# Segment 1: Fundamentals of Causal Inference

## Section 04: The Fundamental Problem

# On "effects of cause" not "cause of effects" A Conversation with DB Rubin (p 446), Stat. Sci. 2014

Fan Li: In the RCM, cause/intervention should always be defined before you start the analysis. In other words, the RCM is a framework to investigate the "effects of a cause," but not the "causes of an effect." Some criticize this as a major limitation. Do you regard this as a limitation? Do you think it is ever possible to draw inference on the causes of effects from data, or is it, per se, an interesting question worth further investigation?

# On "effects of cause" not "cause of effects" A Conversation with DB Rubin (p 446), Stat. Sci. 2014

Donald Rubin: I regard "the cause" of an event topic as more of a cocktail conversation topic than a scientific inquiry, because it leads to an essentially infinite regress. Someone says, "He died of lung cancer because he smoked three packs a day"; then someone else counters, "Oh no, he died of lung cancer because both of his parents smoked three packs a day and, therefore, there was no hope of his doing anything other than smoking three packs a day"; then another one says, "No, no, his parents smoked because his grandparents smoked - they lived in North Carolina where, back then, everyone smoked three packs a day, so the cause is where the grandparents lived," and so on. How far back should you go? You can't talk sensibly about the cause of an event; you can talk about "but for that cause (and there can be many 'but for's),what would have happened?" All these questions can be addressed hypothetically. But the cause? The notion is meaningless to me.

# Foundational Concept: Causal Effect

$Z$ causally affects $Y$ for the $i^{th}$ unit if

$$Y_i^c \neq Y_i^t$$

The causal effect of $Z$ on $Y$ for the $i^{th}$ unit is defined by

$$Y_i^t - Y_i^c$$

(or some other comparison between $Y_i^t$ and $Y_i^c$)

# The Fundamental Problem of Causal Inference

A *causal effect* is defined as a contrast between outcomes for a given unit:

$$Y_i^t \quad \text{vs.} \quad Y_i^c$$

**The fundamental problem of causal inference** is that it is impossible to observe both $Y_i^t$ and $Y_i^c$. Thus, it is impossible to *observe* the causal effect of $t$ (vs. $c$) on $i$.

*The problem...is that we can never observe both...since we cannot return [to the beginning of the study] to give the other treatment. We may have the same unit measured on both treatments in two trials (a repeated measure design), but since there may exist carryover effects (e.g., the effect of the first treatment wears off slowly) or general time trends (e.g., as the [patient] ages, his [susceptibility] increases), we cannot be certain that the unit's responses would be identical at both times.* - Rubin (1974)

# Example: Single Unit

$Z = (t, c)$ for (taking, not taking) an aspirin when I have a headache

| Unit | Initial Headache | Potential Outcomes | | Causal Effect |
|---|---|---|---|---|
| | $X$ | $Y^c$ | $Y^t$ | $Y^t - Y^c$ |
| Cory | 80 | 75 | 25 | -50 |

| | | Gain Scores | | |
|---|---|---|---|---|
| | $X$ | $Y^c - X$ | $Y^t - X$ | $(Y^t - X) - (Y^c - X)$ |
| Cory | 80 | -5 | -55 | -50 |

# Example: Single Unit

$Z = (t, c)$ for (taking, not taking) an aspirin when I have a headache

| Unit | Initial Headache | | Potential Outcomes | | Causal Effect |
|---|---|---|---|---|---|
| | $X$ | $Z$ | $Y^c$ | $Y^t$ | $Y^t - Y^c$ |
| Cory | 80 | $c$ | 75 | ? | ? |

Gain Scores

| | $X$ | $Z$ | $Y^c - X$ | $Y^t - X$ | $(Y^t - X) - (Y^c - X)$ |
|---|---|---|---|---|---|
| Cory | 80 | $c$ | -5 | ? | ? |

Individual-level (or unit-level) causal effects are not identifiable based only observed data

# Example: Hypothetical Dietary Experiment

Gelman, Hill, Vehtari Section 18.1

- Question: Does adding fish oil to diet causally affect blood pressure?
- Study with $n = 8$ participants
  - 
- $n_1 = 4$ add fish oil supplement to diet for 1 year
  - 
- $n_0 = 4$ do not add fish oil supplement
  - 
- Systolic blood pressure is measured after 1 year
  - 

# Example: Hypothetical Dietary Experiment

## Observed Data

Table: Observed Data from the Hypothetical Dietary Experiment

| Unit, $i$ | Treatment, $Z_i$ | Observed Outcome, $Y_i$ |
|---|---|---|
| Audrey | 0 | 140 |
| Anna | 0 | 140 |
| Bob | 0 | 150 |
| Bill | 0 | 150 |
| Caitlin | 1 | 155 |
| Cara | 1 | 155 |
| Dave | 1 | 160 |
| Doug | 1 | 160 |

Compare average *observed* outcomes between groups:

# Example: Hypothetical Dietary Experiment

## Pre-Post

Take Unit $i = 5$, Caitlin:

- We observe that her blood pressure after taking the supplement for a year is 155.
- What if we knew that her blood pressure before the study 160?
  - I.e., her "pre-treatment" blood pressure
- Would we conclude that the supplement caused a reduction?
- What would we need to know to make this conclusion?

# Example: Hypothetical Dietary Experiment

## Potential Outcomes

Table: Potential Outcomes from the Hypothetical Dietary Experiment

| Unit, $i$ | Treatment $Z_i$ | Potential Outcome, $Y_i^c$ | Potential Outcome, $Y_i^t$ | Observed Outcome, $Y_i$ |
|---|---|---|---|---|
| Audrey | 0 | 140 | 135 | 140 |
| Anna | 0 | 140 | 135 | 140 |
| Bob | 0 | 150 | 140 | 150 |
| Bill | 0 | 150 | 140 | 150 |
| Caitlin | 1 | 160 | 155 | 155 |
| Cara | 1 | 160 | 155 | 155 |
| Dave | 1 | 170 | 160 | 160 |
| Doug | 1 | 170 | 160 | 160 |

# Example: Hypothetical Dietary Experiment

## Observed Data

Table: Observed Data from the Hypothetical Dietary Experiment

| Unit, $i$ | Treatment $Z_i$ | Potential Outcome, $Y_i^c$ | Potential Outcome, $Y_i^t$ | Observed Outcome, $Y_i$ |
|---|---|---|---|---|
| Audrey | 0 | 140 | ? | 140 |
| Anna | 0 | 140 | ? | 140 |
| Bob | 0 | 150 | ? | 150 |
| Bill | 0 | 150 | ? | 150 |
| Caitlin | 1 | ? | 155 | 155 |
| Cara | 1 | ? | 155 | 155 |
| Dave | 1 | ? | 160 | 160 |
| Doug | 1 | ? | 160 | 160 |