INTRODUCTION TO CAUSAL INFERENCE

TOKYO DATATHON 2019 @ TMDU

Presented by Satoshi Kimura, Wei-Hung Weng, Ryo Uchimido

Nice to meet you !!!

Satoshi Kimura
MD, MPH
-Harvard T. H. Chan School of Public Health
-Department of Anesthesiology, Okayama University Hospital

Wei-Hung Weng
MD, MMSc, PhD
-MIT EECS and Computer Science and Artificial Intelligence Laboratory

Ryo Uchimido
MD, MPH
-Department of Emergency Medicine, Beth Israel Deaconess Medical Center

MACHINE LEARNING ALGORITHM IS PREDICTIVE, BUT NOT CAUSAL

Prediction is not enough supportive to make a decision in clinical setting

Prediction that identifies patients with heart failure does not assess whether heart transplantation is the best treatment option

Data science is science's second chance to get causal inference right: A classification of data science tasks

Miguel A. Hernán, John Hsu, Brian Healy

(Submitted on 28 Apr 2018 (v1), last revised 9 Oct 2018 (this version, v5))

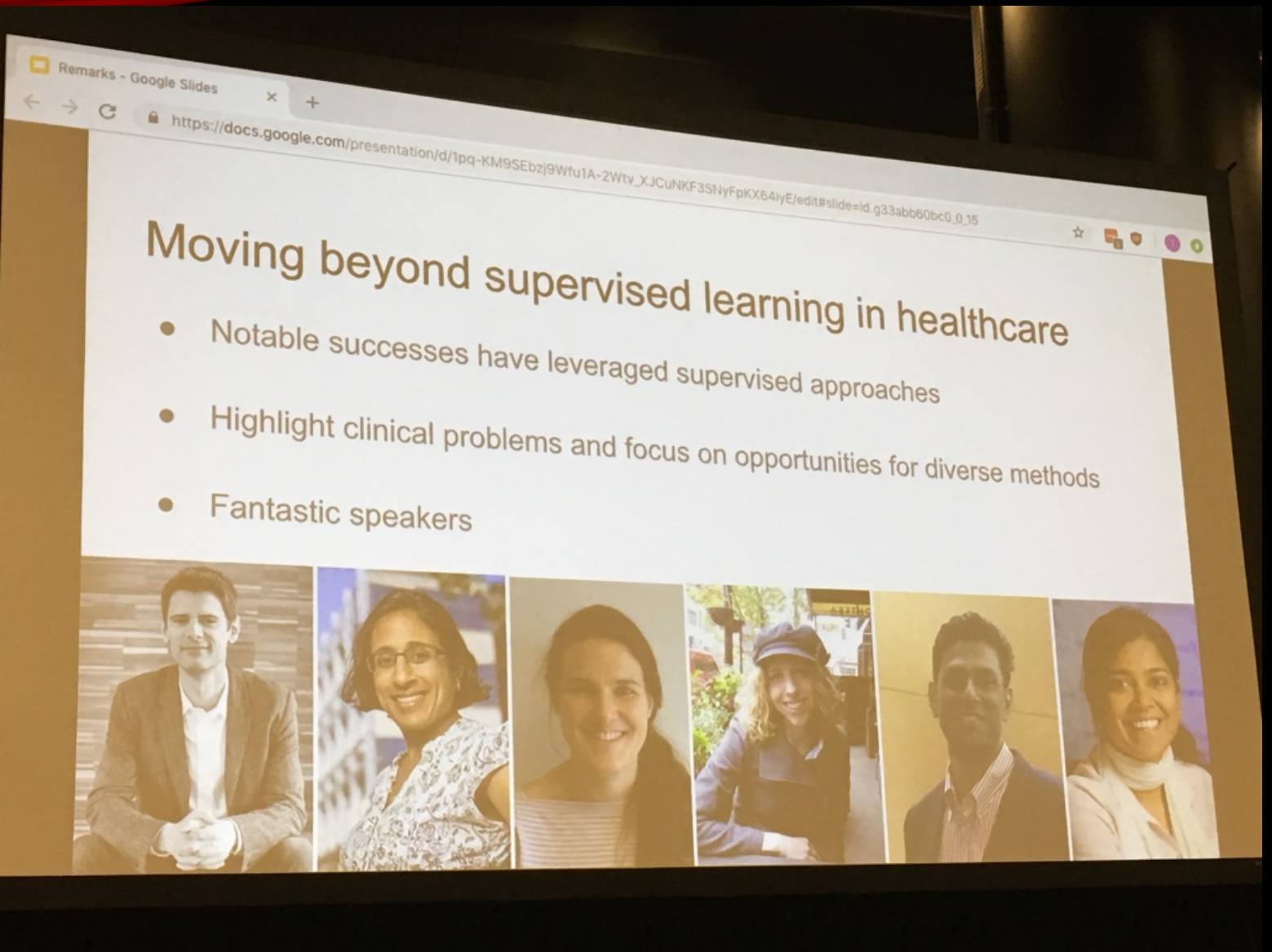
Causal inference from observational data is the goal of many data analyses in the health and social sciences. However, academic statistics has often frowned upon data analyses with a causal objective. The introduction of the term "data science" provides a historic opportunity to redefine data analysis in such a way that it naturally accommodates causal inference from observational data. Like others before, we organize the scientific contributions of data science into three classes of tasks: description, prediction, and causal inference. An explicit classification of data science tasks is necessary to discuss the data, assumptions, and analytics required to successfully accomplish each task. We argue that a failure to adequately describe the role of subject-matter expert knowledge in data analysis is a source of widespread misunderstandings about data science. Specifically, causal analyses typically require not only good data and algorithms, but also domain expert knowledge. We discuss the implications for the use of data science to guide decision-making in the real world and to train data scientists.

MACHINE LEANING MEETS CAUSAL INFERENCE

An episode from a workshop in Neuro IPS 2018/ Machine Learning for Health

ML4H: Machine Learning for Health

Workshop at NeurIPS 2018



Miguel Hernan

Finale Doshi-Velez
Barbara Engelhardt
Katherine Heller
Paul Varghese
Suchi Saria
Peter Schulam

Beyond supervised learning

Causa inference
- Time series data
- Modeling

AGENDA OF THIS LECTURE

1. Ice break - Ryo 5min
2. Lecture: Introduction to causal inference and DAG
   - Satoshi 60min
3. Break 10min – Make small groups
4. Hands on: Quantifying confounding and selection bias with ”dowhy” library in Python – Ryo/Wei-Hung 70min
5. Closing remarks - Ryo 5min

AT THE END OF THIS WORKSHOP YOU WILL BE ABLE TO

1. Lecture

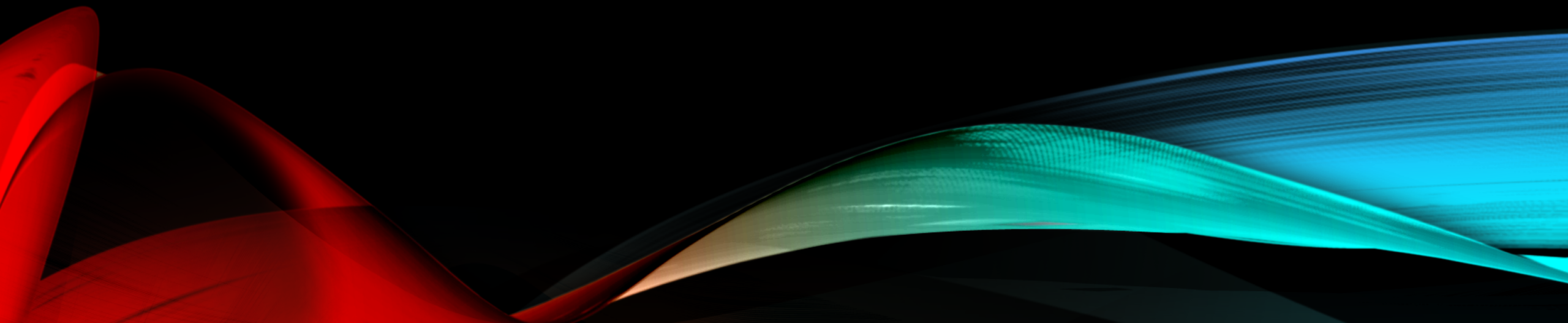
- Write your own DAGs
- Understand why just using computer is not enough to think about causal inference
- Understand why we should not adjust for all of the measured covariates

2. Hands on

- Understand how confounding and selection bias are induced


IF YOU WANT TO GO FAST, GO ALONE.
IF YOU WANT TO FAR, GO TOGETHER.

ANY QUESTION SO FAR?



TOKYO DATATHON 2019

Satoshi Kimura
Wei-Hung Weng
Ryo Uchimido

Introduction of Causal Inference
~Directed Acyclic Graphs~

Critical Data

Introduction of causal inference

- This is a lecture part of this workshop. We don't expect you to use any coding.

