{\gamma} + \epsilon_Y &= (G\hat{\alpha} + \epsilon_X)\beta + \epsilon' \\ G\hat{\gamma} &= G\hat{\alpha}\beta + \cdots \end{aligned}$$

The answer is as simple as

$$\mathbb{E}[\beta] = \frac{\gamma}{\alpha}$$

We will revisit this problem in the GWAS lectures

Today's lecture

- What is causal inference (fundamentals)
- Graphical language for causal inference
- Mendelian Randomization
- Potential outcome framework (what if?)
- Model-based causal inference
- When we don't know an underlying causal graph

The ultimate goal of causal inference

- Causality is tied to an action (doing), rather than an observation (seeing).
- Even the same object/subject/individual at a different time is a different object.
- **What if** we could have multiverse? **What if** we could turn back in time?

The ultimate goal of causal inference

- Causality is tied to an action (doing), rather than an observation (seeing).
- Even the same object/subject/individual at a different time is a different object.
- **What if** we could have multiverse? **What if** we could turn back in time?

Factual (observed) vs. *counterfactual* (what could have happened)

Rubin's causal model ($X \to Y$ with confounder S)

Causal assumptions (Rosenbaum & Rubin, 1983)

Unconfoundedness (ignorability)

$$(Y(0), Y(1)) \perp\!\!\!\perp X | S$$

Blocking out sufficient confounder variables, potential outcome $Y(X)$ is independent of observed exposure $X = x$.

E.g., The chance of unvaccinated myself getting COVID in the other universe has nothing to do with the fact that I had been vaccinated in this universe.

Rubin's causal model ($X \to Y$ with confounder S)

Causal assumptions (Rosenbaum & Rubin, 1983)

Unconfoundedness (ignorability)

$$(Y(0), Y(1)) \perp\!\!\!\perp X | S$$

Blocking out sufficient confounder variables, potential outcome $Y(X)$ is independent of observed exposure $X = x$.

E.g., The chance of unvaccinated myself getting COVID in the other universe has nothing to do with the fact that I had been vaccinated in this universe.

Overlap (smoothness)

$$0 < P(X = 1|S) < 1$$

There is a positive chance of being exposed and not exposed in all values of S, e.g., strata.

E.g., There is always some group who get vaccinated and the other group who decide not to get vaccinated across all conditions, e.g., gender, socioeconomic indexes, age, etc.

SUTVA (Stable Unit Treatment Variable Assumption)

1. The potential outcome of any unit do not vary with the treatment assigned to other units (**No Interference** of causal effects).
2. For each unit, there is no different forms of each treatment level, which lead to different potential outcomes (**No Hidden variation of Treatments**).

Why is it called a “potential” outcome model?

- If my causal statement is: “My headache disappeared because I took a pill of aspirin.”
- Then, we need to investigate counterfactual outcome: “Had I not taken an aspirin, would my headache disappear?”
- Definition of causal effects: Comparison between two (or multiple) different potential outcomes **on the same unit**.

Holland (1986)

The fundamental problem of causal inference is the problem that at most one of the potential outcomes can be realized and thus observed.

Causal inference is fundamentally a missing data problem

For $X_i = 1$,

$$Y_i^{\mathsf{obs}} = Y_i(1), Y_i^{\mathsf{mis}} = Y_i(0)$$

For $X_i = 0$,

$$Y_i^{\mathsf{mis}} = Y_i(1), Y_i^{\mathsf{obs}} = Y_i(0)$$

i	X	Y(0)	Y(1)
1	0	Y	?
2	1	?	Y
3	1	?	Y
4	0	Y	?
5	1	?	Y
...	...	...	...

Average causal effect:

$$\frac{1}{N}\sum_{i=1}^{N} Y_i(1) - Y_i(0)$$

Statistical aspect of causal inference ($N > 1$)

- Learning about causal effects typically involves multiple units.
- A fundamental "mistake" of statistical inference: Considering a unit at different time as the same unit.
- The problem of causal inference is a missing data problem.
- What was the missing data assignment mechanism?
- Prediction of assignment mechanisms by pre-treatment variables.

Causal estimand

- $Y_i(X_i)$: Potential outcome of a unit i with exposure X_i
- S_{ik}: Covariate k for a unit i
- Atomic causal estimand:

$$Y_i(1) - Y_i(0)$$

- Average causal effect:

$$\frac{1}{N}\sum_{i=1}^{N} Y_i(1) - Y_i(0)$$

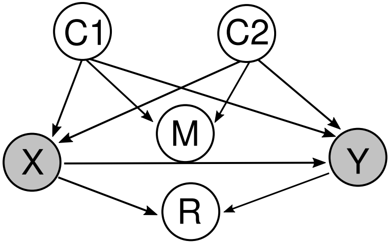
History of causal inference

- Potential outcome: Neyman (1923)
- Randomized assignment: Fisher (1925)
- Regressing out covariates: Yule (1897)

A classification of assignment mechanisms

- Experimental: The assignment mechanism is known and controlled by the researcher.
- Observational: The assignment mechanisms is **not** known to, or **not** under the control of the researcher.

Revisit the same example



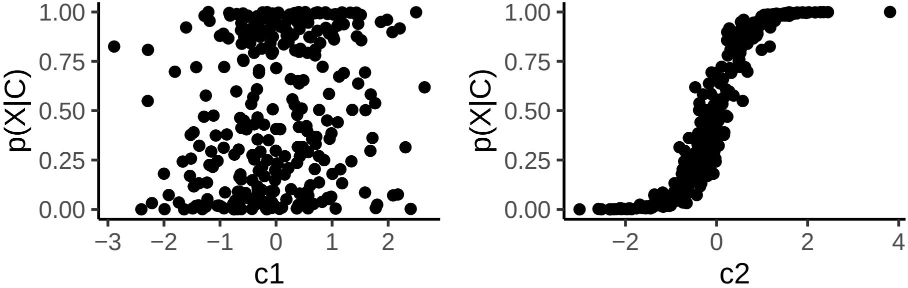
- What are the backdoor variables?
- How did we generate X?

Inverse Propensity Weighting

What is the model for X given confoudners (the backdoor variables)?

