[Recitation 5]

Quasi-Experiment, Natural Experiments, And Observational Studies

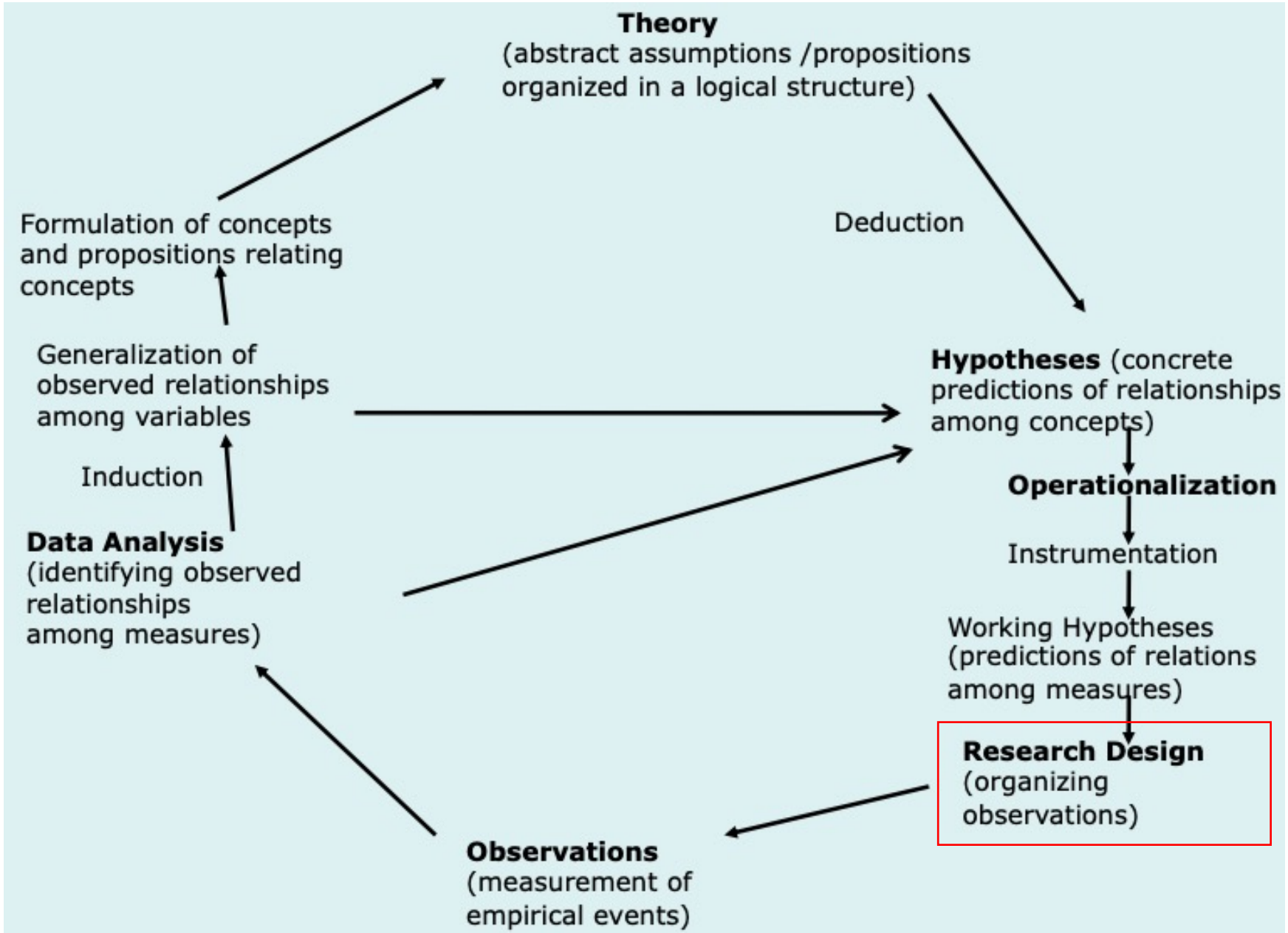
Jungmin Han

The Outline of Today's Recitation

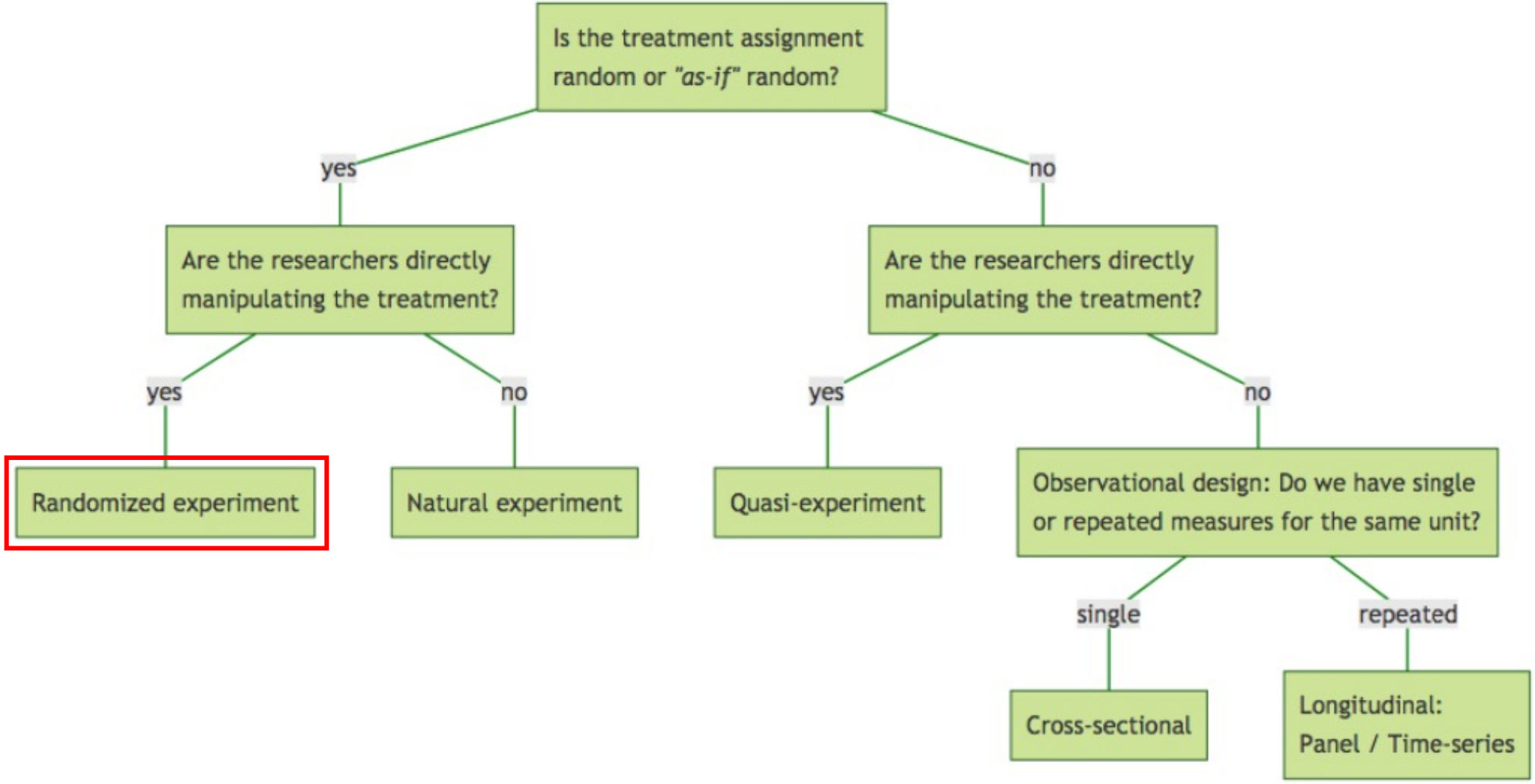
- A Brief Review
  - "True" Experiment vs. Natural Experiment
  - The Strengths and Limitations of Quasi-Experiments
  - "Passive Observation" or "Ex Post Facto" Design
  - The Strengths and Limitations of Passive Observation

- Discussion
  - "Why Demonstrating is Good for Kids" (NYT)
  - "Do 'Fast and Furious' Movies Cause a Rise in Speeding?" (NYT)
  - "Coronavirus Vaccines Work. But This Statistical Illusion Makes People Think They Don't" (Washington Post)