- We hope you understand the concept of DAG, which is the basis of **causal inference （因果推論）**.

- Our aim is to get you as much out of this workshop as possible, so stop us when you’re confused, because chances are other people are as well.

MIT Critical Data

Assess an effect of multivitamin use on birth defects

1. Multivitamin use（ビタミン摂取）
2. Maternal education（母の学歴）
3. Socioeconomic（社会経済的地位）
4. Psychologic characteristics（性格）
5. Family history of birth defect（近親者の先天異常）
6. Birth weight（子の体重）

Which variables are you going to put in your model?

At the end of this Lecture

You will understand

- How to write DAG

- Why just using computer is not enough to think about causal inference,

- Why we should not adjust for all of the measured covariates

Causal inference vs. prediction

Causal inference (主に疫学による“因果推論”)

- Asking what would happen to an outcome as a result of a treatment or intervention.
- The aim is to determine whether a particular independent variable (cause) really affects the outcome and to estimate the magnitude of the effect, if any.

Prediction (主に統計学による“予測”)

- Making comparisons between an outcome across different combinations of values of input variables.
- The goal is to develop a formula for predicting the outcome better.

MIT Critical Data

Epidemiology (疫学) ≠ Statistics (統計学)

Epidemiology	Statistics
Using expert knowledge before analysis	Analytical approach
Causal effect	Association
Product term	Interaction term
Effect measure modification	Interaction term
Stratification of (within levels of ~)	Adjusting for …
Propensity score matching	Adjusting for …
Standardization	Adjusting for …
. . .	. . .

Causal inference（因果推論）

Step 1

- Create a causal DAG based on your expert knowledge

Step 2

- Identification

  : Assessing Identifiability (exchangeability, positivity, consistency)

Step 3

- Estimate the average treatment effect from your observed data by statistical models (using data)

MIT Critical Data

Causal inference（因果推論）

Step 1

- Create a causal **DAG** based on your expert knowledge

Step 2

- Identification

  : Assessing Identifiability (**exchangeability**, positivity, consistency)

→**Decide which variables we should adjust for**

Step 3

- Estimate the average treatment effect from your observed data by statistical models (**using data**)

DAG

- **D**irected **A**cyclic **G**raphs

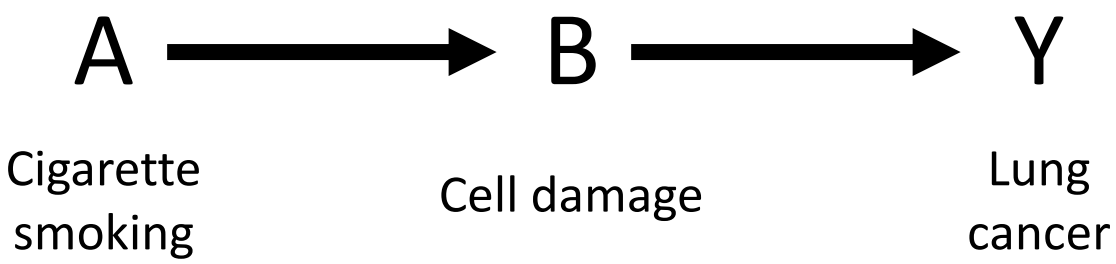
= Causal diagrams = Causal graphs

- Draw your <u>assumptions</u> using expert knowledge from you or your collaborators **before analysis**
  (解析前の「仮定」が重要！)

DAG

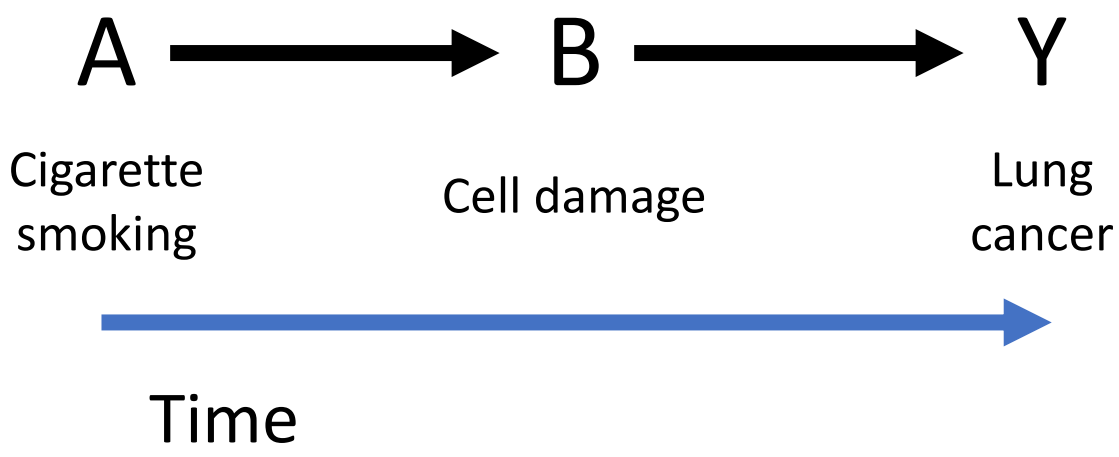
- Directed Acyclic Graphs
  (矢印には方向がある)

- Arrows indicate the direction of causality.
- B causes Y
- Y does not cause B
  (矢印は因果関係を表す)



DAG

- Directed Acyclic Graphs
- There are no cycles.
  (時間は左から右へ)
  (DAGは循環しない)



- If you start at one variable and follow the direction of the arrows, you will never get back to the same variable.

Cf. “Association”

How to find an association between A and Y

1. Smoking and lung cancer are associated if the proportion of individuals with cancer is different among smokers and nonsmokers.
2. A and Y are associated when having information about A allow us to predict Y better on average.

A

Cigarette smoking

vs.

Y

Lung cancer

Causal inference（因果推論）~example~

Step 1. We draw the following DAG (your assumption).

Step 2. We do not think we need any adjustment (in this case).

Step 3. We find an association between smoking and lung cancer (using data).

↓

Then, we think that there is an causal effect of smoking on lung cancer.

Cigarette smoking

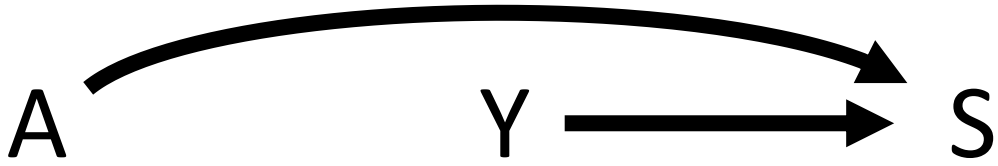
Lung cancer

MIT Critical Data

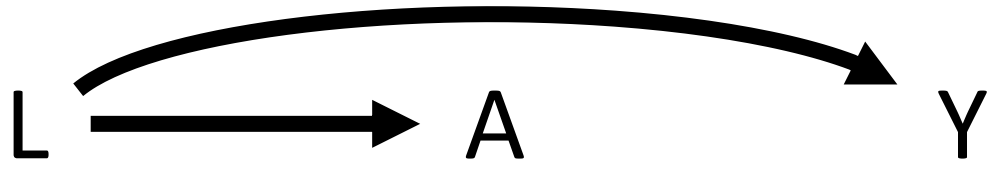
Three basic DAGs
~ Checking causal effect of A on Y ~

1. Intermediator

2. Collider



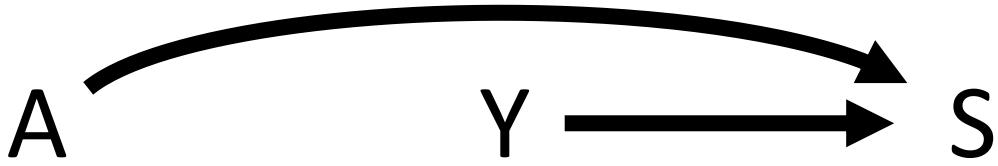
3. Confounder



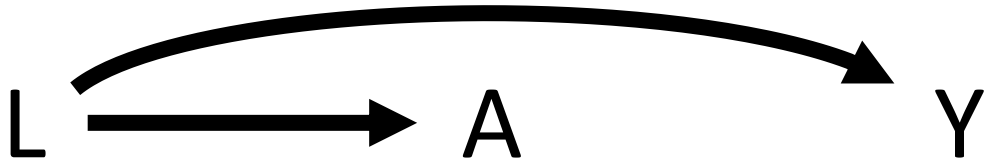
Three basic DAGs
~ Checking causal effect of A on Y ~

1. Intermediator (B)
（介在因子）

2. Collider



3. Confounder



MIT Critical Data

Intermediator

Step 1 and 2

- You need to draw your DAG based on your expert knowledge.
- (Surely, there might be multiple DAGs according to different knowledge and assumptions.)
- Decide whether you should adjust

