Correlation Does Not Imply Causation

Gordon Dahl and the EVH Team
UC San Diego

Correlation versus Causation

In English:

Correlation Does Not Imply Causation

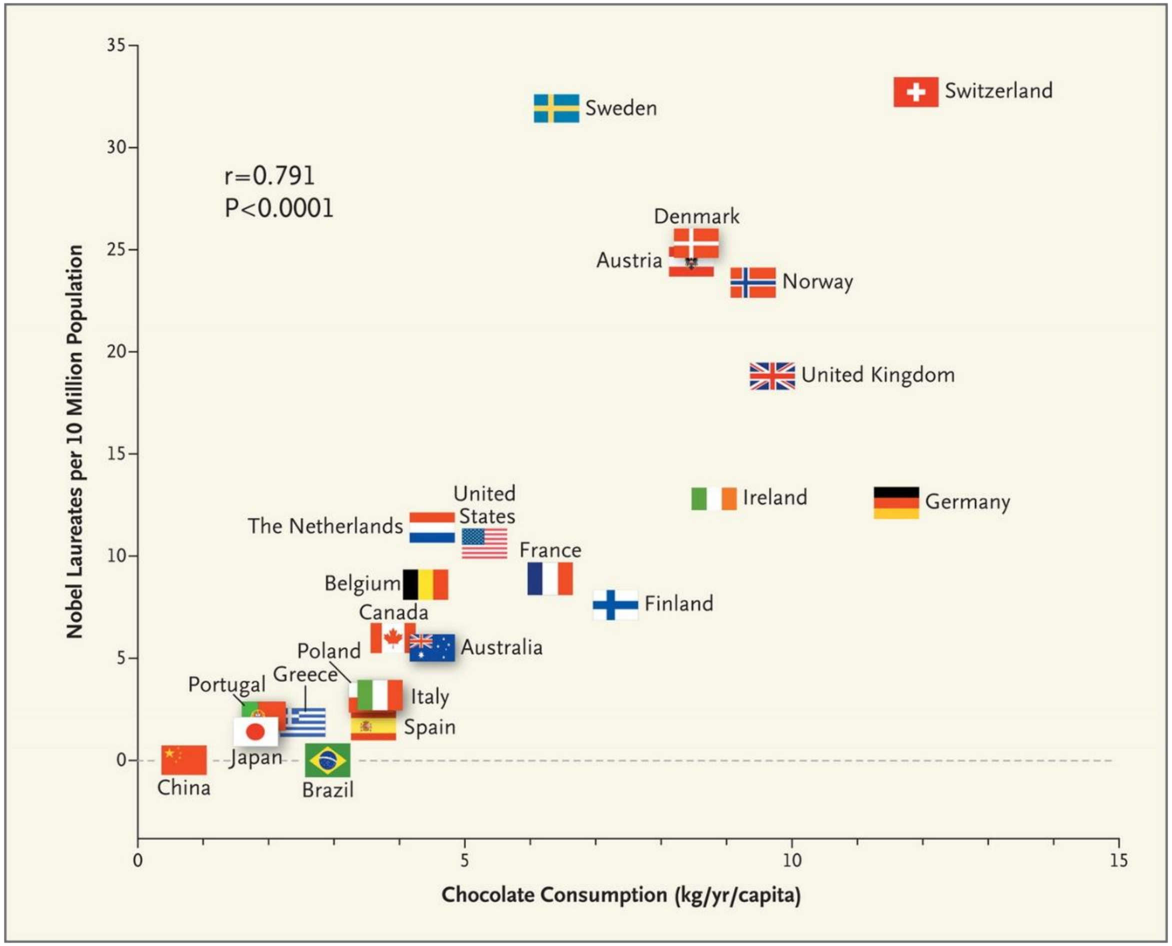
In Latin:

Cum Hoc Ergo Propter Hoc

Similar Idea:

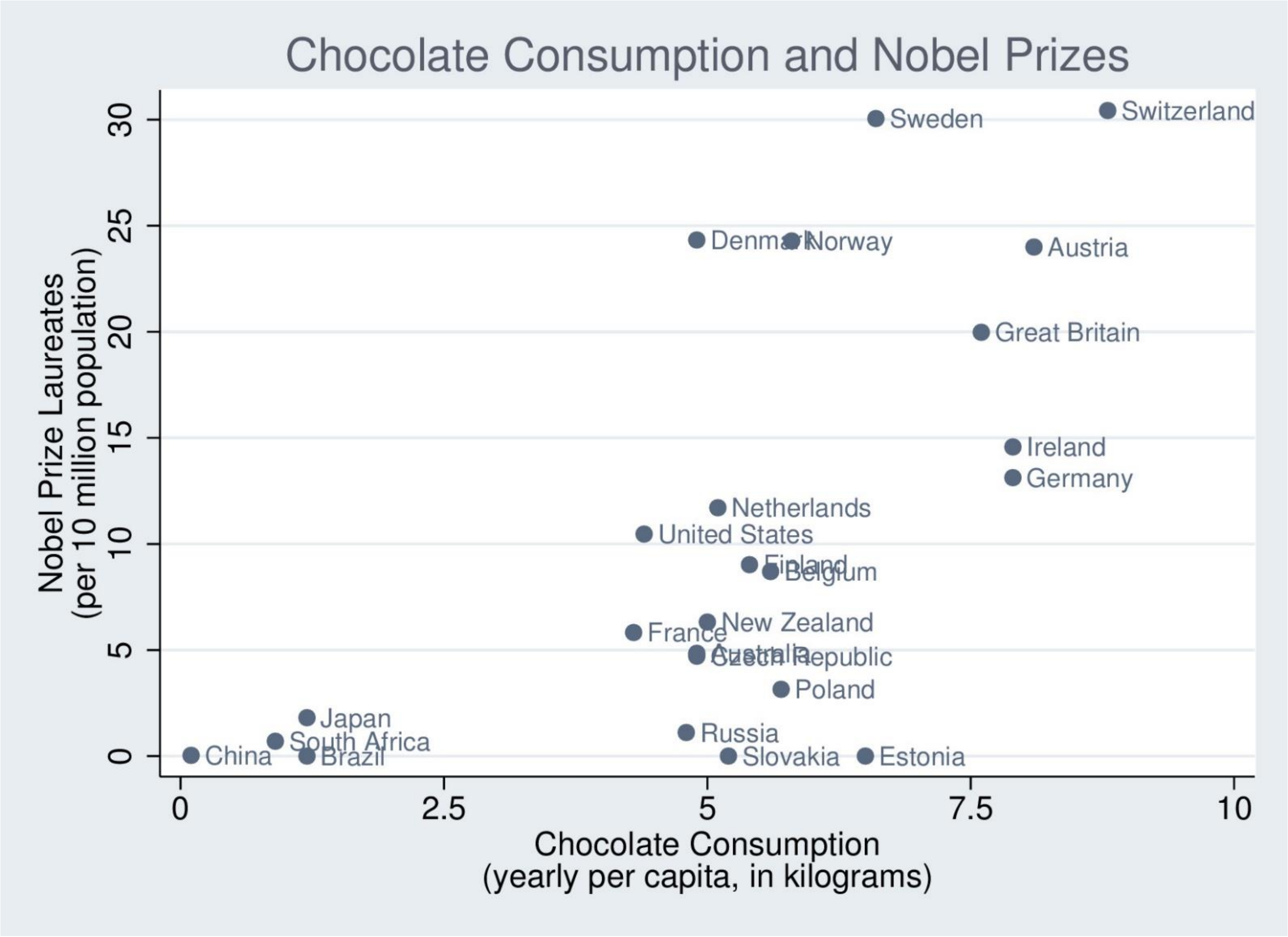
Post Hoc Ergo Propter Hoc

Example 1: Does Chocolate Cause Nobel Prizes?