A Model of the Research Process



Types of Research Design



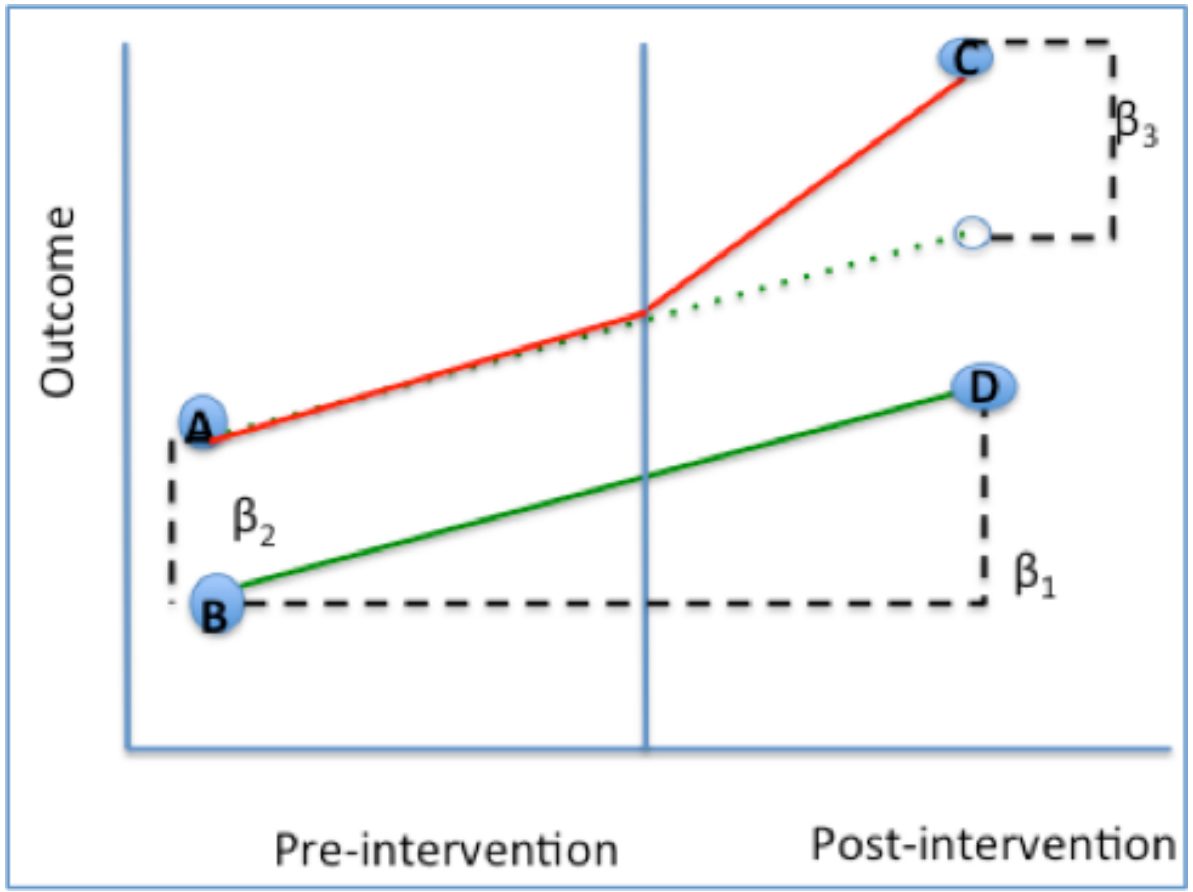
The Classic Experimental Design

		Pre-Test		Post-Test	Difference
Treatment Group	R	M_{1t}	X	M_{2t}	$M_{2t} - M_{1t}$
Control Group	R	M_{1c}		M_{2c}	$M_{2c} - M_{1c}$

1. Create two groups through random assignment
2. Measure the dependent variable
3. Expose one group to the independent variable (Treatment)
4. Measure the dependent variable again