A

Cigarette smoking

B

Cell damage

Y

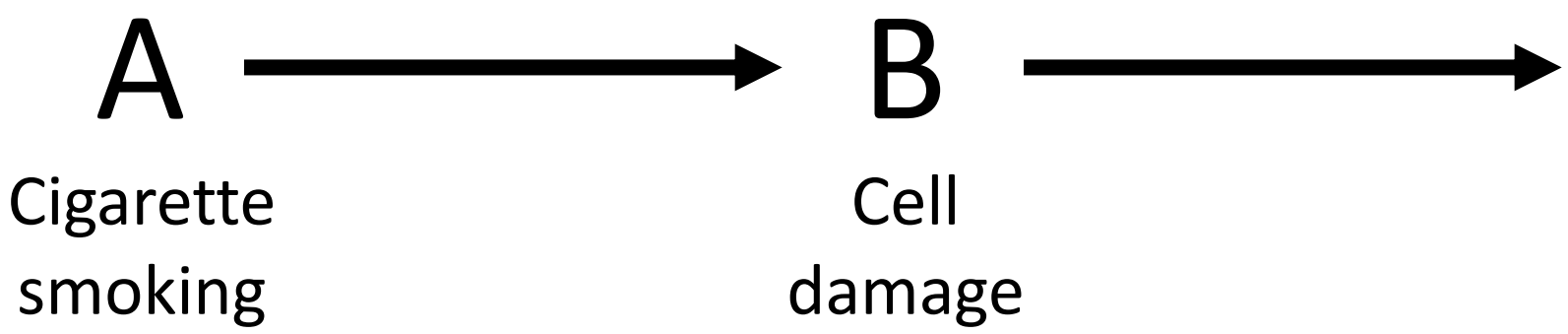
Lung cancer

Intermediator

Step 3

Do we find an association between A and Y?

1. A and Y are associated if the proportion of individuals with Y is different among people with and without A.
2. A and Y are associated when having information about A allow us to predict Y better on average.

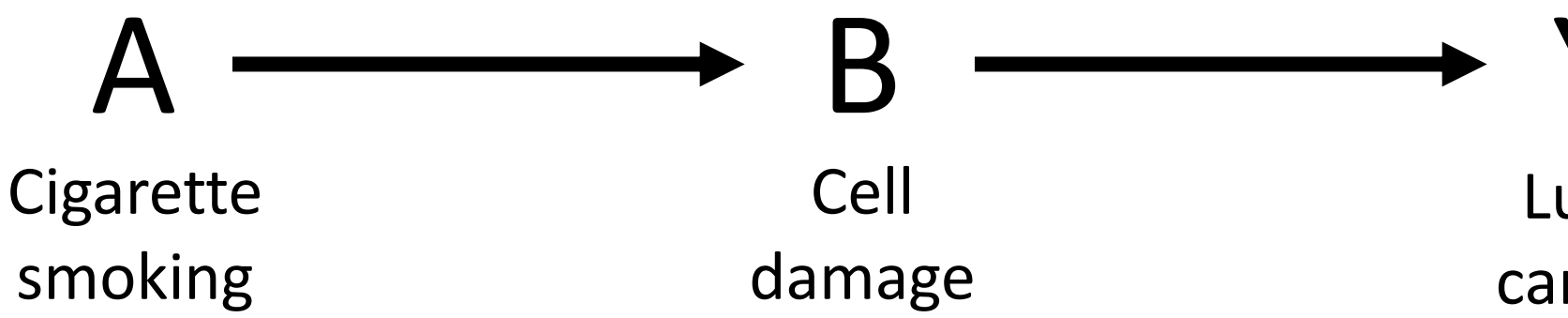


Intermediator

Then, we think that there is an causal effect of smoking on lung cancer.

i.e. “Cigarette smoking causes lung cancer.”

What if we adjust for B (cell damage), skipping step 1 and 2??



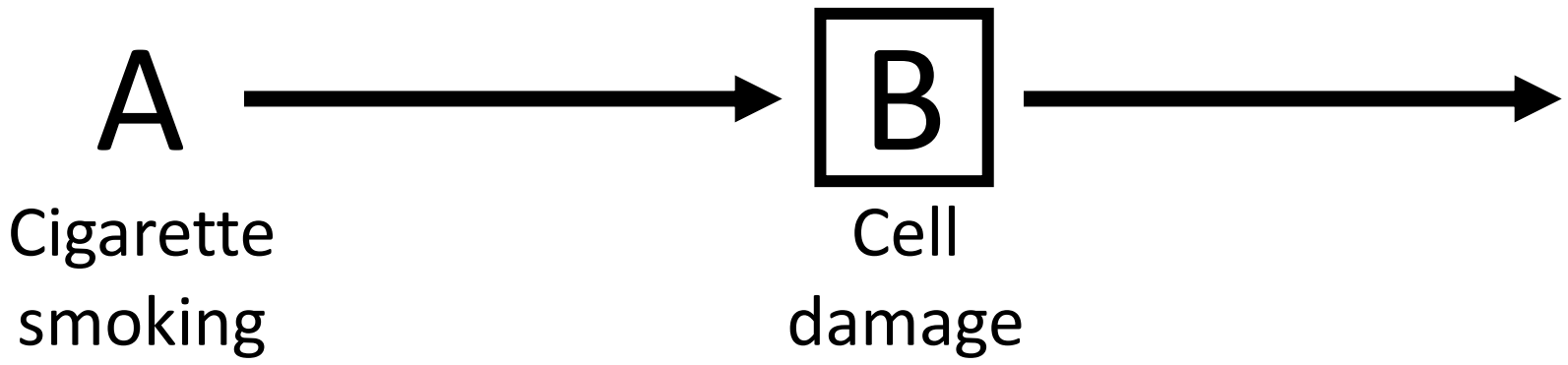
Conditioning （条件付け） in DAG

- To restrict our analysis to a subset of individuals and check whether there is an association between A and Y.

: Are A and Y associated conditional on B or within levels of B?

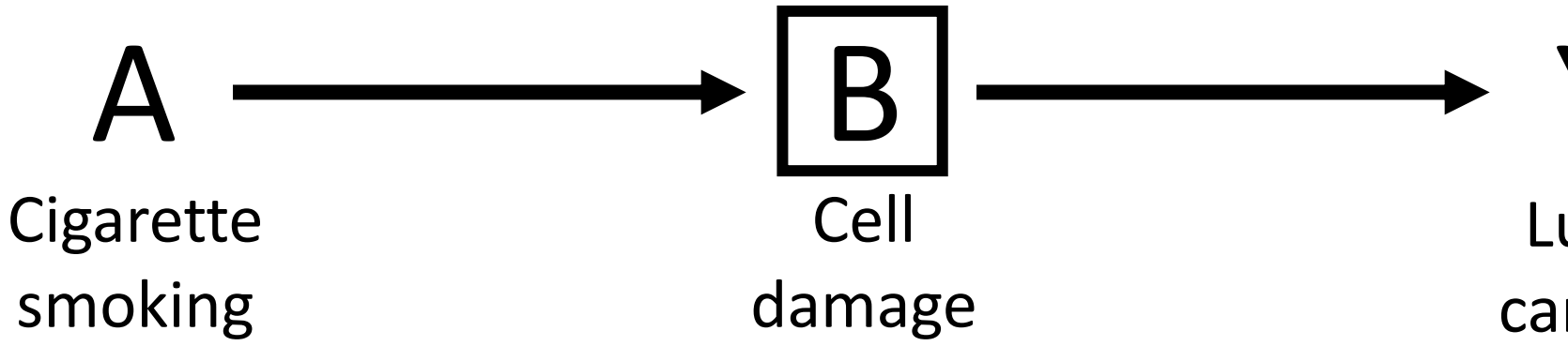
- To represent it graphically that we are conditioning on a particular value of B, we put a square box around B on the graph.