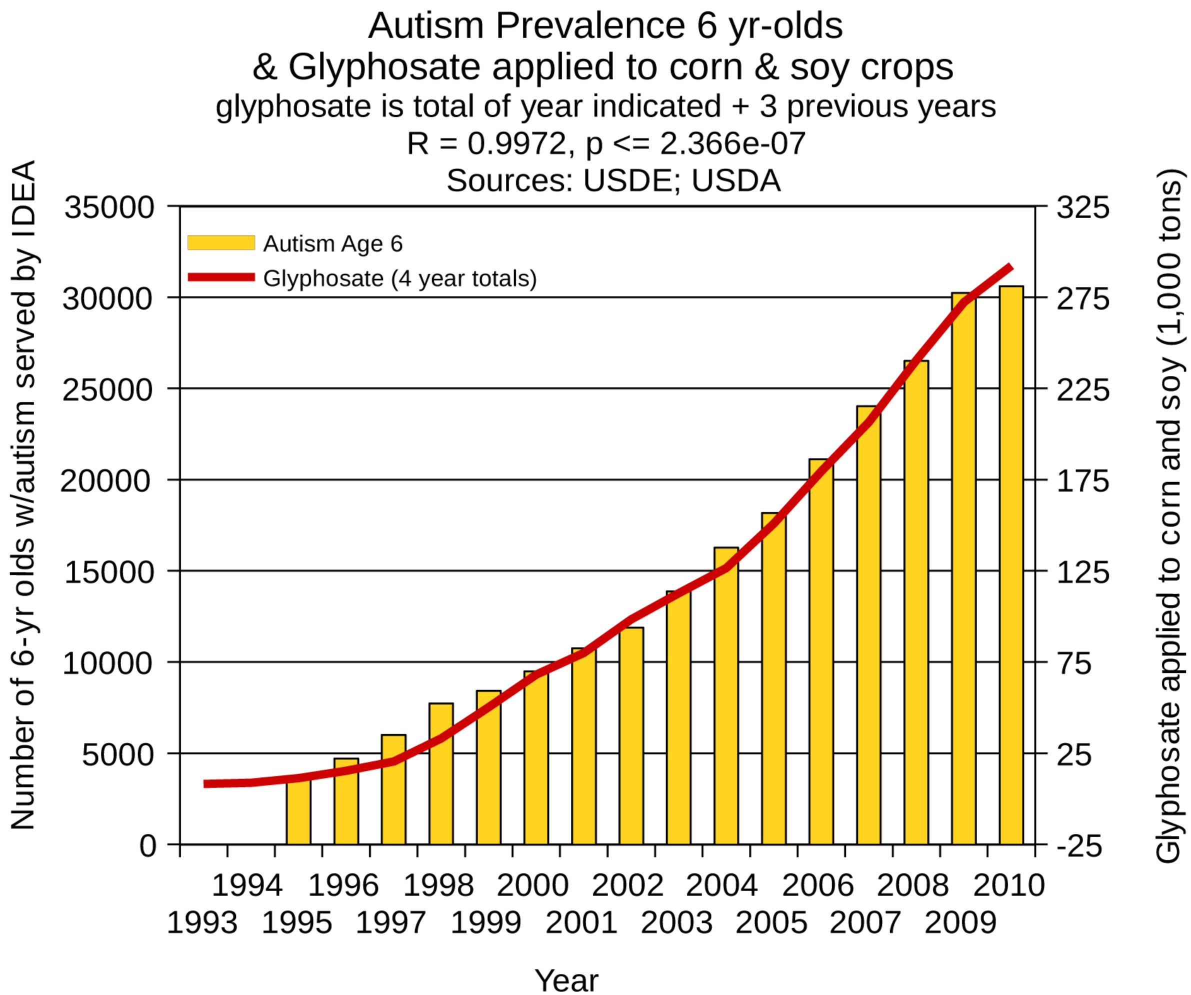
Source: Messerli, Franz H. "Chocolate Consumption, Cognitive Function, and Nobel Laureates," The New England Journal of Medicine, 2012, Vol. 367, No. 16, p. 1563.

Example 1: Does Chocolate Cause Nobel Prizes?



