STAT6061/STAT5008 – Causal Inference

Part 1-1. Introduction to Causal Inference

An-Shun Tai

[1]*Department of Statistics*
National Cheng Kung University

[2]*Institute of Statistics and Data Science*
National Tsing Hua University

Prerequisites

- Basic knowledge of probability theory
- Understanding of statistical inference
- Familiarity with linear regression
- Familiarity with logistic regression
- Experience with R programming
- Basic knowledge of survival analysis (*optional*)

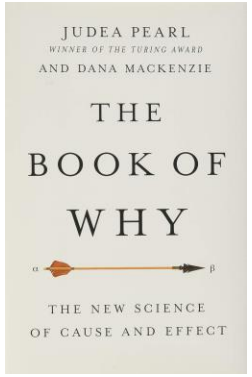
Textbooks

- No specific textbook:

The material is primarily based on lecture notes and various papers.

- Highly recommended readings:

1. Ding, P. (2024). *A First Course in Causal Inference.*
2. Hernán, M. A. & Robins, J. M. (2020). *Causal Inference: What If.*
3. Brumback, B. A. (2021). *Fundamentals of Causal Inference: With R.*
4. Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect.* Basic books.



Does smoking cause lung cancer?

(Doll and Hill, 1950)

TABLE IV.—*Proportion of Smokers and Non-smokers in Lung-carcinoma Patients and in Control Patients with Diseases Other Than Cancer*

Disease Group	No. of Non-smokers	No. of Smokers	Probability Test
Males:			
Lung-carcinoma patients (649)	2 (0·3%)	647	P (exact method) = 0·00000064
Control patients with diseases other than cancer (649) ..	27 (4·2%)	622	
Females:			
Lung-carcinoma patients (60)	19 (31·7%)	41	$\chi^2 = 5{\cdot}76$; n = 1 0·01 < P < 0·02
Control patients with diseases other than cancer (60) ..	32 (53·3%)	28	

When Genius Errs: R. A. Fisher and the Lung Cancer Controversy

(Stolley, 1991)

1. If A is associated with B, then not only is it possible that A causes B, but it is also possible that B is the cause of A. In other words, smoking may cause lung cancer, but it is a logical possibility that lung cancer causes smoking.

2. There may be a genetic predisposition to smoke (and that genetic predisposition is presumably also linked to lung cancer).

3. Smoking is unlikely to cause lung cancer because secular trend and other ecologic data do not support this relation.

4. Smoking does not cause lung cancer because inhalers are less likely to develop lung cancer than are noninhalers.

When Genius Errs: R. A. Fisher and the Lung Cancer Controversy

(Stolley, 1991)

1. **Reverse causation is possible.** However, he provided no data to support this speculation.

2. **Genetic predisposition to smoking.** However, his evidence was weak and based on poorly described twin studies.

3. **Epidemiological trends do not support the link.** However, he failed to compare lung cancer rates between smokers and non-smokers properly.

4. **He pointed out that smokers who inhaled had lower lung cancer rates than those who did not inhale, based on early data from Doll and Hill**. However, he ignored later studies that contradicted this claim and used misleading statistical techniques to exaggerate the effect.

Aphorisms on causal inference in statistics

"Correlation does not imply causation"

"You CANNOT prove causality with statistics"

- These aphorisms are generally true, but advances in causal inference have shown that causation can be inferred from association under specific assumptions.

- In this course, you will learn the "**formal language of causal inference**" and how to "**apply statistical methods to estimate causal effects**" in randomized experiments and observational studies.

The first question in causal inference

"How does causal inference differ from association inference?"

1. **The conditions necessary for valid inference**
2. **Role in data science**
3. **Requirement for intervention**

Three necessary conditions for causal inference

(Shaughnessy, Zechmeister, and Zechmeister, 2000)

1. **Covariation of events**

A statistically significant relationship exists between the cause and effect.

2. **A time-order relationship**

The cause must occur before the effect.

3. **The elimination of plausible alternative causes**

Other potential factors are ruled out, ensuring the observed relationship isn't due to confounding variables.