（囲まれている変数の分類毎に分析を行う）



Intermediator

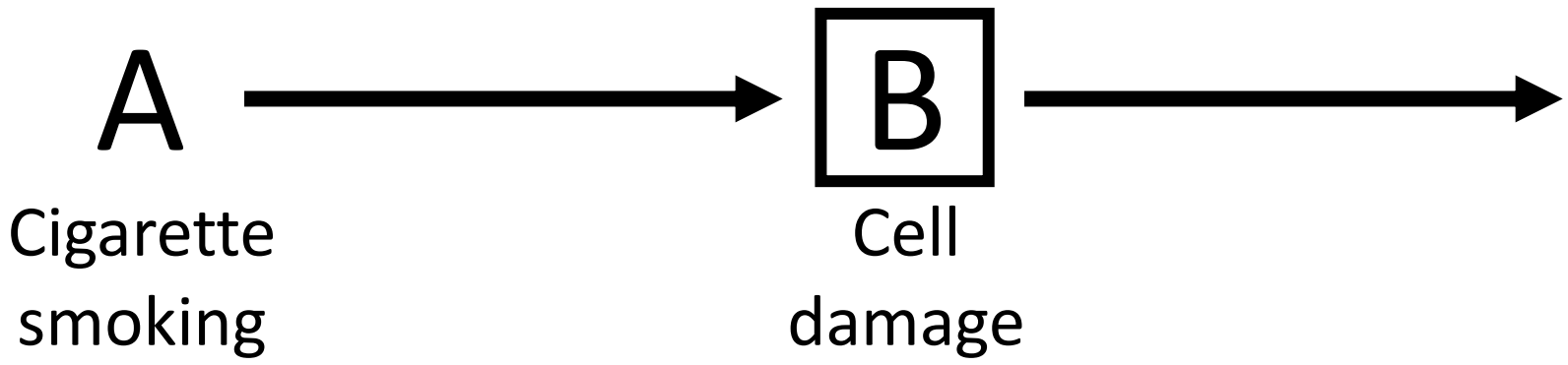
- Among people with cell damage, the proportion of individuals with cancer is not different between smoker and non-smoker.
- If someone has cell damage, then learning that she is a smoker does not provide any additional information with respect to the risk of Y.
- We say that 1) there is no conditional association between A and Y with the levels of B or 2) A and Y are independent conditional on B.



Rule①

- The flow of association between A and Y is interrupted when we condition on the mediator B, even though A has a causal effect on Y.

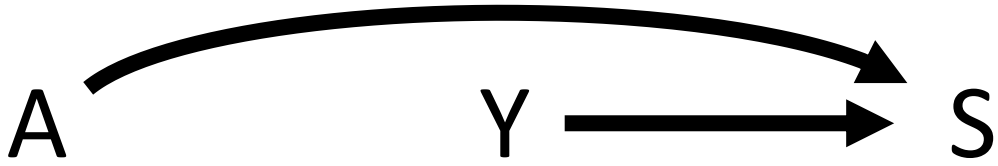
- We usually do not adjust for intermediate variable.



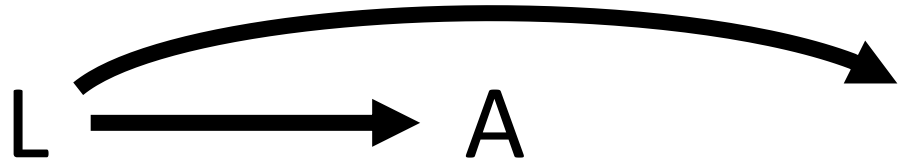
Three basic DAGs
~ Checking causal effect of A on Y ~

1. Intermediator

2. Collider (S)



3. Confounder



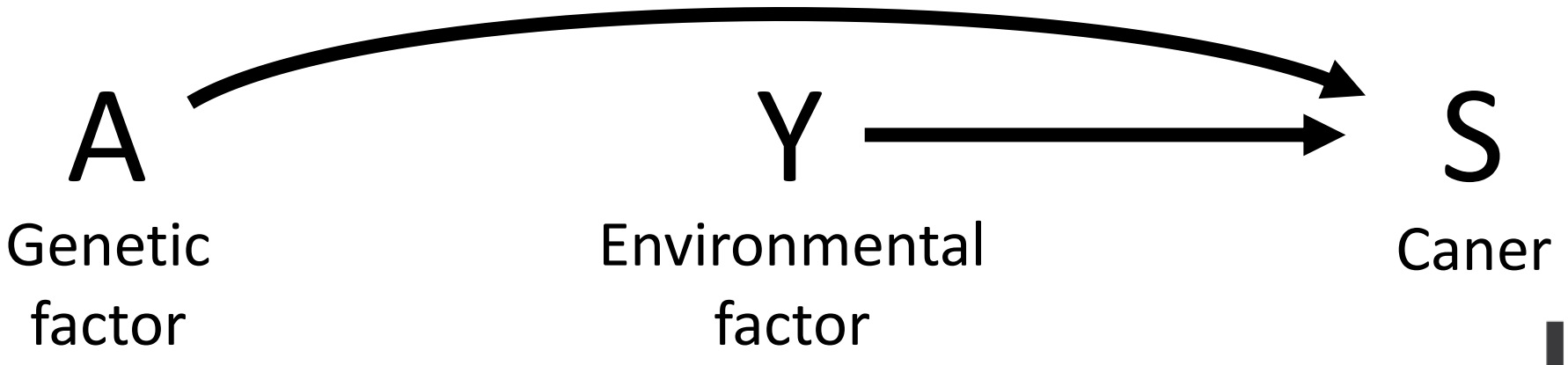
MIT Critical Data

Common effect

Step 1 and 2. Draw your DAG and decide whether you should adjust

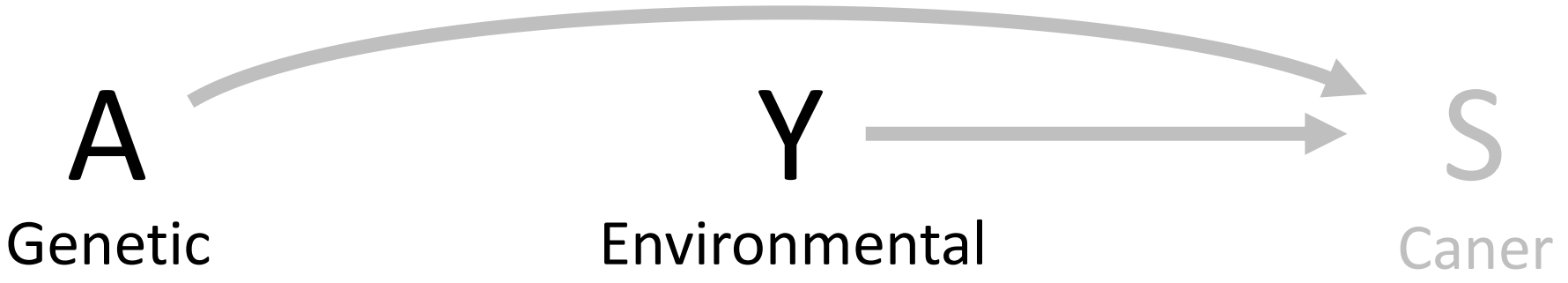
Step 3. Do we find an association between A and Y?

- Both the genetic and the environmental factor may cause cancer in the future.
- The common effect L is referred to as a **collider**.



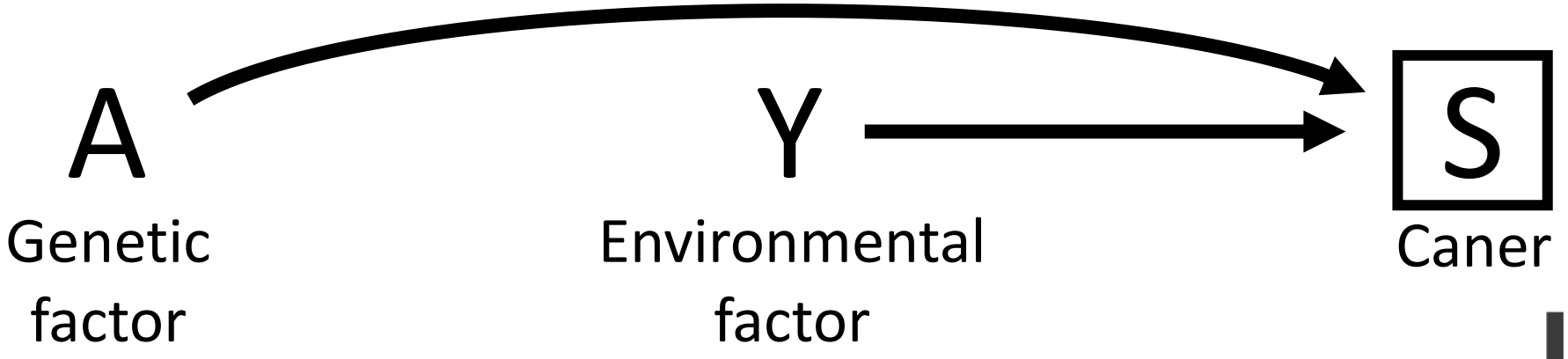
Common effect

- An individual's exposure to high levels of air pollution was not affected by her genes.
- The environmental factor is distributed in the population independently.
- The presence of a common effect of A and Y does not create an association between A and Y.
- What if we adjust for S (cancer)?? (i.e. conditioning on collider)



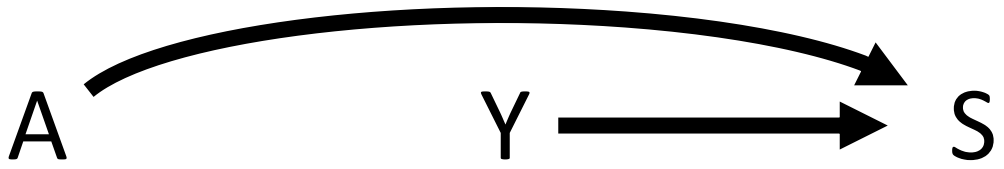
Conditioning on collider

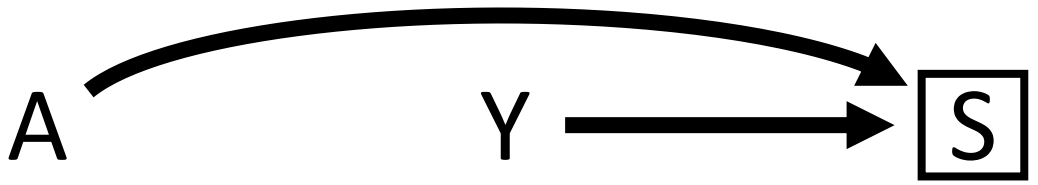
- Someone has cancer.
- Now we learn the person does not have the genetic factor.
- Because she has cancer, something must have caused the cancer.
- One of those causes is the environmental factor.
- When we restrict the analysis to people with cancer, the proportion of people with the environmental factor is higher among people without the genetic factor than among people with the genetic factor.
- There is **an conditional association** between A and Y, even though A has not causal effect on Y (**selection bias**).



