



Lecture 20

Causality

Announcements

- Homework 7 is due Thursday, 03/10
 - Midterm this Friday, March 11, 7-9pm PT
 - [Prep Guide](#), [Past Exams](#)
 - Logistics post will be up today at 5pm PT
 - Room assignments will be out by Wednesday
 - Catch up sessions this week!
-

Weekly Goals

- **Today**
 - Causation
 - Randomized Control Experiments
 - Wednesday
 - P-Value as an Error
 - Examples
 - Friday
 - Midterm review
-

How We've Tested Thus Far

Hypothesis Testing Review

- **1 Sample: One Category** (e.g. percent of flowers that are purple)
 - Test Statistic: `observed_proportion, abs(observed_proportion - null_proportion)`
 - How to Simulate: `sample_proportions(n, null_dist)`
- **1 Sample: More Than 2 Categories** (e.g. ethnicity distribution of jury panel)
 - Test Statistic: `tvd(observed_dist, null_dist)`
 - How to Simulate: `sample_proportions(n, null_dist)`
- **1 Sample: Numerical Data** (e.g. scores in a lab section)
 - Test Statistic: `observed_mean, abs(observed_mean - null_mean)`
 - How to Simulate: `population_data.sample(n, with_replacement=False)`
- **2 Samples: Underlying Values** (e.g. birth weights of smokers vs. non-smokers)
 - Test Statistic: `group_a_mean - group_b_mean, group_b_mean - group_a_mean, abs(group_a_mean - group_b_mean)`
 - How to Simulate: `observed_data.sample(with_replacement=False)`

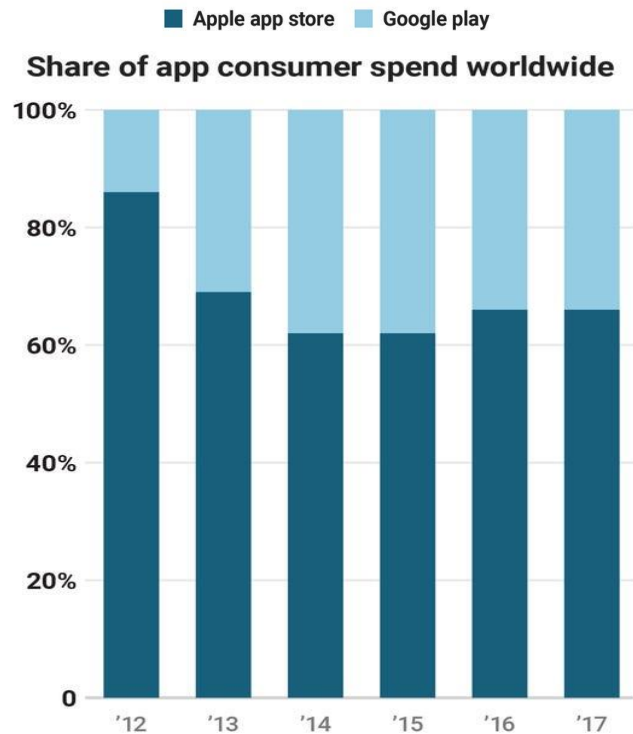
Review: *A/B* Testing

(Demo)

Random Assignment

Importance of Random Assignment

Apple users more willing to pay for apps



Importance of Random Assignment

- iOS users spend 2x as much as Android users on 3rd party apps
 - *Is **higher spending** caused by users owning **iPhone**?*
 - Can't Tell:
 - Users aren't randomly assigned a phone
 - Other factors contribute to their phone purchasing decisions (e.g. income, geography)
-