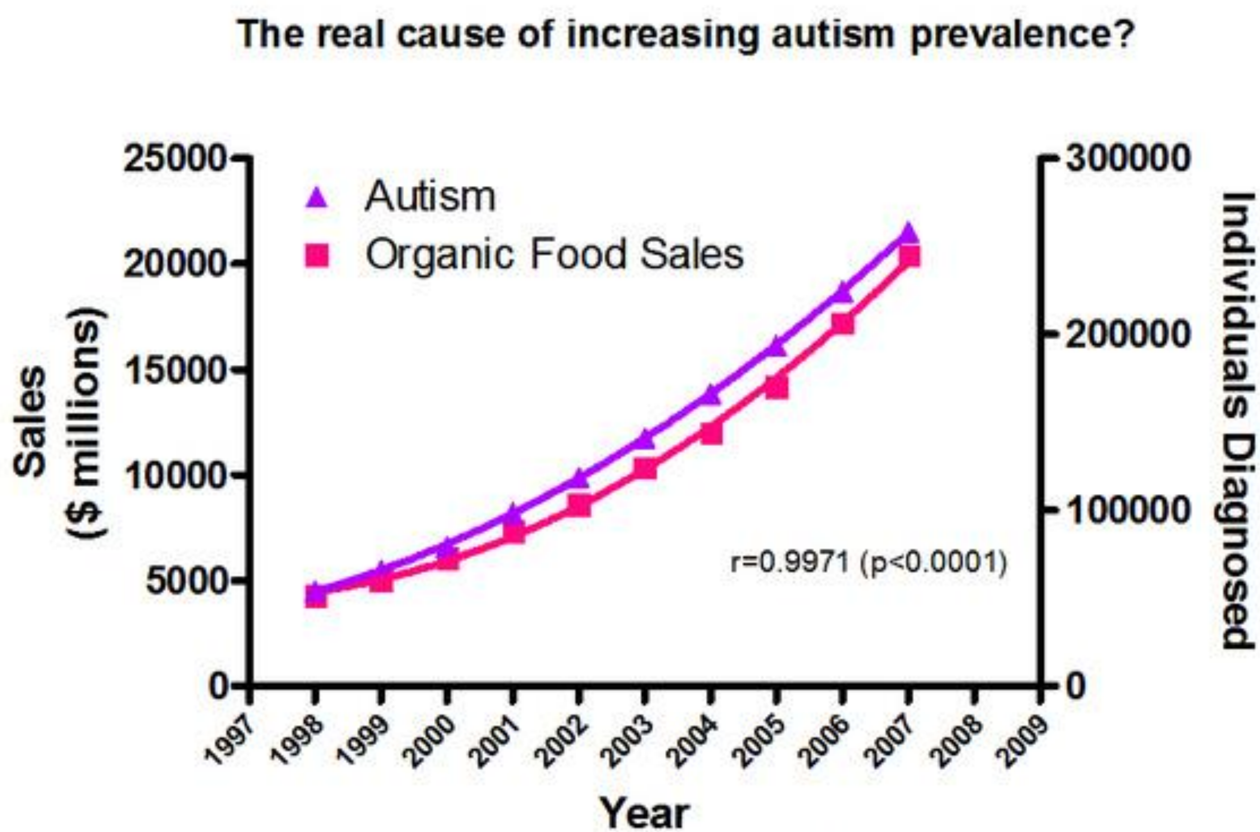
Source: Gordon Dahl's updated analysis based on the idea presented in Messerli, Franz H. "Chocolate Consumption, Cognitive Function, and Nobel Laureates," The New England Journal of Medicine, 2012, Vol. 367, No. 16, p. 1563.

Example 2: Does Glyphosate Cause Autism?



Source: Seneff et al., "Aluminum and Glyphosate Can Synergistically Induce Pineal Gland Pathology: Connection to Gut Dysbiosis and Neurological Disease," Agricultural Sciences, 2015, Vol. 6, p. 52.

Example 3: Does Organic Food Cause Autism?



Sources: Organic Trade Association, 2011 Organic Industry Survey; U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB# 1820-0043: "Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act

Correlation vs. Causation

- Just because two events occur at the same time does not mean one caused the other
- Even when one event occurs before the other, this does not imply causation

Definition of Causality?

- Causality has been the subject of philosophers, going back at least as far as Aristotle
- Different academic fields sometimes have different ideas about causality
  - For example, physics, engineering, biology, epidemiology, psychology, law, history, and sociology often think about and define causality in different ways
  - Some fields distinguish between various types of causality
- Economists and statisticians usually define causality in terms of **potential outcomes**

Epidemiology Definition

- For example, epidemiology has traditionally used a list known as *Hill's criteria for causation* (1965)
  - Strength: The larger the association, the more likely it is to be causal
  - Consistency: Similar findings with different samples
  - Specificity: More specific associations more likely to be causal
  - Temporality: Cause first, then effect
  - Biological gradient: Greater dose should lead to bigger effect
  - Plausibility: Plausible mechanism for cause and effect
  - Coherence: Matches up with lab findings
  - Experiment: Useful if available
  - Analogy: Similar factors find similar effects
- Think about these criteria applied to the example of whether glyphosate causes autism