Rule②

- Colliders block the flow of association along the path that they lie on.
- Conditioning on S creates a flow of association (opens path) between A and Y.
- We do not want to adjust for a collider.

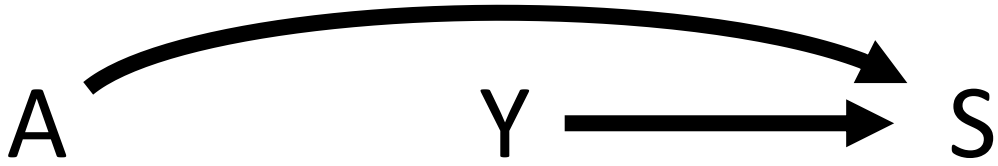




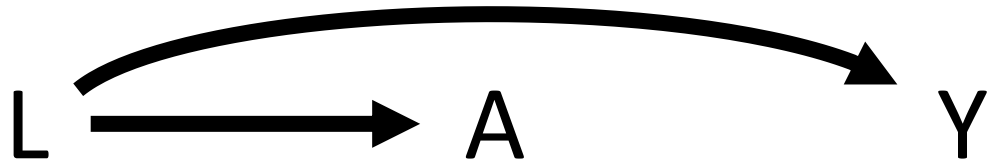
Three basic DAGs
~ Checking causal effect of A on Y ~

1. Intermediator

2. Collider



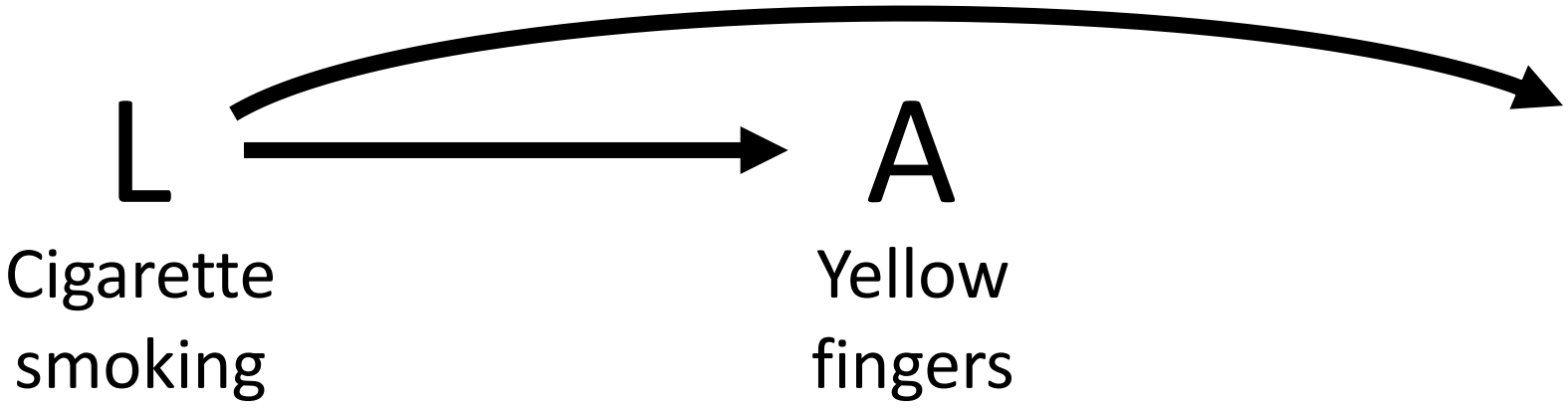
3. Confounder (L)
（交絡因子）



Confounder

Step 1 and 2. Draw your DAG and decide whether you should adjust

Step 3. Do we find an association between A and Y?



Confounder

- People with yellow fingers are more likely have lung cancer than people without yellow fingers.
- Having information about A allows us to predict Y better on average.

→There is an association (good predictor!).

- That is not because yellow fingers cause lung cancer, it is because having yellow fingers is a marker of smoking, which causes lung cancer.
- This association is a **bias**.

Cf. Path and backdoor path

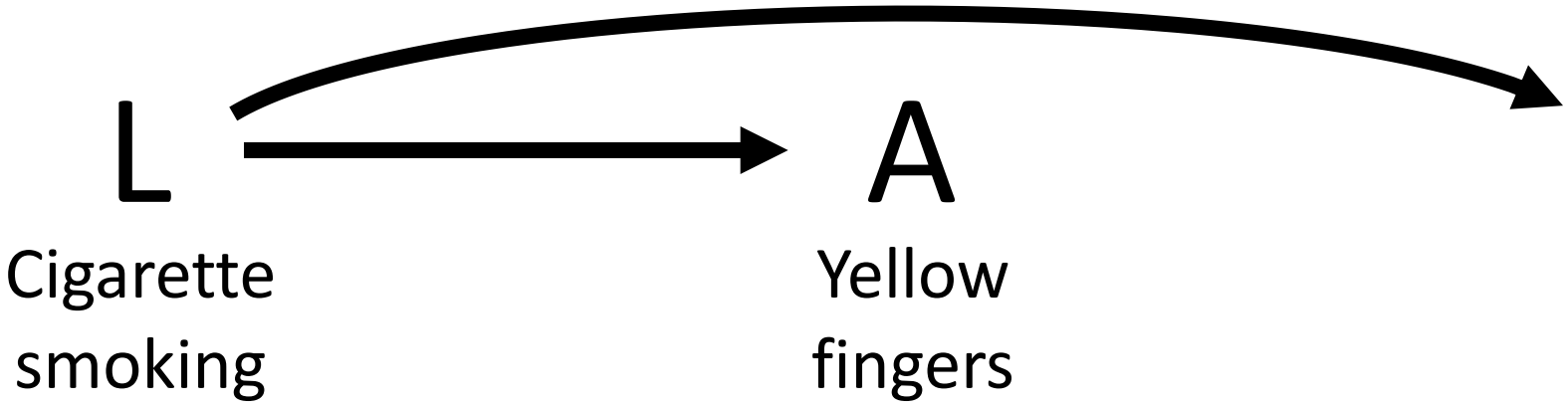
- A path is any arrow-based route between two variables on the graph.

- A backdoor path between A and Y is a path that connects A and Y without using any of the arrows that leave from A.

Confounder

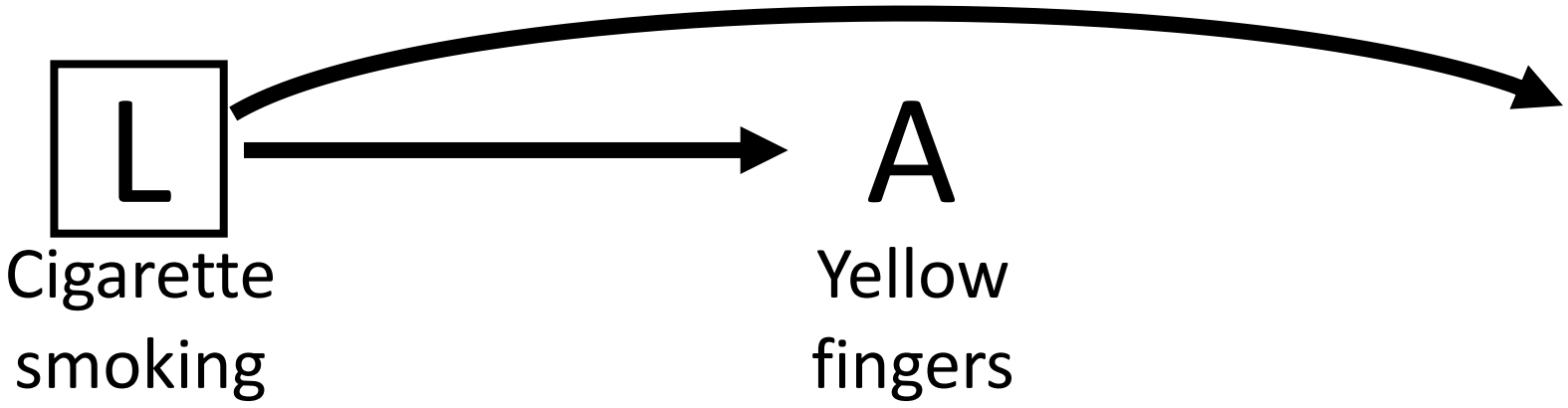
- When there is a component of the association between A and Y that is due to a common cause of A and Y (i.e. L), we say that is confounding.
- One of the most important goals of causal inference is **to eliminate bias due to confounding**.