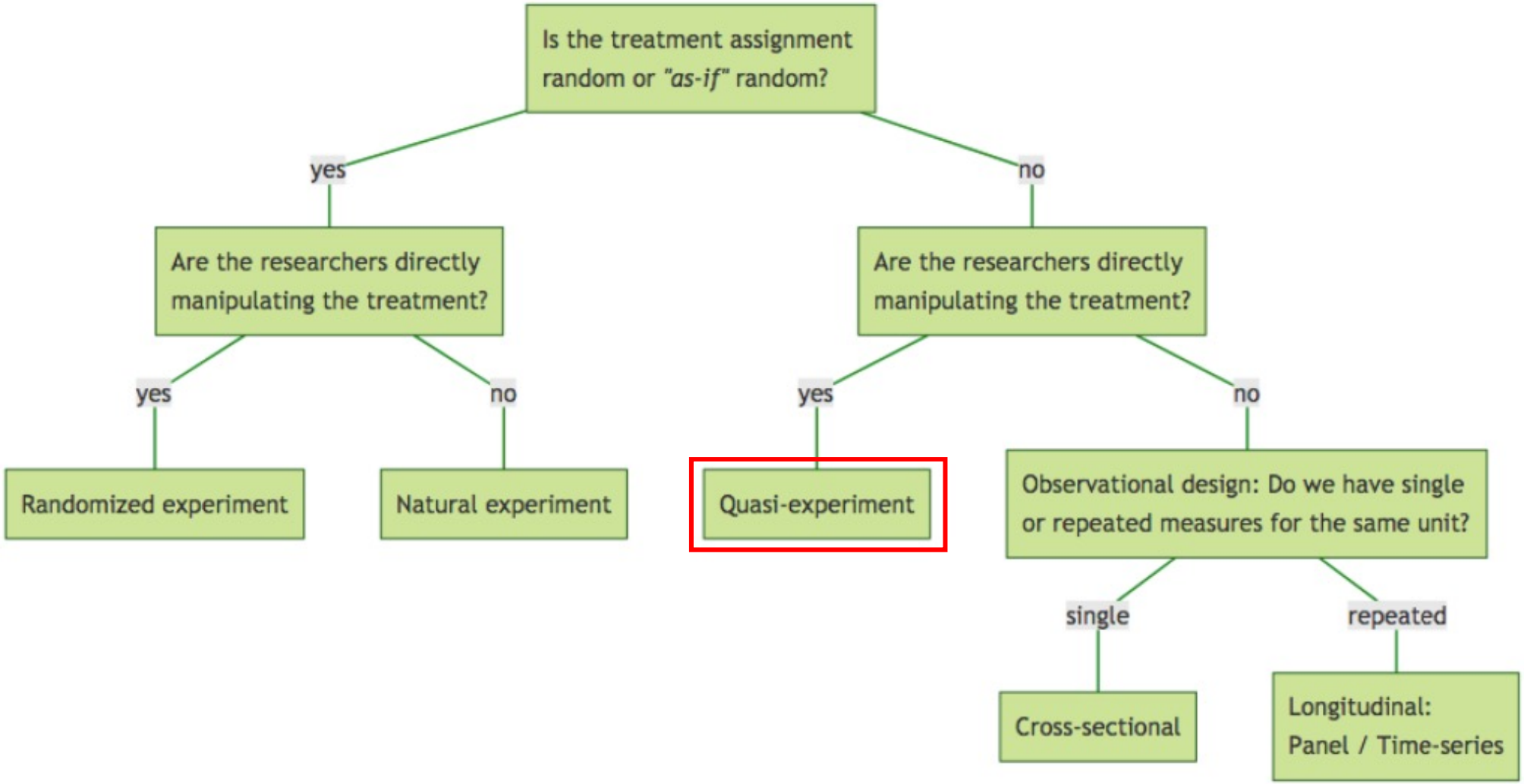
Parallel Trend

Coefficient	Calculation	Interpretation
β_0	B	Baseline average
β_1	D-B	Time trend in control group
β_2	A-B	Difference between two groups pre-intervention
β_3	(C-A)-(D-B)	Difference in changes over time