Key Difference 1: Conditions for Inference

Causal inference succeeds only when all three conditions are met, whereas establishing an association requires only evidence of covariation (the first condition).

Data science task: To explain or to predict?

(Hernán, Hsu, and Healy, 2019; Shmueli, 2010)

- **Description**

Using data to provide a quantitative summary of certain features of the world.

- **Prediction (Inference of association)**

Using data to map some features of the world to other features of the world.

- **Counterfactual prediction (Causal inference)**

Using data to predict certain feature of the world if the world had been different.

Key Difference 2: Roles in Data Science

"Causal inference" and "Inference of association" serve fundamentally different roles in data science.

Example of lung cancer studies

- **Description**

How can lung cancer patients be grouped into distinct classes based on their individual clinical and demographic characteristics?

- **Prediction (Inference of association)**

What is the predicted one-year survival rate for lung cancer patients with specific profiles?

- **Counterfactual prediction (Causal inference)**

Does initiating smoking, on average, increase the risk of mortality among patients with those characteristics?

The causal effect of interventions

➢ **Interventionist definition of causation**

A variable **A** causes **Y** if and only if changing **A** leads to a change in **Y**, while holding all other factors constant.

➢ **Example: Front Yard vs. Back Yard**

- The front yard and back yard are always wet or dry at the same time, showing an association.
- However, they do not have a causal relationship because intervening in the status of the front yard (e.g., covering it with a tarp) does not change the condition of the back yard.

Key Difference 3: Intervention Requirement

Causal inference necessitates actively manipulating or intervening in the treatment/exposure to assess its impact, whereas association analysis relies solely on observational data.

Association and Causation

- The underlying conceptual distinctions between association and causation lie in

1. the conditions necessary for valid inference,
2. their distinct roles in data science, and
3. the requirement for intervention.

- From a statistical perspective, what distinguishes association from causation?

Association and Causation

- Comparing two linear regression results of toy examples

Case 1: Y~A

```
Call:
lm(formula = Y ~ A)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3502 -0.8277 -0.0003  0.8274  5.7833

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0007395  0.0012249  -0.604    0.546
A            0.4999218  0.0008669 576.698   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Case 2: Y~A+C

```
Call:
lm(formula = Y ~ A + C)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1409 -0.4741 -0.0014  0.4715  3.0196

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept)  0.0008452  0.0022324    0.379    0.705
A           -0.4992008  0.0027331 -182.648   <2e-16 ***
C            0.5005031  0.0015844  315.894   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

A significant association exists between A (treatment) and Y (outcome).

- Key Questions:

- Does statistical significance in both cases indicate a causal relationship?
- If not, what could explain the observed associations in each case? (*Simpson's paradox*, *Berkson's paradox*)

Simpson's paradox

- **Edward H. Simpson formally described this phenomenon in a 1951 technical paper, though Karl Pearson (1899) and Udny Yule (1903) had observed similar effects earlier.**

- **Example of Simpson's paradox: Kidney stone treatment**

- The paradoxical result is that treatment A is more effective for both small and large stones individually, yet treatment B appears more effective when both sizes are considered together.

- **How could it be true?**

Table. *Success rat for kidney stone treatment*

Stone Size	**Treatment**	
	Treatmen A	Treatment B
Small kidney Stones	93% (81/87)	87% (234/270)
Large kidney Stones	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)

Simpson's paradox

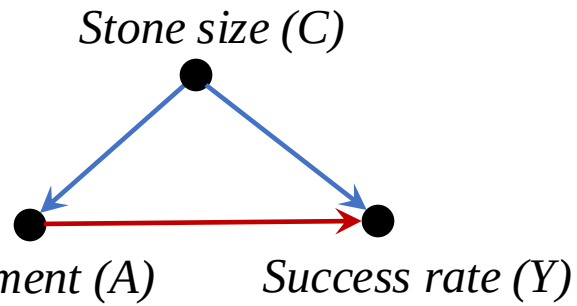
- Stone size clearly influences the success rate of treatments.
- Doctors tend to prescribe treatment A for larger stones and treatment B for smaller stones

  → Stone size influences the assignment of treatments.