What can we do?



Conditioning on Confounder

- We will now consider the conditional association between A and Y within levels of L.
- We can restrict the analysis to the subset of individuals who are never smokers.



Crude analysis (before conditioning)

	Lung cancer		Total	Cumulative incidence
	+	-		
Yellow fingers +	420	180	600	420/600
Yellow fingers -	240	810	1050	240/1050

Risk ratio = (420/600) / (240/1050) = 3.06

Does having Yellow fingers cause lung cancer?

MIT Critical Data

Conditioning on confounder

	Lung cancer		Total
	+	-	
Yellow fingers +	420	180	600
Yellow fingers -	240	810	1050

Smoking (-)

	Lung cancer		Total
	+	-	
Yellow fingers +	20	80	100
Yellow fingers -	200	800	1000

Risk ratio = (20/100) / (200/1000) = 1

Smoking (+)

	Lung cancer		Total
	+	-	
Yellow fingers +	400	100	500
Yellow fingers -	40	10	50

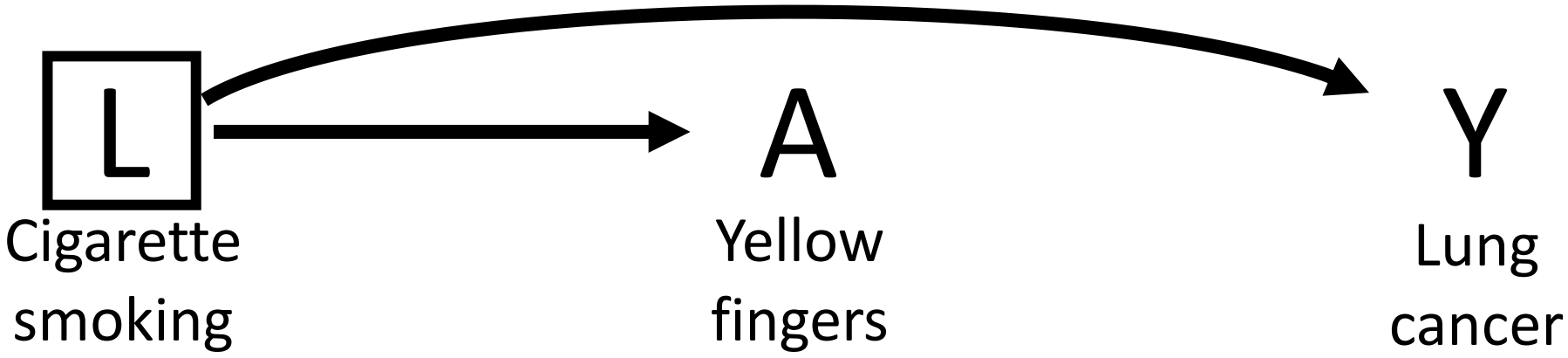
Risk ratio = (400/500) / (40/50) = 1

Rule③

- Flow of association between A and Y (backdoor path from A to Y) is interrupted when we condition on their common cause L.

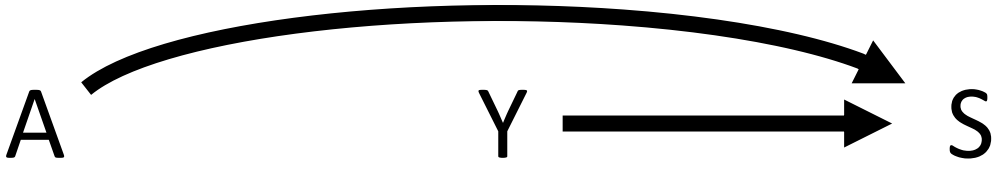
=The box around L blocks the association between A and Y.

- We want to adjust for confounder.

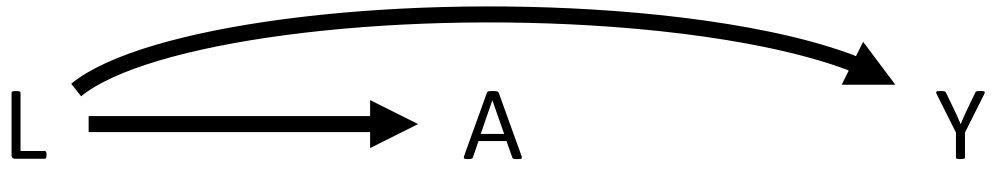


Summary of DAG

- The flow of association between A and Y is interrupted when we condition on the mediator B, even though A has a causal effect on Y.



- Conditioning S creates a flow of association between A and Y.



- Flow of association between A and Y is interrupted when we condition on their common cause L.

MIT Critical Data

Cf. Definition of confounding/confounder

Non-structured

- "Change in estimate" definition: Only using data

- "Conventional" definition: Using DAG

Structured approach

- Only using DAG

“Change in estimate” definition

- Sill popular but not good for causal inference
- A variable is a confounder of the effect of A on Y if adjusting for it alters the association between A and Y.
- If the measure associations are different, then L is a confounder.

ex. Adjusting for L changes the risk ratio by more than 10%

ex. Effect of smoking on baby’s weight: (200-150) > 200*0.1)

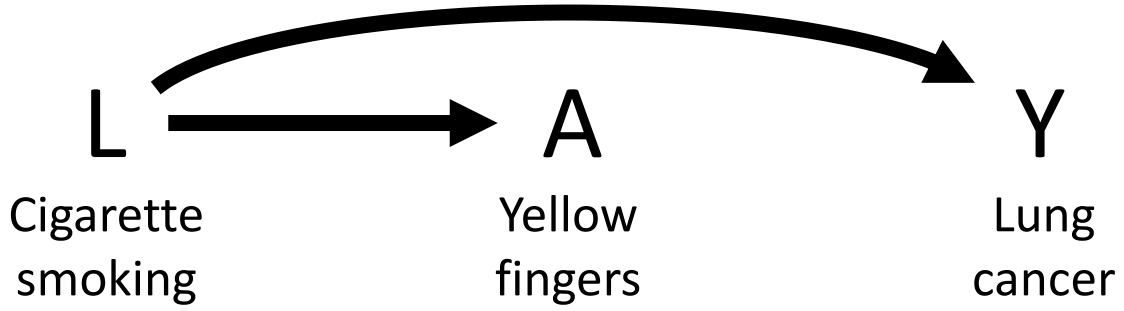
<Before>

Covariate	β coefficient (se)	P-value
Intercept	2800	<0.001
Smoking	**-200** (20)	0.0001

<After>

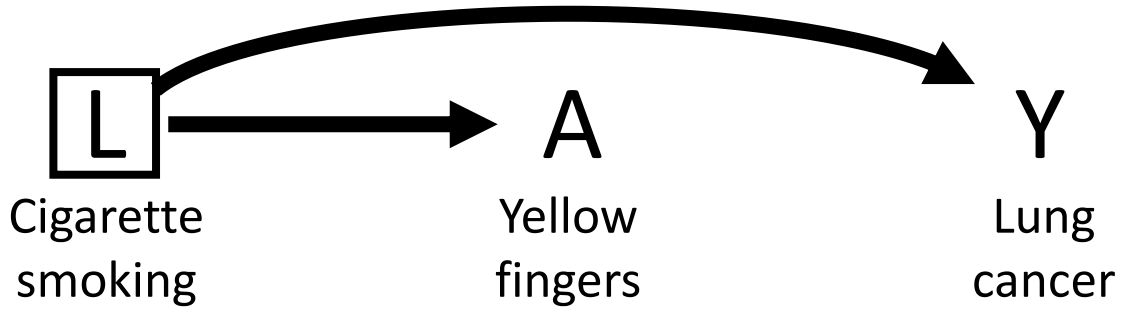
Covariate	β coefficient (se)	P-value
Intercept	2920	<0.0001
Smoking	**-150** (25)	0.02
Drinking alcohol	-80	0.06

“Change in estimate” definition



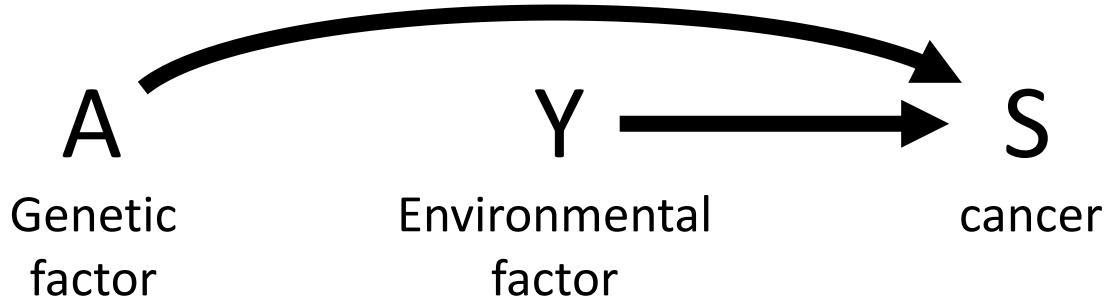
- We expect an association between A and Y.