Causality

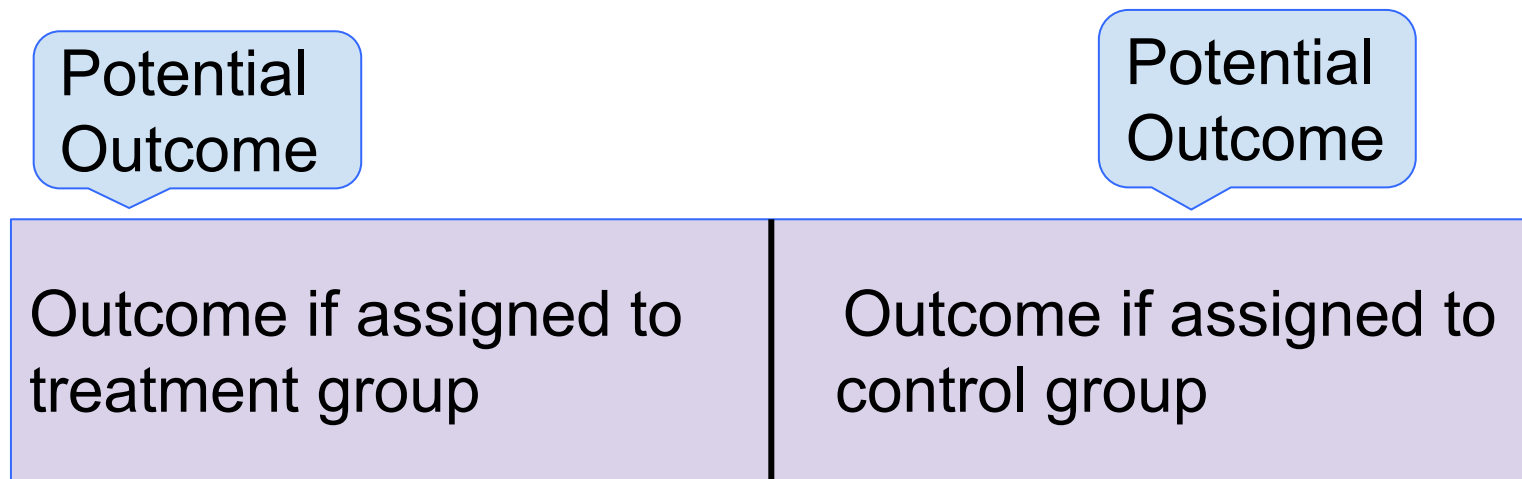
Randomized Controlled Experiment

- Sample A: **control group**
- Sample B: **treatment group**
- **If the treatment and control groups are selected at random, then you can make causal conclusions.**
- Any difference in outcomes between the two groups could be due to
 - chance
 - the treatment

(Demo)

Before the Randomization

- In the population there is one imaginary ticket for each of the 31 participants in the experiment.
- Each participant's ticket looks like this:



The Data

16 randomly picked tickets show:

	Outcome if assigned to control group
--	--------------------------------------

The remaining 15 tickets show:

Outcome if assigned to treatment group	
--	--

The Hypotheses

- **Null:**

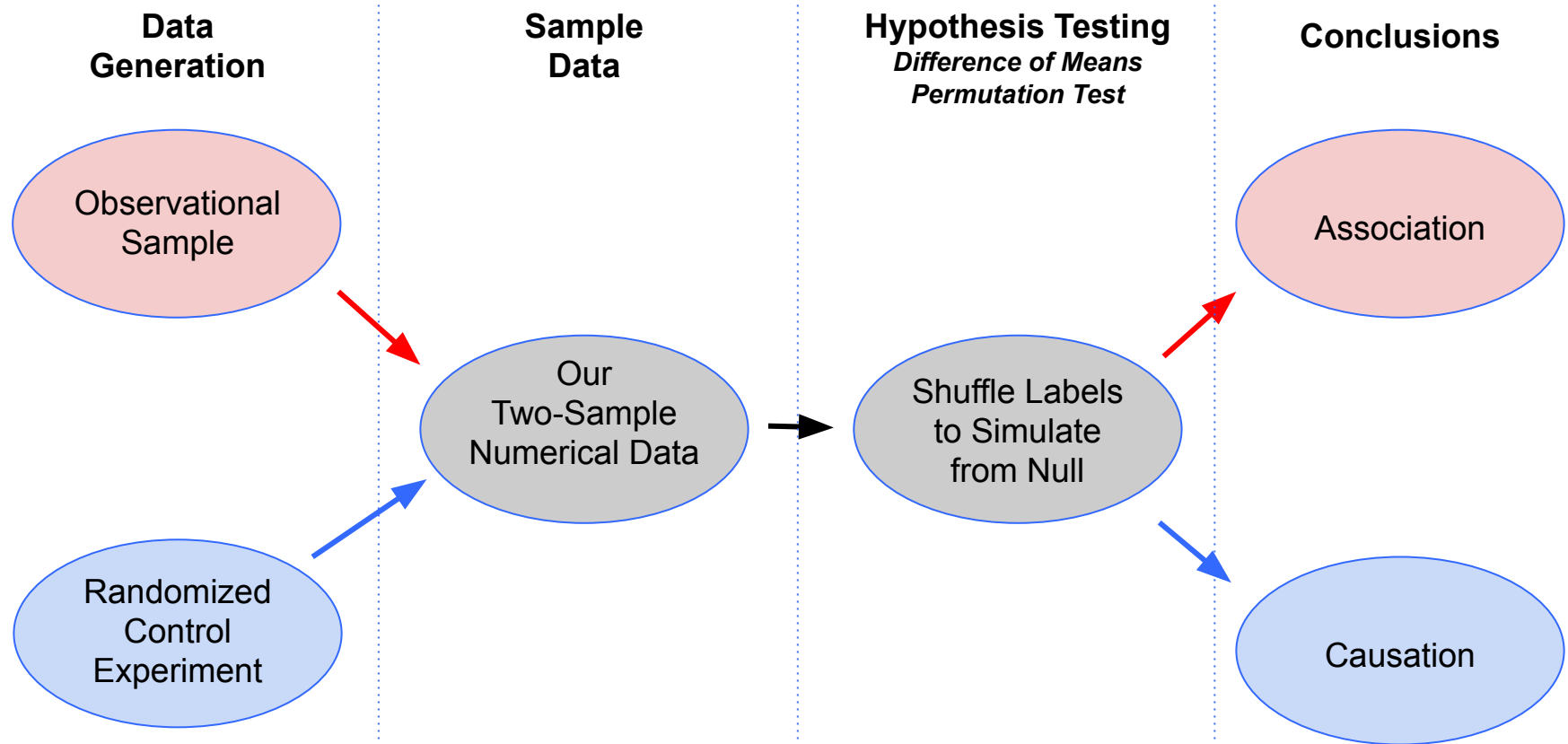
- In the population, the distribution of all potential control scores is the same as the distribution of all potential treatment scores.
- tl;dr the treatment has no effect