- Thus, stone size serves as a **confounding variable/common cause** (*C*) between treatment (*A*) and success rate (*Y*), as illustrated in the directed acyclic graph (DAG).

Table. *Success rat for kidney stone treatment*

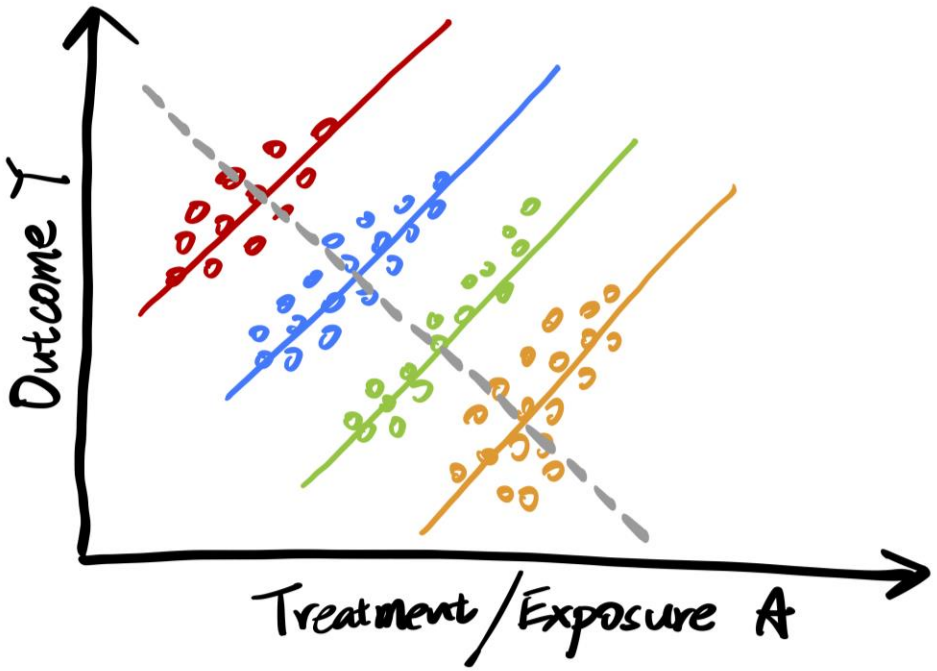
Stone Size	**Treatment**	
	Treatmen A	Treatment B
Small kidney Stones	93% (81/87)	87% (234/270)
Large kidney Stones	73% (192/263)	69% (55/80)
Both	78% (273/350)	83% (289/350)



A Directed Acyclic Graph (DAG) is a graphical representation of causal relationships, where directed edges indicate causal influence, and the structure contains no cycles. **(Part 6)**

Visual explanation for Simpson's paradox

- **Simpson's paradox: A pattern seen in separate groups may vanish or reverse when the groups are aggregated.**
- Confounding is a major challenge in distinguishing association from causation.
- Conventionally, confounding in linear regression is adjusted by including confounders as covariates to control their impact.



Berkson's paradox

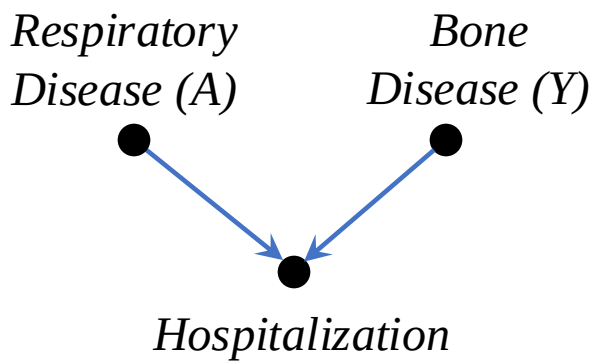
- In 1946, biostatistician Joseph Berkson identified a bias in hospital-based observational studies.
- Even if two diseases are unrelated in the general population, they may appear associated in hospital patients.