- The association disappears if we condition on L.

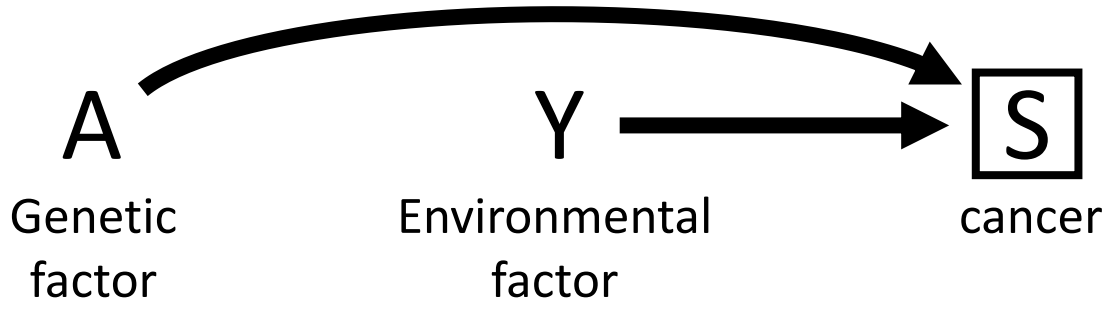


- L is a confounder.

“Change in estimate” definition



- There is no association between A and Y.
- If we adjust for al, then the association measure changes from zero association to non-zero association.



- Adjustment for L introduces bias (selection bias).

MIT Critical Data

Structured approach

- Best way for causal inference

- Are there any common causes of A and Y?

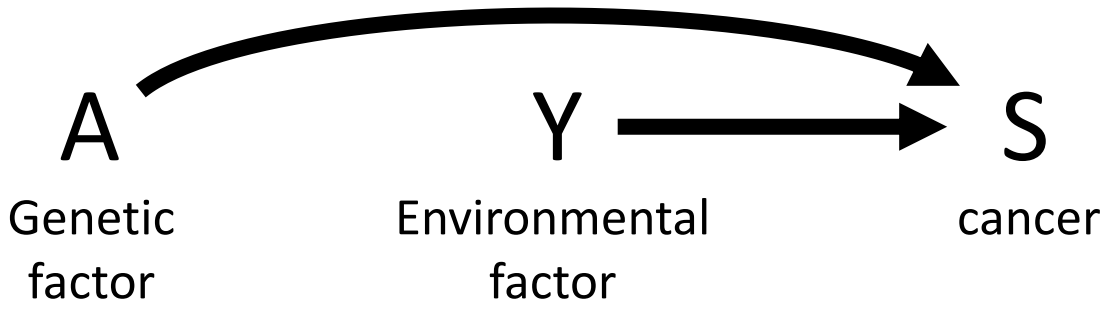
= Are there any backdoor paths between A and Y?

→If yes, then there is confounding.

- Can the backdoor path between A and Y be blocked by conditioning on some of the measured variables?

→If yes, then we need to block adjust for the variable.

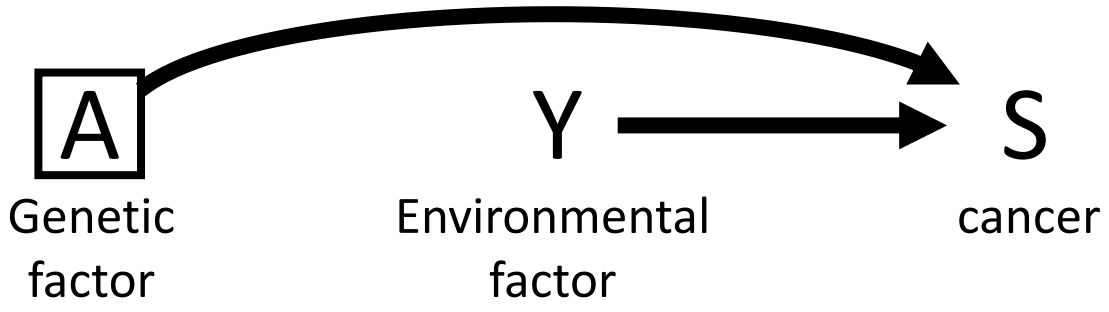
Structured approach



- Are there any common causes of A and Y?

→No.

Then there is no confounding!



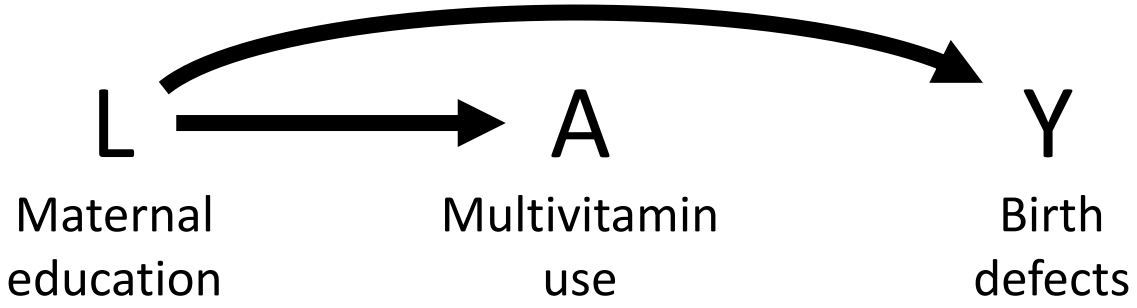
Any question?

Q. Assess an effect of multivitamin use on birth defects

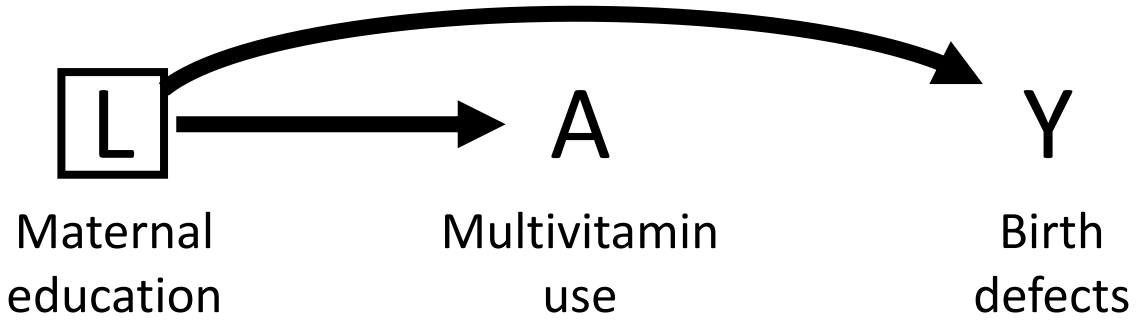
1. Multivitamin use（ビタミン摂取）
2. Maternal education（母の学歴）
3. Socioeconomic（社会経済的地位）
4. Psychologic characteristics（性格）
5. Family history of birth defect（近親者の先天異常）
6. Birth weight（子の体重）

Which variables are you going to put in your model?