Propensity $=$ probability of assignment $X = 1$:

$$p(X|C_1, C_2) \approx \frac{1}{1 + \exp(-\beta_0 - \beta_1 C_1 - \beta_2 C_2)}$$



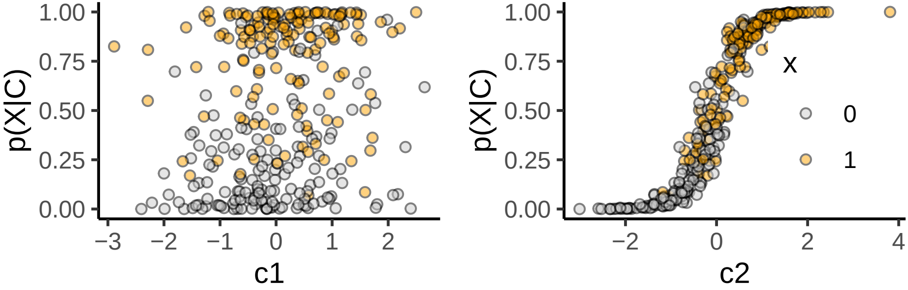
Horvitz and Thomson, *JASA*, 1952

Inverse Propensity Weighting

What is the model for X given confoudners (the backdoor variables)?

Propensity $=$ probability of assignment $X = 1$:

$$p(X|C_1, C_2) \approx \frac{1}{1 + \exp(-\beta_0 - \beta_1 C_1 - \beta_2 C_2)}$$



Horvitz and Thomson, *JASA*, 1952

Intuition behind IPW

What if we have assigned $X = 1$ unconfounded by C?

Intuition behind IPW

What if we have assigned $X = 1$ unconfounded by C?

In other words, what if samples could be drawn more from the underrepresented group?

Intuition behind IPW

What if we have assigned $X = 1$ unconfounded by C?

In other words, what if samples could be drawn more from the underrepresented group?

Likewise, what if samples could be dropped in the overrepresented group?

Inverse Propensity Weighting "inverse" confounded assignment

$$\hat{Y}_i^{(1)} = \frac{X_i Y_i}{\hat{p}(X_i = 1 | C_i)}$$

$$\hat{Y}_i^{(0)} = \frac{(1 - X_i) Y_i}{1 - \hat{p}(X_i = 1 | C_i)}$$

```
p.xc <-
    glm(x~cc, family="binomial") %>%
    predict() %>%
    sigmoid() %>%
    clamp() # avoid 0 or 1

ww <- x / p.xc + (1-x) / (1-p.xc)
```

Inverse Propensity Weighting "inverse" confounded assignment

$$\hat{Y}_i^{(1)} = \frac{X_i Y_i}{\hat{p}(X_i = 1|C_i)}$$

$$\hat{Y}_i^{(0)} = \frac{(1 - X_i) Y_i}{1 - \hat{p}(X_i = 1|C_i)}$$

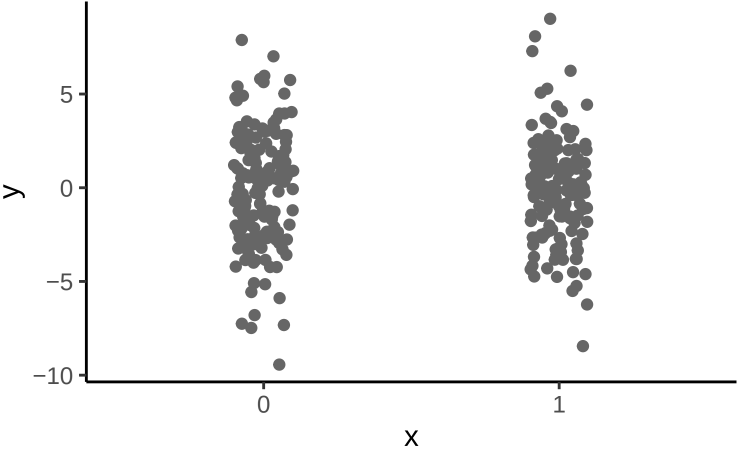
equivalently give weights for $\forall$ i

$$W_i \propto \begin{cases} 1/p(X_i = 1|C_i) & X_i = 1 \\ 1/p(X_i = 0|C_i) & X_i = 0 \end{cases}$$

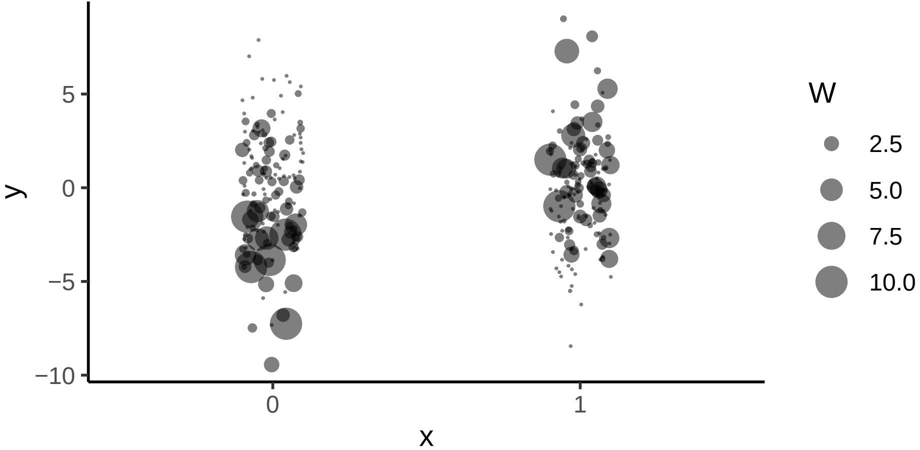
```
p.xc <-
    glm(x~cc, family="binomial") %>%
    predict() %>%
    sigmoid() %>%
    clamp() # avoid 0 or 1

ww <- x / p.xc + (1-x) / (1-p.xc)
```