	General Population		
Respiratory Disease?	*Bone Disease?*		
	Yes	No	% Yes
Yes	17	207	7.6
No (control)	184	2,376	7.2

	Hospitalized in Last Six Months		
Respiratory Disease?	*Bone Disease?*		
	Yes	No	% Yes
Yes	5	15	25.0
No (control)	18	219	7.6

Berkson's paradox

- Berkson's paradox occurs when **selection bias**, also known as **collider-stratification bias,** creates a **spurious association** between two independent variables due to conditioning on a common effect (a **collider**).



	General Population		
Respiratory Disease?	*Bone Disease?*		
	Yes	No	% Yes
Yes	17	207	7.6
No (control)	184	2,376	7.2

	Hospitalized in Last Six Months		
Respiratory Disease?	*Bone Disease?*		
	Yes	No	% Yes
Yes	5	15	25.0
No (control)	18	219	7.6

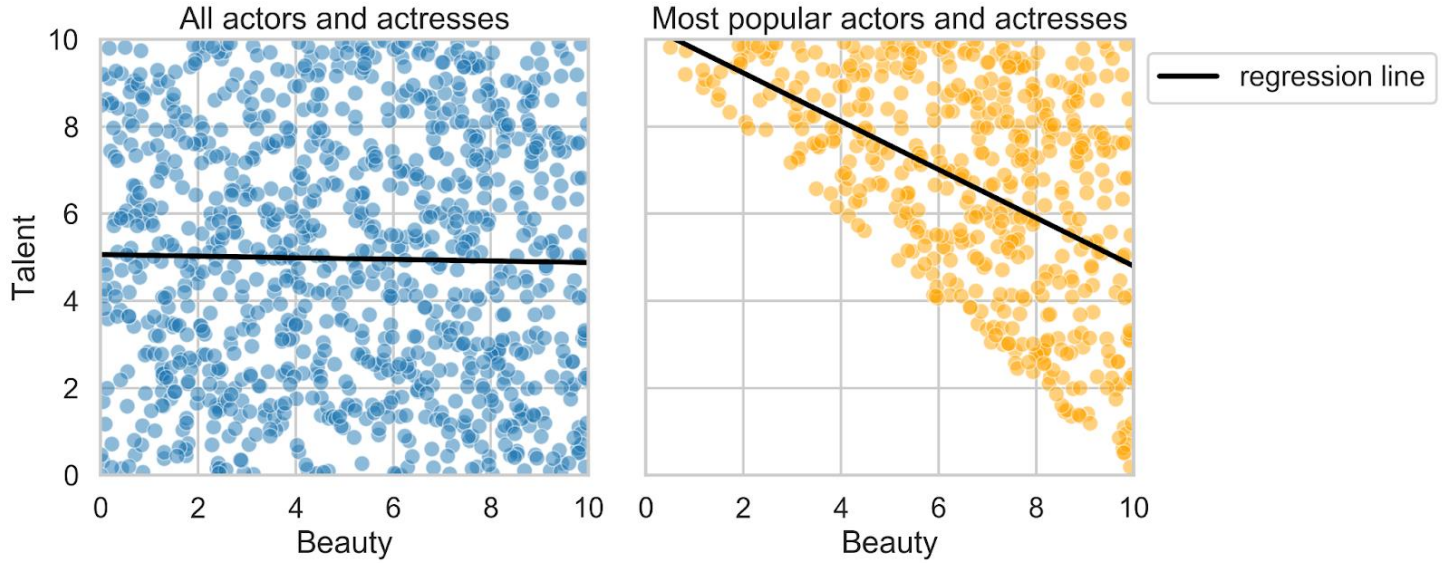
Visual explanation for Berkson's paradox (Stratification)

- **Left plot: All actors and actresses**

- Beauty and talent are independent (no correlation).
- The regression line is flat, indicating no relationship.