Granger's Definition

- *Granger causality* (1969)
  - Statistical concept based on prediction
  - Time series data: Y_t, X_t for t=1,…, T
  - Forecast Y_{t+1} using past terms of Y_t
  - Next forecast Y_{t+1} using past terms of Y_t and X_t
  - If second forecast is more successful, then X_t "Granger causes" Y_{t+1}
- Limitations
  - What if X_t and Y_{t+1} both correlated with an omitted variable?
  - What about instantaneous effects, expectations about the future or non-linear relationships?
- Two examples
  - Inflation and employment
  - Rain and umbrellas

Statistician's Definition Using Potential Outcomes

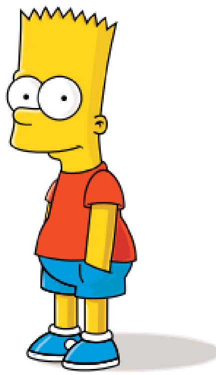
- Potential outcomes model
  - *Rubin Causal Model* (1974)
  - Origins in work by *Neyman and Fisher* in the early 1920's related to randomized experiments
- Thought experiment with two parallel worlds
  - One with treatment imposed and the other without
  - Compare the outcomes in the two worlds for each individual
  - Captures the ceteris paribus idea: everything else is the same in the two worlds, except for treatment status of the individual

Fundamental Challenge

versus

- Suppose we want to know the causal effect on grades of reading the class textbook in advance of the test
- Fundamental challenge: can’t observe both states of the world for the same individual
  - Only get to see an individual’s “potential outcome” for the world they actually live on

Fundamental Challenge



versus?

- Suppose we want to know the causal effect on grades of reading the class textbook in advance of the test
- Fundamental challenge: can’t observe both states of the world for the same individual
  - Only get to see an individual’s “potential outcome” for the world they actually live on

Fundamental Challenge

- Fundamental challenge: can’t observe both states of the world for the same individual
  - We only get to see data for the world an individual actually lives on
  - We’d like to know an individual’s “potential outcome” for the world they don’t live on, but that data is missing

Data Sources: Chocolate Consumption and Nobel Prizes

Conway, Jan: (2019, May). Per capita chocolate consumption worldwide in 2017, by country (in kilograms). Retrieved from https://www.statista.com/statistics/819288/worldwide-chocolate-consumption-by-country/

List of countries by Nobel laureates per capita. Retrieved 2019 July 10, from https://en.wikipedia.org/wiki/List_of_countries_by_Nobel_laureates_per_capita

Potential Outcomes and Causality

Gordon Dahl and the EVH Team

UC San Diego

Causality and Treatment Effects

- Often we are interested in evaluating the **causal** effect of treatment on some outcome
- Example: Does aspirin causally reduce head pain?
- In this example
  - taking aspirin is the treatment
  - whether a person has a headache is the outcome

Treatment Effects: Notation

- Denote **treatment** with a binary variable: $D_i = \{0, 1\}$
  - the variable equals 1 if individual i is treated and 0 if not treated
- Outcome variable of interest: Y_i
- Question: Does D_i causally influence Y_i?

Potential Outcomes: Parallel Worlds

- Causality can be defined in the “potential outcomes framework”
- Thought experiment with two parallel worlds
  - In one world treatment is imposed, but not in the other
  - In this thought experiment, the same individual is observed on both parallel worlds
  - Captures the ceteris paribus idea: everything else is the same in the two worlds, except for the treatment status of the individual

Potential Outcomes: Notation

- Potential outcomes

$$\text{Potential outcome} = \begin{cases} Y_i^1 & \text{in the treated world} \\ Y_i^0 & \text{in the untreated world} \end{cases}$$

- The superscripts on Y denote what the outcome would be for individual i if they were on the treated world and the untreated world

Causal Treatment Effects

- If we could observe every individual i on both worlds, then we could compare what happens with and without treatment for each individual

$$\text{Treatment effect for individual } i = Y_i^1 - Y_i^0$$

- $Y_i^1 - Y_i^0$ can be different for different people. We are often interested in the average treatment effect (ATE).

$$ATE = E\left[Y_i^1 - Y_i^0\right]$$

Causal Treatment Effects

- Sometimes we want to know the average treatment effect on the treated (ATT), rather than for everyone

$$ATT = E\left[Y_i^1 - Y_i^0 | D_i = 1\right]$$

- Both the ATE and the ATT, defined in terms of potential outcomes, capture causal effects

ATE and ATT are Population Parameters

- Recall the ***expectation*** operator

$$\mathsf{E}\left[Y\right] = \int_{-\infty}^{\infty} y f_Y\left(y\right) dy$$

- And the ***conditional expectation*** operator

$$\begin{aligned}\mathsf{E}\left[Y|X = x\right] &= \int_{-\infty}^{\infty} y f_{Y|X}\left(y|X = x\right) dy \\ &= \int_{-\infty}^{\infty} y \frac{f_{X,Y}\left(x, y\right)}{f_X\left(x\right)} dy\end{aligned}$$

- ATE and ATT are just an expectation and a conditional expectation, respectively
- ATE and ATT are population parameters
- Can we estimate these population parameters using a sample, particularly if we only observe Y_i^1 *or* Y_i^0, but not both, for a given individual?