Take samples inversely proportional to propensity

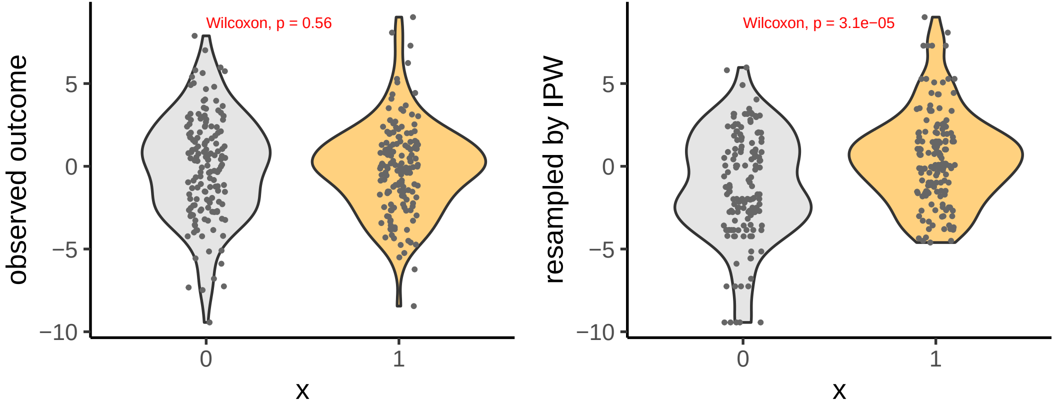


Take samples inversely proportional to propensity



$$W_i \propto \begin{cases} 1/p(X_i = 1|C_i) & X_i = 1 \\ 1/p(X_i = 0|C_i) & X_i = 0 \end{cases}$$

Take samples inversely proportional to propensity



$$W_i \propto \begin{cases} 1/p(X_i = 1|C_i) & X_i = 1 \\ 1/p(X_i = 0|C_i) & X_i = 0 \end{cases}$$

Why IPW works? Unbiased estimate potential outcome

Letting $e(z) = \hat{p}(X = 1|C = z)$,

we can prove $\mathbb{E}[XY/e(X)] \to \mathbb{E}\left[Y^{(1)}\right]$

using

- Strong ignorability
- Smoothness
- Stable Unit Treatment (exposure) Variable

Why IPW works? Unbiased estimate potential outcome

Letting $e(z) = \hat{p}(X = 1|C = z)$,

$$\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\right] \quad =$$

Why IPW works? Unbiased estimate potential outcome

Letting $e(z) = \hat{p}(X = 1 | C = z)$,

$$\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)} \middle| C_i\right]\right] \quad \text{(law of total expectation)}$$

Why IPW works? Unbiased estimate potential outcome

Letting $e(z) = \hat{p}(X = 1 | C = z)$,

$$\begin{aligned}
\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\middle| C_i\right]\right] && \text{(law of total expectation)} \\
\text{(C is sufficient backdoor)} \quad &= \mathbb{E}\left[\mathbb{E}\left[\frac{X_i Y_i^{(1)}}{e(C_i)}\middle| C_i\right]\right]
\end{aligned}$$

Why IPW works? Unbiased estimate potential outcome

Letting $e(z) = \hat{p}(X = 1|C = z)$,

$$\begin{aligned}
\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\middle| C_i\right]\right] && \text{(law of total expectation)} \\
\text{(C is sufficient backdoor)} \quad &= \mathbb{E}\left[\mathbb{E}\left[\frac{X_i Y_i^{(1)}}{e(C_i)}\middle| C_i\right]\right] \\
\text{(strong ignorability)} \quad &= \mathbb{E}\left[Y_i^{(1)} \mathbb{E}\left[\frac{X_i}{e(C_i)}\middle| C_i\right]\right] && Y^{(1)} \perp\!\!\!\perp X|C
\end{aligned}$$

Why IPW works? Unbiased estimate potential outcome

Letting $e(z) = \hat{p}(X = 1 | C = z)$,

$$\begin{aligned}
\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\middle| C_i\right]\right] && \text{(law of total expectation)} \\
\text{(C is sufficient backdoor)} \quad &= \mathbb{E}\left[\mathbb{E}\left[\frac{X_i Y_i^{(1)}}{e(C_i)}\middle| C_i\right]\right] \\
\text{(strong ignorability)} \quad &= \mathbb{E}\left[Y_i^{(1)} \mathbb{E}\left[\frac{X_i}{e(C_i)}\middle| C_i\right]\right] && Y^{(1)} \perp\!\!\!\perp X | C \\
\text{(smoothness)} \quad &= \mathbb{E}\left[Y_i^{(1)} \frac{1}{\cancel{e(C_i)}} \cancel{\mathbb{E}[X_i | C_i]}\right] && 0 < e(C) < 1
\end{aligned}$$

Why IPW works? Unbiased estimate potential outcome

Letting $e(z) = \hat{p}(X = 1 | C = z)$,

$$\begin{aligned}
\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{X_i Y_i}{e(C_i)}\middle| C_i\right]\right] && \text{(law of total expectation)} \\
\text{(C is sufficient backdoor)} &= \mathbb{E}\left[\mathbb{E}\left[\frac{X_i Y_i^{(1)}}{e(C_i)}\middle| C_i\right]\right] \\
\text{(strong ignorability)} &= \mathbb{E}\left[Y_i^{(1)}\mathbb{E}\left[\frac{X_i}{e(C_i)}\middle| C_i\right]\right] && Y^{(1)} \perp\!\!\!\perp X | C \\
\text{(smoothness)} &= \mathbb{E}\left[Y_i^{(1)}\frac{1}{\cancel{e(C_i)}}\cancel{\mathbb{E}[X_i | C_i]}\right] && 0 < e(C) < 1 \\
&= \mathbb{E}\left[Y^{(1)}\right]
\end{aligned}$$

When do we use IPW to estimate potential outcomes?

- Backdoor variables are sufficiently characterized

When do we use IPW to estimate potential outcomes?

- Backdoor variables are sufficiently characterized
- We have an unbiased way to estimate the propensity model

When do we use IPW to estimate potential outcomes?

- Backdoor variables are sufficiently characterized
- We have an unbiased way to estimate the propensity model
- Smooth overlap $1 > e(X) > 0$

Today's lecture

What is causal inference (fundamentals)

Graphical language for causal inference

Mendelian Randomization

Potential outcome framework (what if?)