https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation

Types of Research Design

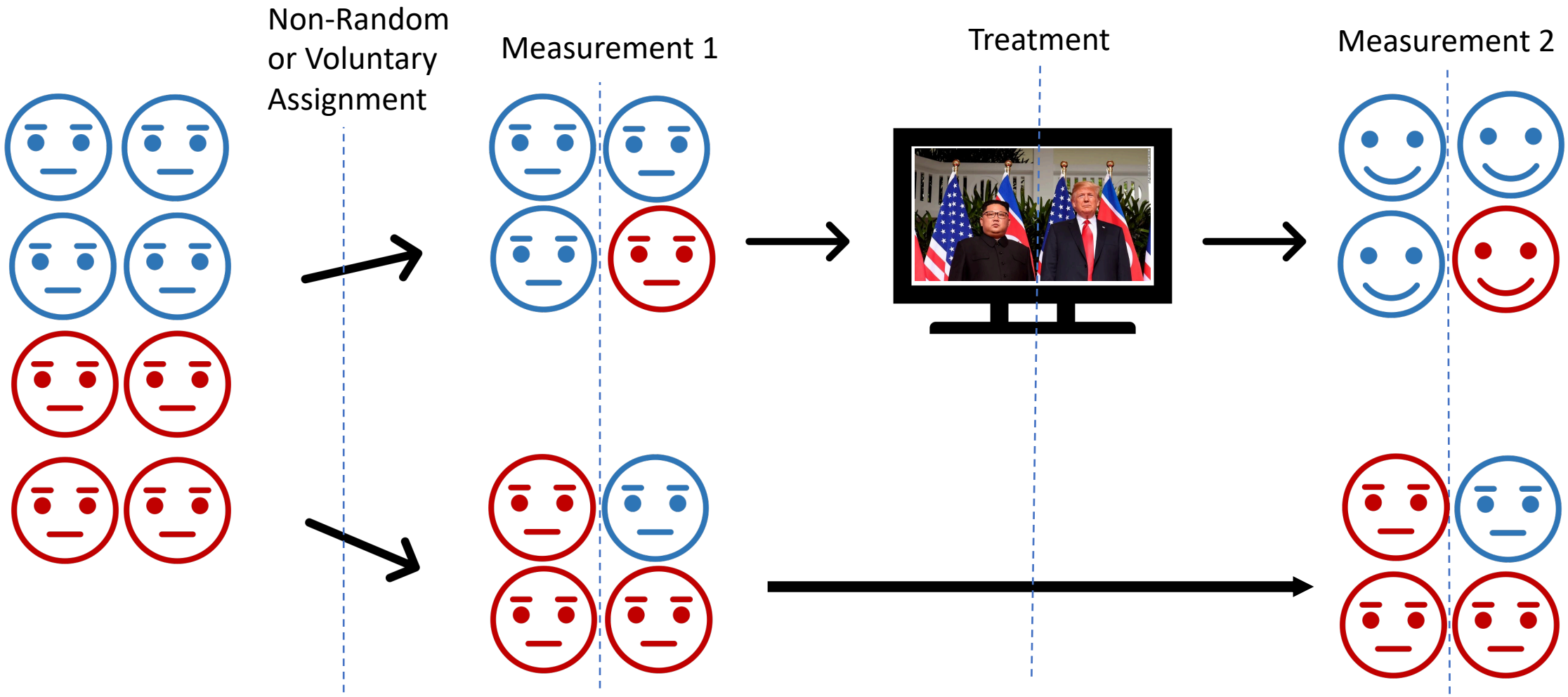


The Classic Quasi-Experimental Design

	Pre-Test		Post-Test	Difference
Treatment Group	M_{1t}	X	M_{2t}	$M_{2t} - M_{1t}$
Control Group	M_{1c}		M_{2c}	$M_{2c} - M_{1c}$

1. Measure the dependent variable
2. Expose one group to the independent variable (Treatment)
3. Measure the dependent variable again

The Classic Quasi-Experimental Design (Example)



Benefits of Quasi-Experiments

- QE studies with pre-tests can **control** for the baseline levels of Y for all groups – the starting points on the dependent variable.

- The researcher directly observes **changes** in Y, and(usually) knows that X came before the changes in Y were observed.