- **Right plot: Only the most popular actors and actresses**

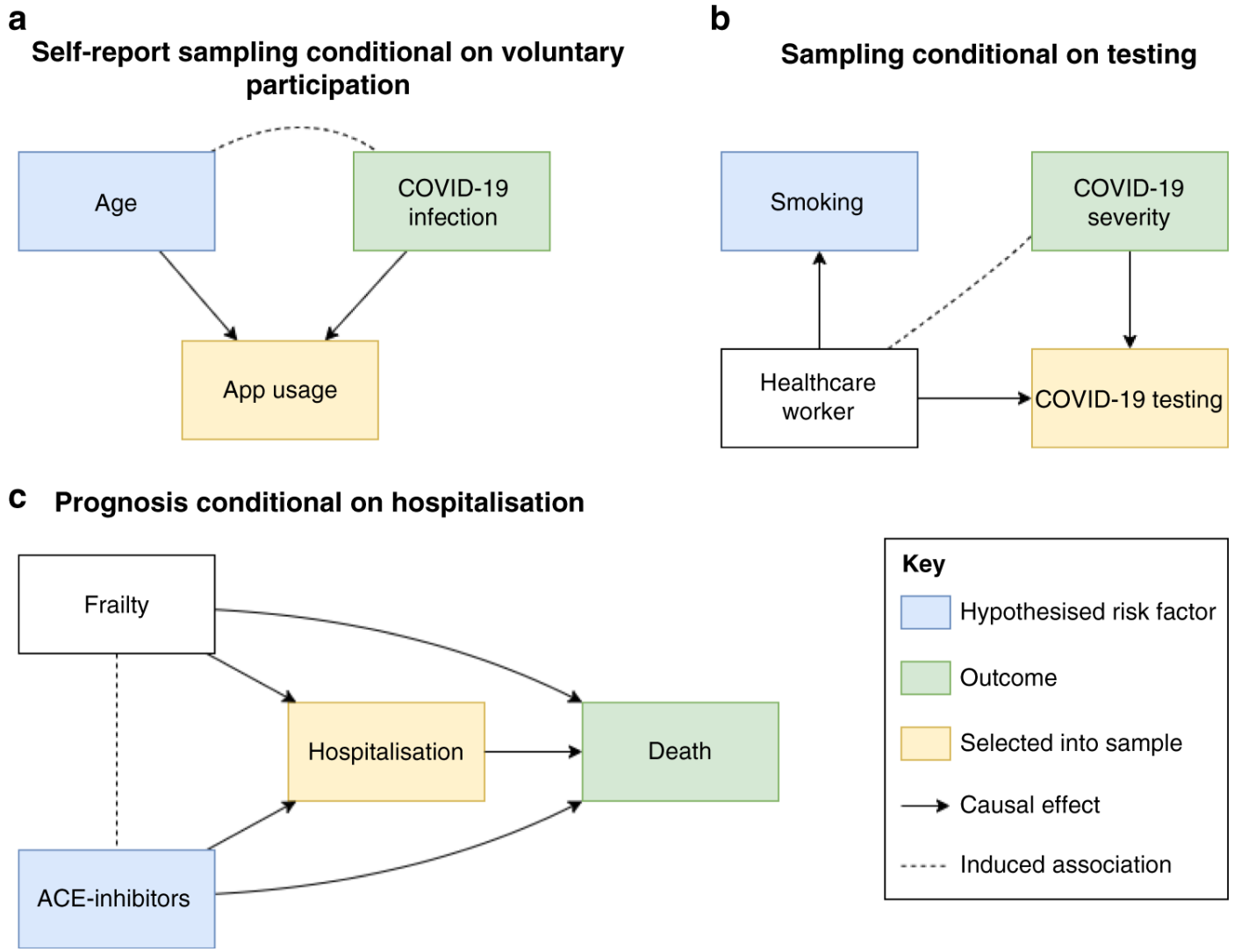
- The selection criteria introduce a spurious negative correlation between beauty and talent.
- The regression line now slopes downward, implying a trade-off that doesn't exist in the full population.