Model-based causal inference

When we don't know an underlying causal graph

G-formula bridges two different worlds (DAG meets PO)

- Can we *directly* estimate $p(Y^{(1)})$ and $p(Y^{(0)})$?
- In other words, $p(Y|\mathrm{do}(X=1))$ and $p(Y|\mathrm{do}(X=0))$.
- Why is it interesting? Knowing $p(Y^{(x)})$, we know $\mathbb{E}\left[Y^{(x)}\right]$ (aka potential outcome).

The "G"-formula?

"G" for "generalized" method (James Robins, 1986)

Suppose we have characterize potential backdoor variables S for closing off unwanted paths $Y \to X$ to identify a causal path $X \to Y$.

For each condition for exposure variable, $X = 0$ or $X = 1$,

- $p(Y^{(1)}) = \int_S p(Y|X = 1, S)p(S)dS$
- $p(Y^{(0)}) = \int_S p(Y|X = 0, S)p(S)dS$

How do we derive G-formula?

Bayesian integration results in causal estimand?

$$p(Y^{(1)}) \quad = \quad \int\int p(Y^{(1)}|X,S)p(X|S)p(S)dXdS$$

How do we derive G-formula?

Bayesian integration results in causal estimand?

$$
\begin{aligned}
p(Y^{(1)}) &= \int\int p(Y^{(1)}|X,S)p(X|S)p(S)dXdS \\
&= \int_S\int_X \underbrace{p(Y^{(1)}|S)p(X|S)}_{\text{ignorability}} p(S)dXdS \quad \Longleftarrow (Y^{(1)}, Y^{(0)} \perp\!\!\!\perp X|S)
\end{aligned}
$$

How do we derive G-formula?

Bayesian integration results in causal estimand?

$$
\begin{aligned}
p(Y^{(1)}) &= \int\int p(Y^{(1)}|X,S)p(X|S)p(S)dXdS \\
&= \int_S\int_X \underbrace{p(Y^{(1)}|S)p(X|S)}_{\text{ignorability}}p(S)dXdS \quad \Longleftarrow (Y^{(1)},Y^{(0)} \perp\!\!\!\perp X|S) \\
&= \int_S p(Y^{(1)}|S)\underbrace{\left[\int_X p(X|S)dX\right]}_{\text{sum to 1}}p(S)dS \quad \Longleftarrow P(X|S) > 0
\end{aligned}
$$

How do we derive G-formula?

Bayesian integration results in causal estimand?

$$
\begin{aligned}
p(Y^{(1)}) &= \int\int p(Y^{(1)}|X,S)p(X|S)p(S)dXdS \\
&= \int_S\int_X \underbrace{p(Y^{(1)}|S)p(X|S)}_{\text{ignorability}} p(S)dXdS \quad \Longleftarrow (Y^{(1)}, Y^{(0)} \perp\!\!\!\perp X|S) \\
&= \int_S p(Y^{(1)}|S) \underbrace{\left[\int_X p(X|S)dX\right]}_{\text{sum to 1}} p(S)dS \quad \Longleftarrow P(X|S) > 0 \\
&= \int_S p(Y|X=1,S)p(S)dS
\end{aligned}
$$

G-formula justifies outcome regression models

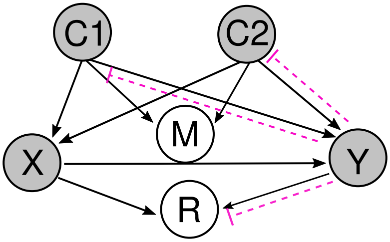
How do we estimate "causal" expectation

$$\begin{aligned}
\mathbb{E}\left[Y^{(1)}\right] &= \int_y y\, p(Y^{(1)} = y) dy \\
&= \int_y y \int_S p(Y = y | X = 1, S) p(S) dS\, dy \\
&= \int_S \mathbb{E}[Y | X = 1, S] dS
\end{aligned}$$

Likewise,

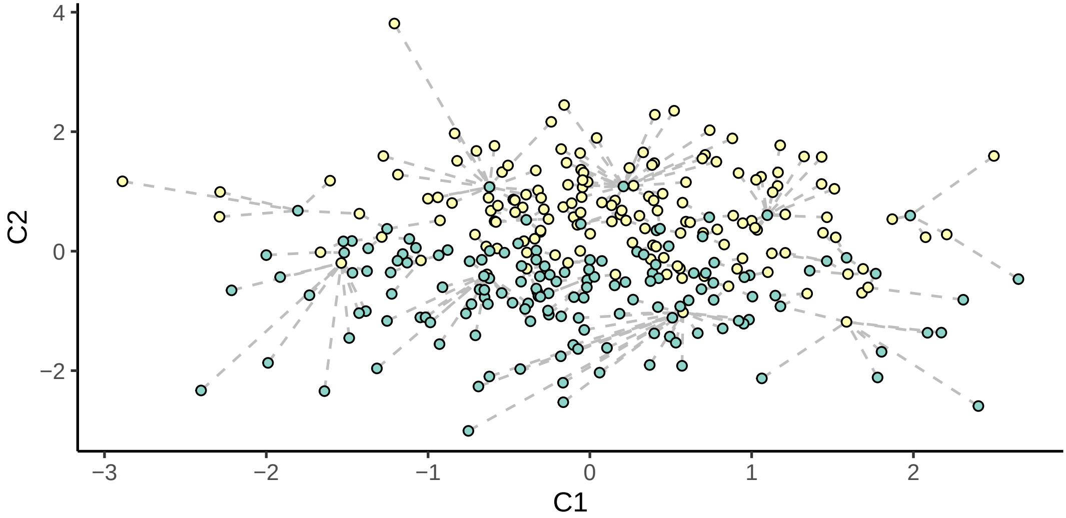
$$\mathbb{E}\left[Y^{(0)}\right] = \int_S \mathbb{E}[Y | X = 0, S] dS$$