Bias from Self Selection

Gordon Dahl and the EVH Team

UC San Diego

Reminder: Causal Treatment Effect Parameters

- We are often interested in the average treatment effect for all individuals (**ATE**)

$$ATE = E\left[Y_i^1 - Y_i^0\right]$$

- Sometimes we want to know the average treatment effect on the treated (**ATT**), rather than for everyone

$$ATT = E\left[Y_i^1 - Y_i^0 | D_i = 1\right]$$

- These treatment effects, defined in terms of potential outcomes (think of two parallel worlds), capture causal effects

The Missing Data Problem

- In reality, people do not live in two parallel worlds
- This means we never observe both Y_i^1 *and* Y_i^0
- Instead we observe

$$
\begin{aligned}
Y_i &= \begin{cases} Y_i^1 & \text{if } D_i = 1 \\ Y_i^0 & \text{if } D_i = 0 \end{cases} \\
&= Y_i^0 + \left(Y_i^1 - Y_i^0 \right) D_i
\end{aligned}
$$

- Hence, we have a "missing data" problem
  - We observe one of the outcomes for each individual, but not the other
- Stated differently, we don't observe the **counterfactual** for an individual

Aspirin Example

- Example: Head pain is the outcome and aspirin is the treatment

- Consider the ATE

$$ATE = E\left[Y_i^1 - Y_i^0\right]$$

- Causal estimate depends on
  - Y_i^1: whether individual i would have a headache if they take aspirin, *irrespective* of whether they actually take aspirin
  - Y_i^0: whether individual i would have a headache if they don't take aspirin, *irrespective* of whether they actually take aspirin

- We don't observe both Y_i^1 and Y_i^0, but only Y_i
  - I.e., we only observe Y_i^1 if $D_i = 1$ and Y_i^0 if $D_i = 0$
  - This is a “missing data” problem

Selection Bias in ATT with Observational Data

- Does taking the observed difference in outcomes between treated and untreated individuals estimate either ATE or ATT?
  - **No!**

- First consider ATT

$$ATT = E\left[Y_i^1 - Y_i^0 | D_i = 1\right]$$

- Decompose the observed difference

$$\underbrace{E\left[Y_i | D_i = 1\right] - E\left[Y_i | D_i = 0\right]}_{\text{Observed difference}} = E\left[Y_i^1 | D_i = 1\right] - E\left[Y_i^0 | D_i = 0\right]$$

$$= \underbrace{E\left[Y_i^1 | D_i = 1\right] - E\left[Y_i^0 | D_i = 1\right]}_{\text{ATT}}$$

$$\underbrace{+ E\left[Y_i^0 | D_i = 1\right] - E\left[Y_i^0 | D_i = 0\right]}_{\text{Selection bias}}$$

Selection Bias in ATT with Observational Data

$$Selection\, bias = E\left[Y_i^0|D_i = 1\right] - E\left[Y_i^0|D_i = 0\right]$$

- **Selection bias** arises if individuals choose treatment based "expected benefits" or "expected gains"

- Aspirin example
  - Woud individuals who chose to take aspirin ($D_i = 1$) have experienced headaches had they not taken aspirin (Y_i^0) at the same rate as individuals who did not choose to take aspirin ($D_i = 0$)?
  - It is likely that individuals who took aspirin did so because they expected a gain from doing so – maybe they already had a headache, and thought aspirin would help
    - This violates the ceteribus paribus assumption needed for a caual estimate
  - What is the likely sign of the bias?