Q. Maternal education

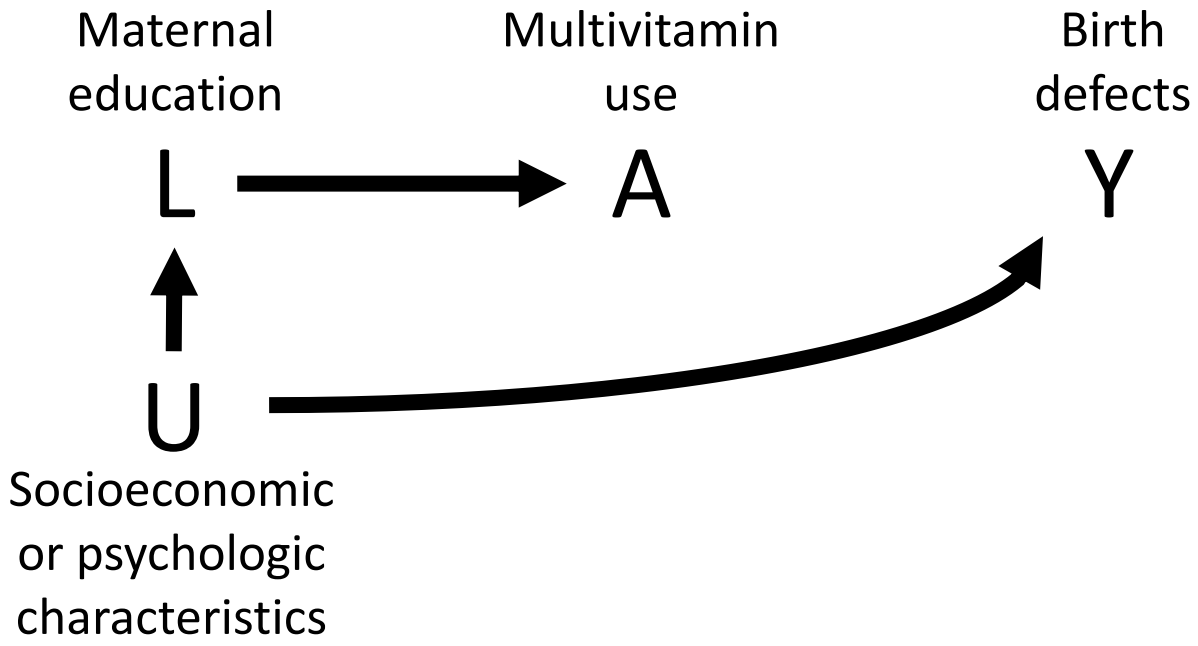


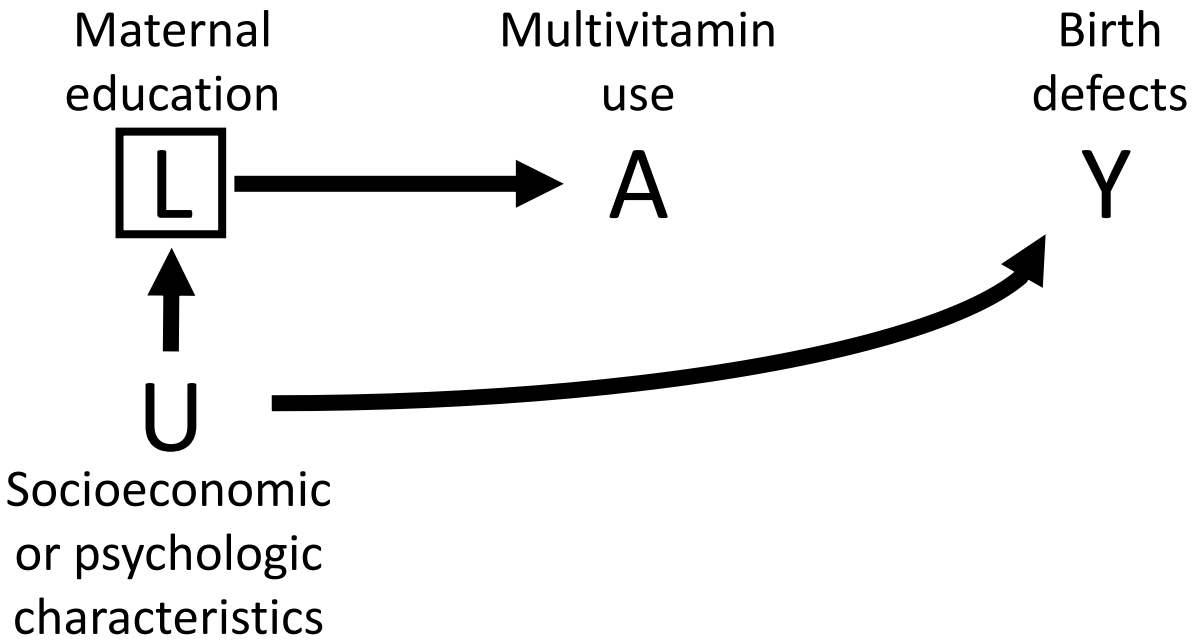
- Low maternal education affects the use of multivitamins.
- Low maternal education, through multiple pathways, increases the risk of birth defects.



- We could adjust for L because conditioning on L block the backdoor path between A and Y.

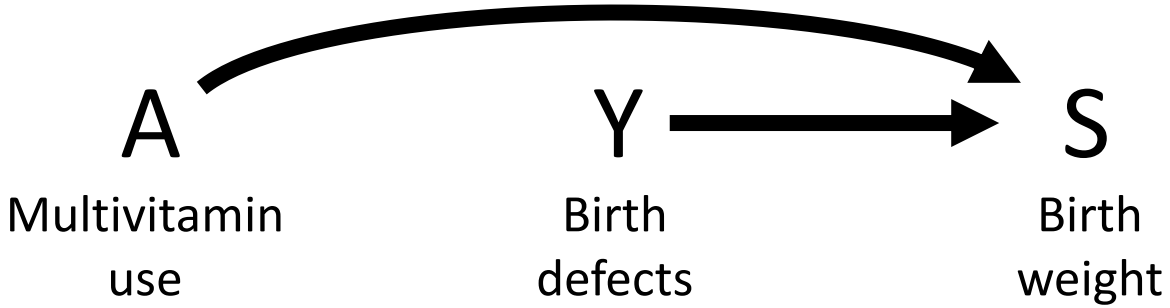
MIT Critical Data

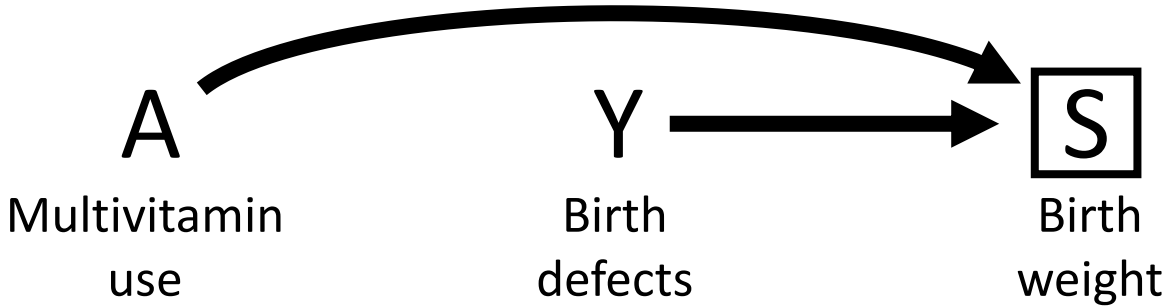




- Maternal education is a result of unmeasured socioeconomic or psychological characteristics U that directly affect defects through high risk behaviors.
- We would also adjust for L because conditioning on L blocks the backdoor path between A and Y.
- Same as family history of birth defect or maternal diet.

Q. Birth weight





- S cannot possibly be a common cause of A and Y because S happens after A and Y.
- Birth weight can be affected both by use of multivitamin use and by the presence of a birth defect.
- We would not adjust for S because conditioning on S does not block any backdoor path between A and Y.
- Conditioning on the collider S opens a non-causal path, introducing selection bias.

MIT Critical Data

We have learned

- What DAG is.

- Three basic DAG
  1) Intermediator
  2) Collider
  3) Confounder

- Which variable we should adjust for.

Take home message

In order to assess causal inference,

- Draw DAG before analysis by using expert knowledge to determine if we should adjust for a variable.

- Statistical criteria (ex. Forward selection, stepwise selection etc) are insufficient to characterize confounding and confounders and to determine their adjustments.

Reference

1. Causal Diagrams Lesson. EPI201, EPI289. HSPH.

Thank you

skimura@hsph.harvard.edu

Critical Data

CLOSING REMARK

WHAT DID YOU LEAEN IN THIS WORK SHOP?

- I appreciate if you sharing your thoughts

- The slides and codes will be shared in GitHub

WHAT WE DIDN'T COVER TODAY

- Counter factual model
- Exchangeability, Consistency, and Positivity -> Identifiability
- The difference between Conditioned treatment effect and Marginal treatment effect
- Other methodology for conditional exchangeability
  - Propensity score analysis
  - Inverse Probability Weighting (non-stabilized and stabilized) for conditional exchangeability
    - For confounding
    - **For censoring**
  - Standardization and G-formula
  - Marginal Structure Model
  - Instrumental Variable

ADVANCED TOPICS OF CAUSAL INFERENCE

- Time dependent treatment effect estimation
  - Time varying confounder
  - Time series data

- New modeling methodology for the effect estimation based on Machine learning
  - Super learner

Please keep in touch with us !!!

Satoshi Kimura
kimsato1034@hotmail.co.jp

Wei-Hung Weng
ckbjimmy@mit.edu

RYO UCHMIDO
ruchimd@bidmc.Harvard.edu

We are happy to do a lecture in Japan !!!
Please Let Us Know If You're Interested