Estimating potential outcome by matching

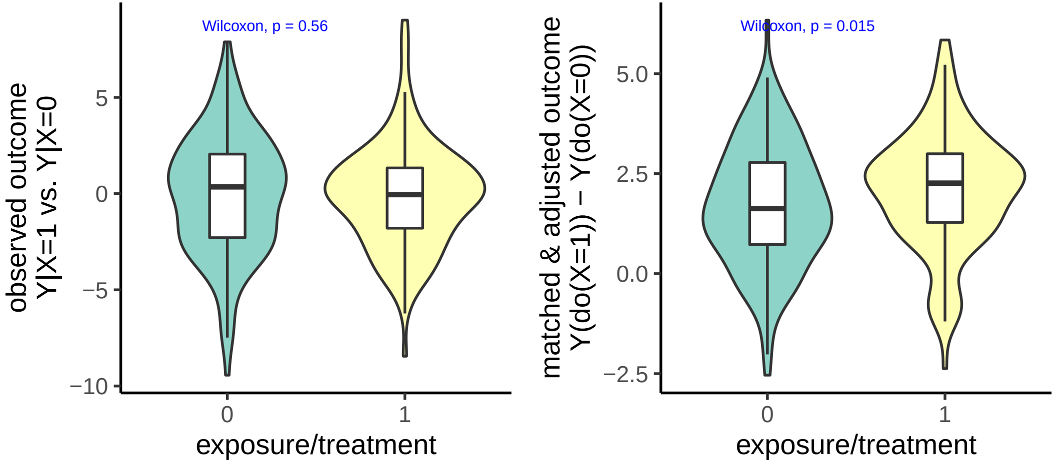


- Estimtae $\mathbb{E}\left[Y_i^{(0)}|C_{i1}, C_{i2}\right]$ for $X_i = 1$ to compare with $\mathbb{E}[Y_i|X_i = 1]$
- Estimtae $\mathbb{E}\left[Y_i^{(1)}|C_{i1}, C_{i2}\right]$ for $X_i = 0$ to compare with $\mathbb{E}[Y_i|X_i = 0]$

Estimating potential outcome by matching



Estimating potential outcome by matching



Bayesian Additive Regression Tree (BART) approach

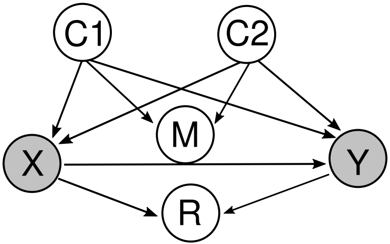
- G-formula $\rightarrow$ outcome regression models
- Regression model for $Y \sim S$ for each $X = 1$ and $X = 0$ using BART

$$
\begin{aligned}
\mathbb{E}\left[Y^{(1)}\right] &= \int_S \mathbb{E}[Y|X=1,S]dS \\
\mathbb{E}\left[Y^{(0)}\right] &= \int_S \mathbb{E}[Y|X=0,S]dS
\end{aligned}
$$

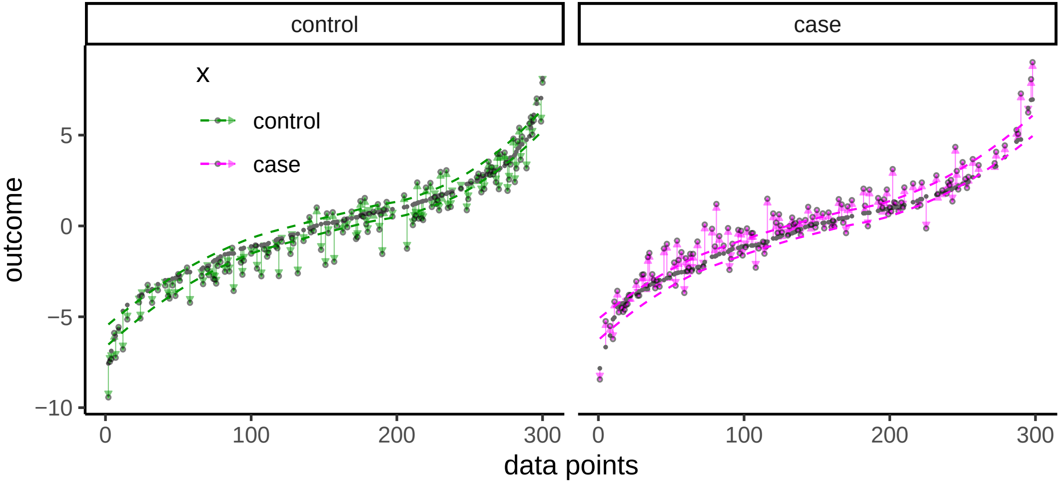
- Estimate causal effect: $\mathbb{E}\left[Y^{(1)}\right] - \mathbb{E}\left[Y^{(0)}\right]$

Hill, *Bayesian Nonparametric Modeling for Causal Inference* (2011)

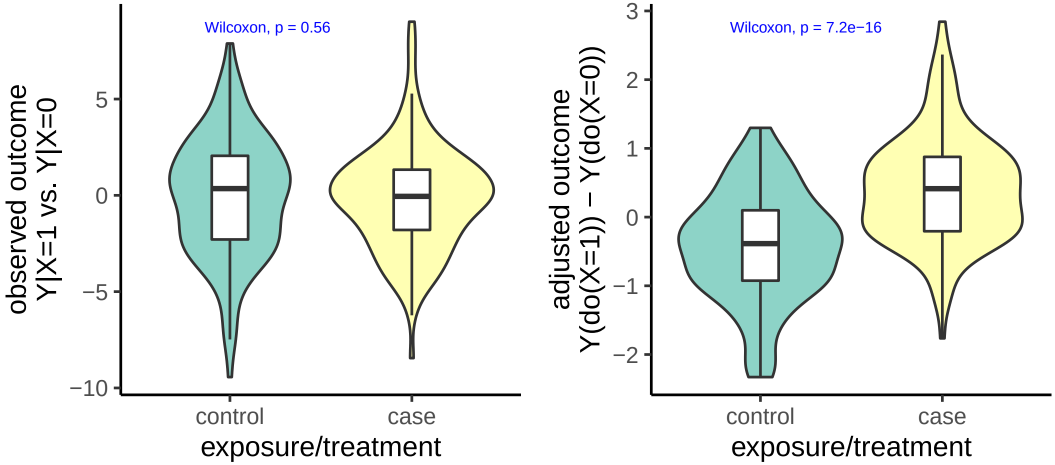
The same example with causal relationship from X to Y