Detour: Law of Iterated Expectations

- The law of iterated expectations says that for two random variables X and Y

$$E[Y] = E[E[Y|X]]$$

- If X is a discrete random variable which takes on just two values, 0 and 1, this can be written as

$$E[Y] = Pr(X = 1)E[Y|X = 1] + Pr(X = 0)E[Y|X = 0]$$

Selection Bias in ATE with Observational Data

- Decompose the observed difference another way

$$\underbrace{E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]}_{\text{Observed difference}}=E\left[Y_i^1|D_i=1\right]-E\left[Y_i^0|D_i=0\right]$$

$$\begin{aligned}=&Pr(D_i=1)E\left[Y_i^1|D_i=1\right]+Pr(D_i=0)E\left[Y_i^1|D_i=1\right]\\ &-Pr(D_i=1)E\left[Y_i^0|D_i=0\right]-Pr(D_i=0)E\left[Y_i^0|D_i=0\right]\end{aligned}$$

$$\begin{aligned}=&\underbrace{Pr(D_i=1)E\left[Y_i^1|D_i=1\right]+Pr(D_i=0)E\left[Y_i^1|D_i=0\right]}_{E[Y_i^1]}\\ &\underbrace{-Pr(D_i=0)E\left[Y_i^0|D_i=0\right]-Pr(D_i=1)E\left[Y_i^0|D_i=1\right]}_{-E[Y_i^0]}\\ &+Pr(D_i=1)(E\left[Y_i^0|D_i=1\right]-E\left[Y_i^0|D_i=0\right])-Pr(D_i=0)(E\left[Y_i^1|D_i=0\right]-E\left[Y_i^1|D_i=1\right])\end{aligned}$$

$$\begin{aligned}=&\underbrace{E\left[Y_i^1\right]-E\left[Y_i^0\right]}_{\text{ATE}}\\ &+\underbrace{Pr(D_i=1)(E\left[Y_i^0|D_i=1\right]-E\left[Y_i^0|D_i=0\right])+Pr(D_i=0)(E\left[Y_i^1|D_i=1\right]-E\left[Y_i^1|D_i=0\right])}_{\text{Selection bias}}\end{aligned}$$

Selection Bias in ATE with Observational Data

$$Selection\,bias = Pr(D_i=1)(E\left[Y_i^0|D_i=1\right]-E\left[Y_i^0|D_i=0\right])+Pr(D_i=0)(E\left[Y_i^1|D_i=1\right]-E\left[Y_i^1|D_i=0\right])$$

- As before, **selection bias** arises if individuals choose treatment based "expected gains"
- *Note: The selection bias terms for ATE and ATT are not the same, because ATE and ATT are different parameters*
- Aspirin example
  - Woud individuals who chose to take aspirin ($D_i = 1$) have experienced headaches had they not taken aspirin (Y_i^0) at the same rate as individuals experiencing headaches had they taken aspirin (Y_i^1) who did not choose to take aspirin ($D_i = 0$)?
    - Whew, that's a mouthful! But the answer is clearly no!
  - What is the likely sign of the bias?

Empirical Example

- Example: Do hospitals make people healthier?

- "Would you say your health in general is excellent (5), very good (4), good (3), fair (2), poor (1)?"

Group	Sample Size	Mean Health Status	Std. Error
Hospital	7,774	3.21	0.014
No Hospital	90,049	3.93	0.003

- *Source: National Health Interview Survey (NHIS) 2005, via Mostly Harmless Econometrics (Angrist and Pischke, 2009)*

- Difference in means is -0.72 (t-statistic = 50)
- So do hospitals make people have *worse* health?

Empirical Example

- Based on the table, we observe:

$$E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]=3.21-3.93=-0.72$$

  - This likely does not equal either ATT or ATE

- Consider ATT and decompose the observed difference

$$\underbrace{E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]}_{\text{Observed difference in average health}}=\underbrace{E\left[Y_i^1|D_i=1\right]-E\left[Y_i^0|D_i=1\right]}_{\text{ATT}}$$
$$+\underbrace{E\left[Y_i^0|D_i=1\right]-E\left[Y_i^0|D_i=0\right]}_{\text{Selection bias}}$$

  - People who go to the hospital ($D_i=1$) likely would have been in even worse health had they not gone to the hospital (Y_i^0), while people who did not go to the hospital ($D_i=0$) likely have had good health even without going to the hospital (Y_i^0)

Randomized Controlled Trials

Gordon Dahl and the EVH Team

UC San Diego

Gold Standard in Scientific Research

- Appeal of a randomized experiment to estimate causal outcomes
- Split sample into two groups randomly
- All characteristics should be the same on average in the two groups
- While many factors may still influence outcomes, these factors are “balanced” across treatment and control