Another example of Berkson's paradox

(Griffith et *al*., 2020)

- Collider bias induced by conditioning on a collider in three scenarios relating to COVID-19 analysis



Source of Biases

(Hernán and Robins, 2020; Zhao, Keele, and Small, 2018)

Causal Bias =

Systematic bias	+	**Misspecification bias**	+	**Random Variability**
(Hidden bias) - Confounding bias **(Simpson's paradox)** - Selection bias **(Berkson's paradox)** - Measurement bias		- Due to parametric modeling		- Finite sample bias

The distinction between association and causation primarily arises from the need to account for potential systematic biases (hidden biases)

Back to the toy examples

Case 1: Y~A

```
Call:
lm(formula = Y ~ A)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3502 -0.8277 -0.0003  0.8274  5.7833

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0007395  0.0012249  -0.604    0.546
A            0.4999218  0.0008669 576.698   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
# Example of Simpson's paradox
n <- 10^6
C <- rnorm(n)
A <- C+rnorm(n)
Y <- C+rnorm(n)

summary(lm(Y~A))
```

Case 2: Y~A+C

```
Call:
lm(formula = Y ~ A + C)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1409 -0.4741 -0.0014  0.4715  3.0196

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept)  0.0008452  0.0022324    0.379    0.705
A           -0.4992008  0.0027331 -182.648   <2e-16 ***
C            0.5005031  0.0015844  315.894   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
# Example of Berkson's paradox (Stratification)
n <- 10^5
A <- rnorm(n)
Y <- rnorm(n)
C <- A+Y+rnorm(n)

summary(lm(Y~A+C))
```

- In both cases, A (treatment) is causally independent of Y (outcome).

Why do we need to learn causal inference?

- **Establishes a formal mathematical language for causal effects (Part 1)**
- **Enhances insights into standard statistical models (Parts 2, 3, and 4)**
- **Strengthens data-driven analytical skills (Parts 5 and 6)**
- **Uncovers underlying mechanisms that go beyond what association-based analyses can identify (Parts 7 and 8)**

Course Roadmap

Part 1. Introduction
1-1. Introduction to Causal Inference
1-2. Counterfactual Framework and Causal Estimands
1-3. Assumption and Identification

Part 2. Randomized Experiments
2-1. Assignment mechanisms
2-2. Classical randomized experiments

Part 3. Observational Studies with Measured Confounding
3-1. Stratification via covariates
3-2. Standardization
3-3. Propensity-score methods

Part 4. Observational Studies with Unmeasured Confounding
4-1. Front-door criterion
4-1. Difference-in-Differences (DiD) method
4-2. Instrumental variable

Part 5. Sensitivity Analysis
5-1. Evaluating sensitivity to exchangeability assumption violations
5-2. Evaluating sensitivity to positivity assumption violations

Part 6. Causal Directed Acyclic Graphs (DAGs)
6-1. DAGs for selection bias and confounding bias
6-2. DAGs for measurement bias
6-3. DAGs for interaction and effect modification

Part 7. Causal Mediation Analysis
7-1. Product method and difference method
7-2. Mediational G-formula
7-3. Estimation for causal mediation analysis

Part 8. Causal Inference for Longitudinal Data
8-1 G-method for time-vary treatments and confounders
8-2 Censoring and truncation
8-3 Causal survival analysis and survival average causal effect

References

Charig, C. R., Webb, D. R., Payne, S. R., & Wickham, J. E. (1986). Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *Br Med J (Clin Res Ed)*, *292*(6524), 879-882.

Doll, R., & Hill, A. B. (1950). Smoking and carcinoma of the lung. *British medical journal*, *2*(4682), 739.

Griffith, G. J., et al. (2020). Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nature communications*, *11*(1), 5749.

Hernán, M. A., Hsu, J., & Healy, B. (2019). A second chance to get causal inference right: a classification of data science tasks. *Chance*, *32*(1), 42-49.

Hernán M.A., Robins J.M. (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC

References

Shmueli, G. (2010). To Explain or to Predict?. *Statistical Science*, *25*(3), 289-310.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2000). *Research methods in psychology*. McGraw-Hill.

Stolley, P. D. (1991). When genius errs: RA Fisher and the lung cancer controversy. *American Journal of Epidemiology*, *133*(5), 416-425.

Zhao, Q., Keele, L. J., & Small, D. S. (2019). Comment: will competition-winning methods for causal inference also succeed in practice?. *Statistical Science*.