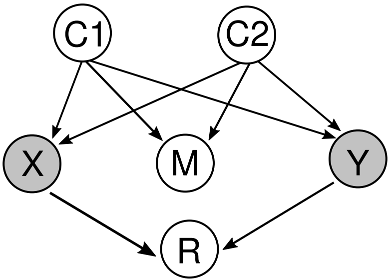
BART: a regression model to impute potential outcomes



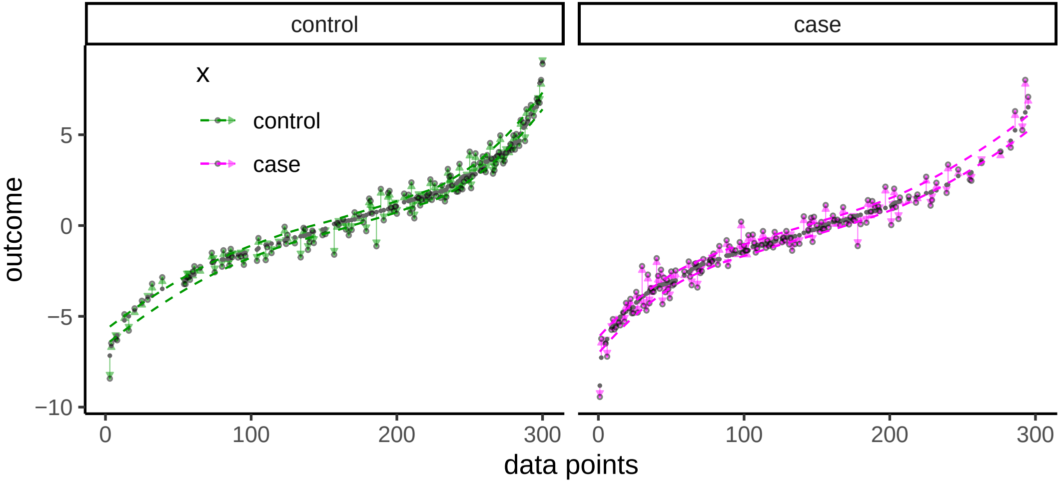
BART: a regression model to impute potential outcomes



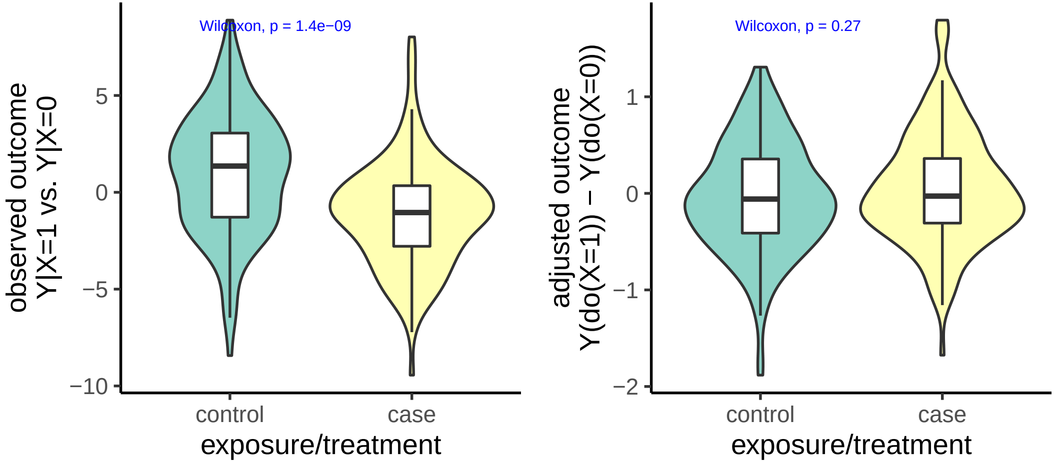
Revisit the example without causal relationship from X to Y



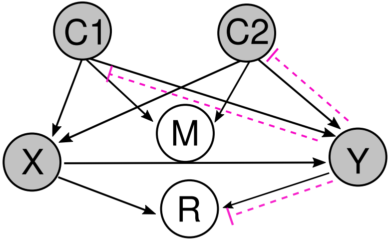
BART correctly adjusts spurious correlation



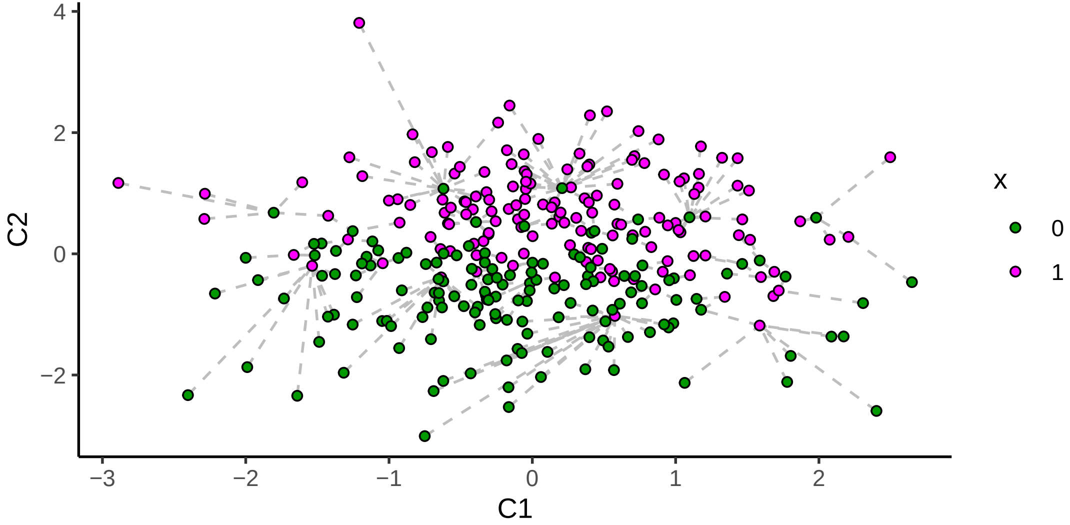
BART correctly adjusts spurious correlation