- **Alternative:**

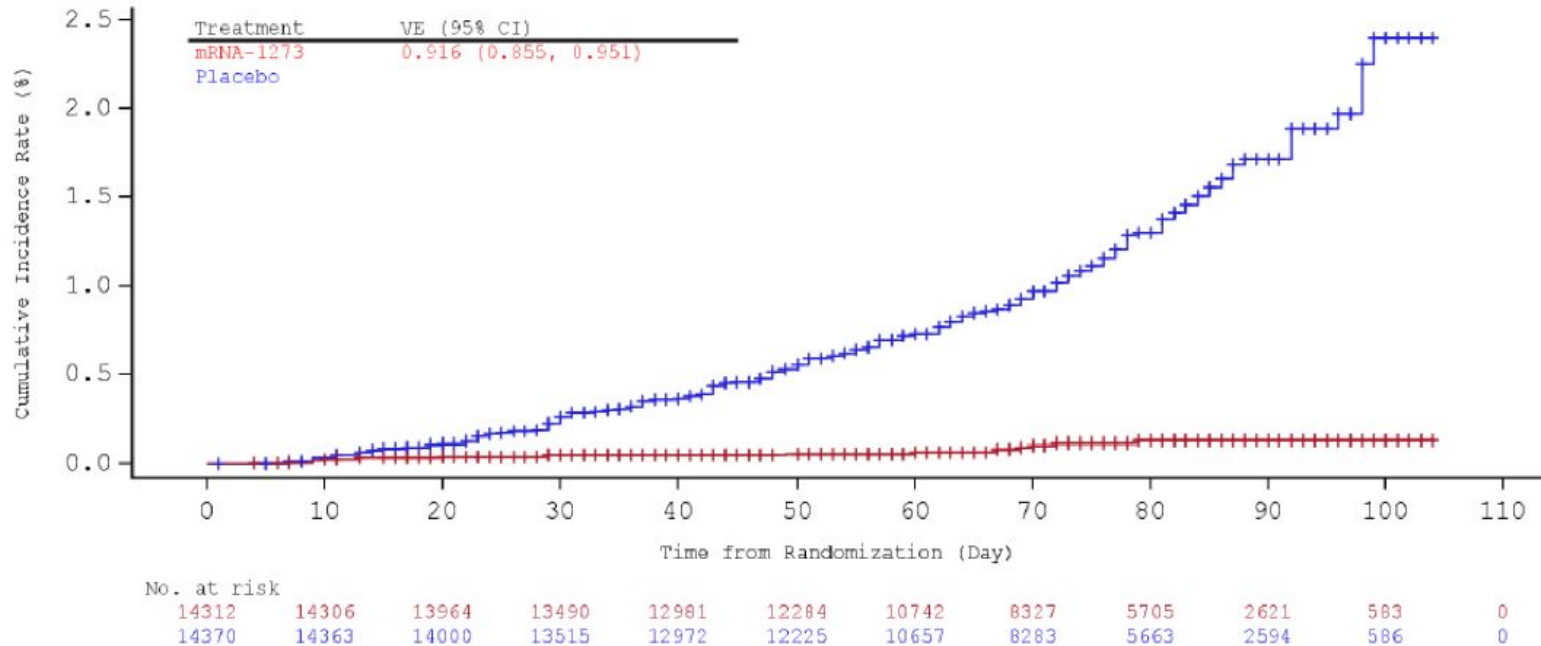
- In the population, more of the potential **treatment** scores are 1 (pain improves) than the potential **control** scores.

(Demo)

Random Assignment & Shuffling



Causality in the Real World



[Source: FDA](#)