Randomized Controlled Trials

- Randomized experiments are often called “randomized control trials” (**RCT**)
- *Random*: Subjects are randomly assigned to treatment and control groups
- *Controlled*: There is a control group which does not receive treatment
- *Trial*: Treatment is assigned as part of the experiment, with subjects having no choice

Randomization Implies Independence

- Randomly assign some units to **treatment** $(D_i = 1)$ and others to **control** $(D_i = 0)$
- Randomization implies that treatment, D_i, is *independent* of the potential outcomes, Y_i^1 and Y_i^0
- This means that

$$E\left[Y_i^1 | D_i = 1\right] = E\left[Y_i^1 | D_i = 0\right] = E\left[Y_i^1\right]$$

$$E\left[Y_i^0 | D_i = 1\right] = E\left[Y_i^0 | D_i = 0\right] = E\left[Y_i^0\right]$$

Randomization Eliminates Selection Bias in ATT

- Sample is individuals who would like to participate in treatment if given the choice

- With randomization to treatment and control of these individuals, the observed difference estimates ATT

$$\underbrace{E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]}_{\text{Observed difference}} = E\left[Y_i^1|D_i=1\right]-E\left[Y_i^0|D_i=0\right]$$

$$= \underbrace{E\left[Y_i^1|D_i=1\right]-E\left[Y_i^0|D_i=1\right]}_{\text{ATT}}$$

$$+ \underbrace{E\left[Y_i^0|D_i=1\right]-E\left[Y_i^0|D_i=0\right]}_{\text{Selection bias (=0)}}$$

  - since randomization implies $E\left[Y_i^0|D_i=1\right]=E\left[Y_i^0|D_i=0\right]$

- There is no selection bias, because individuals are not choosing whether or not to be treated

Randomization Eliminates Selection Bias in ATE

- Sample is the entire population

- With randomization to treatment and control, the observed difference estimates ATE

$$\underbrace{E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]}_{\text{Observed difference}} = E\left[Y_i^1|D_i=1\right]-E\left[Y_i^0|D_i=0\right]$$

$$= \underbrace{E\left[Y_i^1\right]-E\left[Y_i^0\right]}_{ATE}$$

  - since randomization implies $E\left[Y_i^1|D_i=1\right]=E\left[Y_i^1\right]$ and $E\left[Y_i^0|D_i=0\right]=E\left[Y_i^0\right]$

- Again, there is no selection bias, because individuals are not choosing whether or not to be treated

Example: Fertilizer and Crop Yields

- Effect of fertilizer on crop yields
  - 200 fields used to grow corn
  - Randomly choose 100 fields to receive fertilizer (treatment group, $D_i = 1$), and 100 fields not to be fertilized (control group, $D_i = 0$)

- Observe whether crop yields are higher in the treated versus the control fields

$$\underbrace{E\left[Y_i | D_i = 1\right] - E\left[Y_i | D_i = 0\right]}_{\text{Observed difference}} = \underbrace{E\left[Y_i^1\right] - E\left[Y_i^0\right]}_{\text{ATE}}$$

  - Because of randomization, there is no selection bias term to worry about

Example, continued

- Manage the 200 fields identically
  - Use the same amount of corn seed, plant the seeds at the same time, etc.
  - Even so, there will be individual differences across fields
    - Some fields will have better soil, receive more rain, etc.

- But all of these individual differences should be the same “on average” between treatment and control fields
  - These other factors are “balanced” on average due to randomization
  - For example, the amount of rain the 100 treated fields get on average should be similar to the amount of rain the 100 control fields get on average, even though individual fields will get more or less rain

Humans Make Running Experiments Harder

- People make choices
  - Farm fields do what they are told, people (including farmers!) often do not
  - Fewer ethical concerns with farm fields than people
- Example #1: Does smoking lower birthweight?
- Example #2: Do prison education programs reduce recidivism?
- In these examples, would people do what they are told?
- In these examples, would it be ethical, affordable, and politically feasible to run an RCT?

Natural Experiments and Social Experiments

- Experiments have limitations and challenges when they involve humans
  - We will discuss this in the context of “Social Experiments”
- Often, we can’t conduct an actual experiment, but can take advantage of naturally occuring variation which mimics an experiment
  - We will discuss how social scientists use these “Natural Experiments”