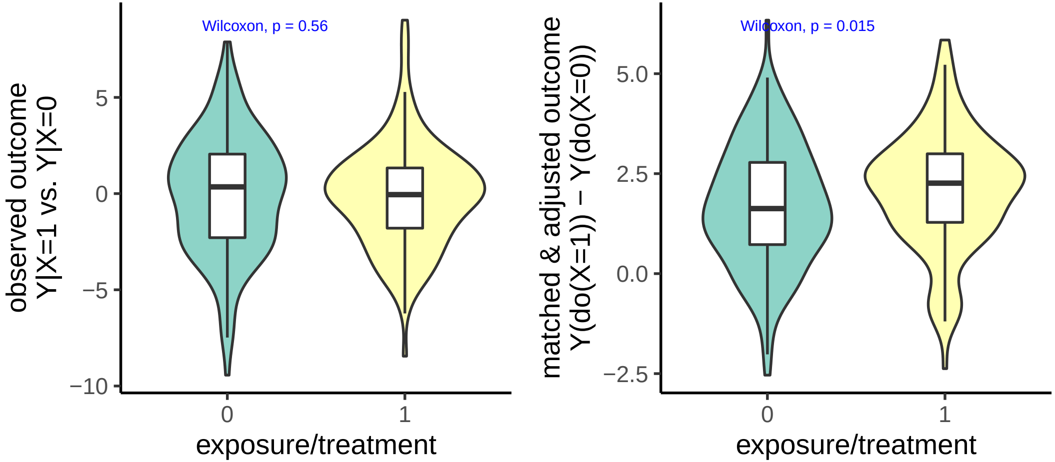
Estimating potential outcome by matching - 1



Estimating potential outcome by matching - 1



Estimating potential outcome by matching - 1



Today's lecture

What is causal inference (fundamentals)

Graphical language for causal inference

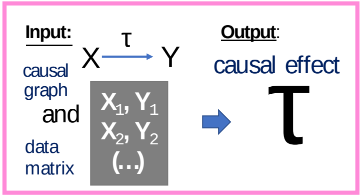
Mendelian Randomization

Potential outcome framework (what if?)

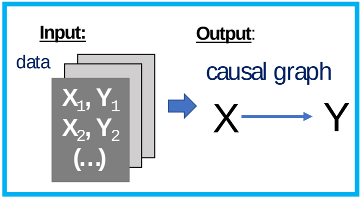
Model-based causal inference

When we don't know an underlying causal graph

Causal Effect Inference



Causal structure discovery



Causal DAG from data: an intractiable, yet very tempting problem

Intractability

Large-sample learning learning of a Bayesian Network is NP-hard.

Chickering and Heckerman, *UAI*, (2002)

Causal DAG from data: an intractiable, yet very tempting problem

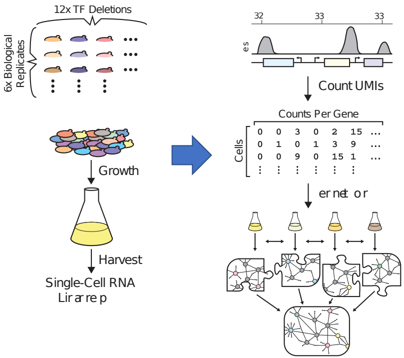
Intractability

Large-sample learning learning of a Bayesian Network is NP-hard.

Chickering and Heckerman, *UAI*, (2002)

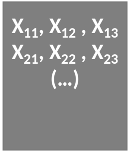
But there is hope if we have "intervention" data

Can we borrow the awesome power of genetics?



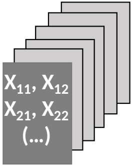
Jackson *et al.* (2020)

Early 2000's



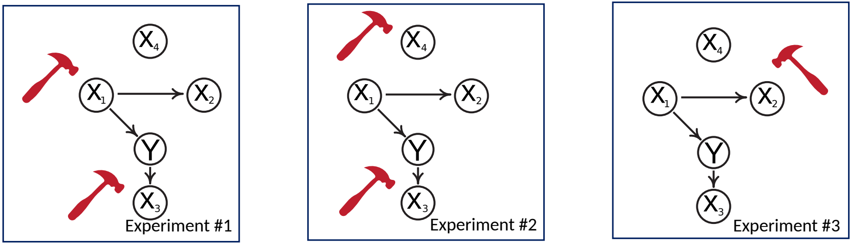
A single or only a handful of conditions

After perturb-seq



Massive (random) perturbation assays à hundreds of data matrices

Causal Structure Discovery by Invariance



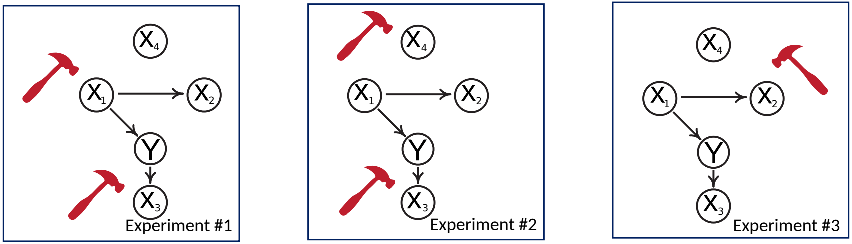
Knock-out do(X=0)

The premise: There is a causal structure that remain **invariant** across multiple experimental conditions.

Peters, Bühlmann, Meinshausen, *J. Royal Stat. B* (2016)

Causal Structure Discovery by Invariance

Testing hypothesis (model): $Y \sim X_1 + \epsilon$



Knock-out do(X=0)

Exp #1:
$Y = \epsilon_Y$ and
$X_1 = 0$

Exp #2:
$Y = X_1 + \epsilon_Y$ and
$X_1 = \epsilon_1$

Exp #3:
$Y = \epsilon_Y$ and
$X_1 = 0$

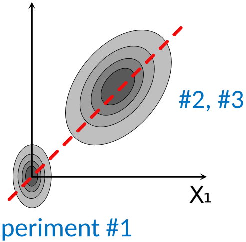
Does this model make sense in all the experiments?

Observation:

$Y = \varepsilon_y$ Experiment #1

$Y = X_1 + \varepsilon_y$ Experiment #2

$Y = X_1 + \varepsilon_y$ Experiment #3



Regression Residuals:

$\boxed{\varepsilon_y} \sim 0$ #1

$\cancel{X_1} + \boxed{\varepsilon_y} \sim \cancel{X_1}$ #2

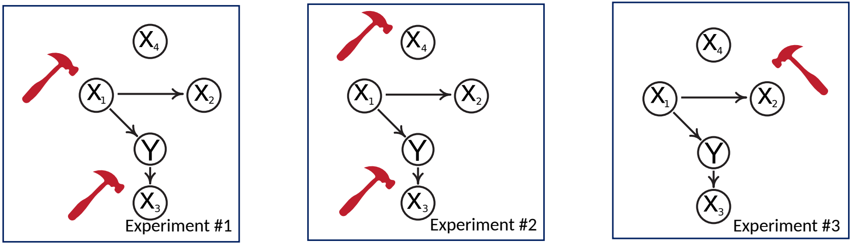
$\cancel{X_1} + \boxed{\varepsilon_y} \sim \cancel{X_1}$ #3

The same distribution of the residuals

Cannot reject any experiments

Causal Structure Discovery by Invariance - 2

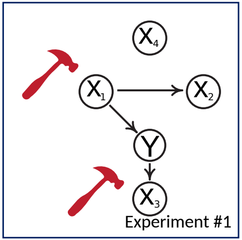
Testing hypothesis (model): $Y \sim X_1 + X_3 + \epsilon$

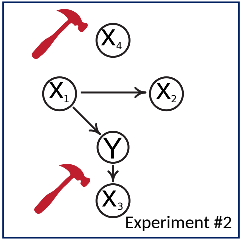


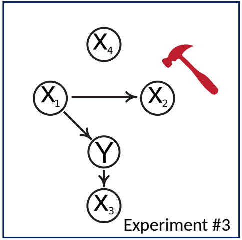
Knock-out do(X=0)

Causal Structure Discovery by Invariance - 2

Testing hypothesis (model): $Y \sim X_1 + X_3 + \epsilon$







Exp #1:
$Y = \epsilon_Y$ and
$X_1 = 0, X_3 = 0$

Exp #2:
$Y = X_1 + \epsilon_Y$ and
$X_1 = \epsilon_1, X_3 = 0$

Knock-out do(X=0)

Exp #3:
$Y = \epsilon_Y$ and
$X_1 = 0, X_3 = Y + \epsilon_3$

Does this model make sense in all the experiments?

Observation:

#1 $Y = \varepsilon_y$ $X_3 = 0$

#2 $Y = X_1 + \varepsilon_y$ $X_3 = 0$

#3 $Y = X_1 + \varepsilon_y$ $X_3 = \varepsilon_3$

Regression Residuals:

$\boxed{\varepsilon_y} \sim 0$ #1

$\cancel{X_1} + \boxed{\varepsilon_y} \sim \cancel{X_1} + 0$ #2

$\cancel{X_1} + \cancel{\varepsilon_y} \sim X_1 + Y + \varepsilon_3$ **#3**

$\sim \cancel{X_1} + (X_1 + \cancel{\varepsilon_y}) + \varepsilon_3$

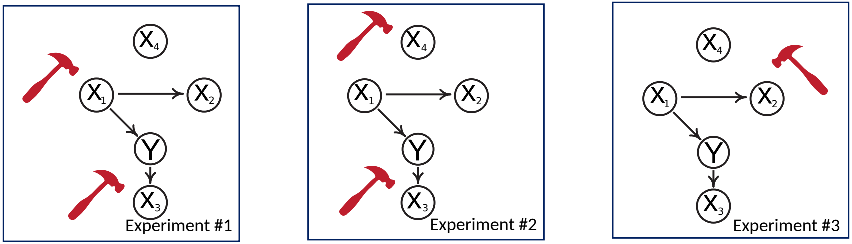
Invariance violated

Reject the experiment #3

→ Reject the invariance of X_3

Causal Structure Discovery by Invariance - 3

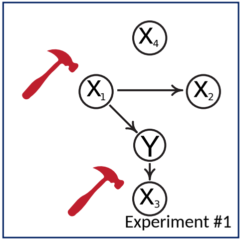
Testing hypothesis (model): $Y \sim X_2 + \epsilon$

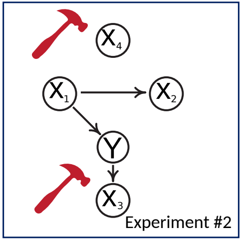


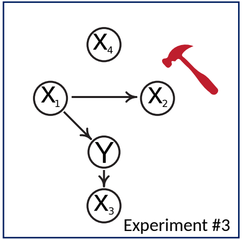
Knock-out do(X=0)

Causal Structure Discovery by Invariance - 3

Testing hypothesis (model): $Y \sim X_2 + \epsilon$







Exp #1:
$Y = \epsilon_Y$ and
$X_2 = \epsilon_2$

Exp #2:
$Y = X_1 + \epsilon_Y$ and
$X_2 = X_1 + \epsilon_2$

Knock-out do(X=0)

Exp #3:
$Y = \epsilon_Y$ and
$X_2 = 0$

Does this model make sense in all the experiments?

Observation:

#1 $Y = \varepsilon_y$ $\quad X_2 = \varepsilon_2$

#2 $Y = X_1 + \varepsilon_y$ $\quad X_2 = X_1 + \varepsilon_2$

#3 $Y = X_1 + \varepsilon_y$ $\quad X_2 = 0$

Regression Residuals:

$\boxed{\varepsilon_y} \sim \varepsilon_2$ #1

$\cancel{X_1} + \boxed{\varepsilon_y} \sim \cancel{X_1} + \varepsilon_2$ #2

$\boxed{X_1} + \boxed{\varepsilon_y} \sim \varepsilon_1$ #3

Invariance violated

Reject the experiment #3

$\rightarrow$ Reject the invariance of X_2

Summary

- Introduction to causal inference
- Causal structural model (DAG)
- Confounding factor adjustment
- Potential outcome model
- Causal DAG learning