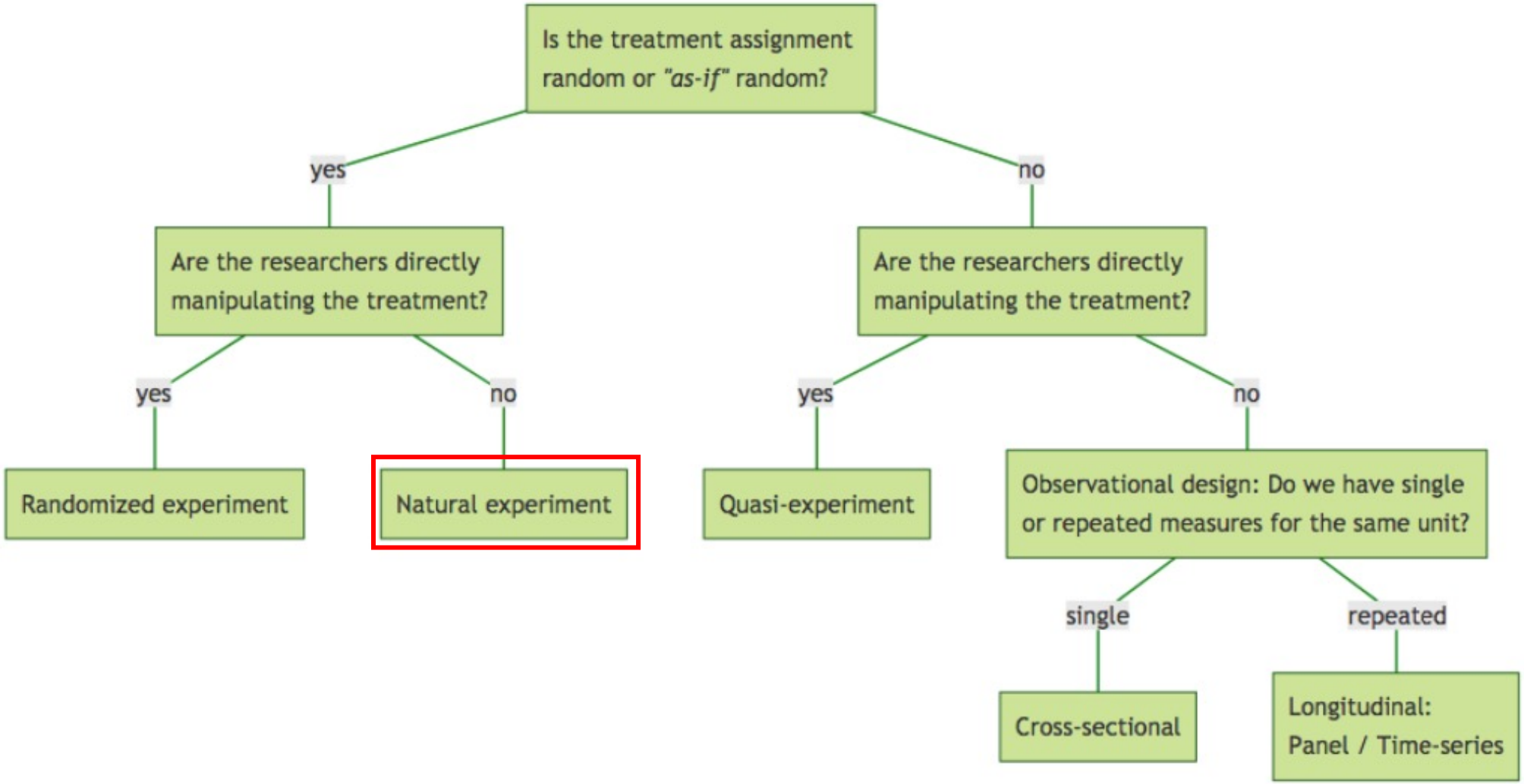
- QE studies often have relatively high **external validity**, as they involve real world observations, without laboratory or other constraints imposed through classic experimental methods.

Limitations of Quasi-Experiments

- As with the classic experiment, the QE researcher may not have a problem of..
  - Not knowing **the true length of time it takes for X to cause Y**
  - Possible **non-compliance** and **drop-out**
  - Possible **unknown variation** in treatment
  - Possible "**contamination**" of the treatment and control groups through **spillover effects**
  - Possibly **unrepresentative sample** sites

- A bigger problem? "Selection Bias" (History, Maturation, and Regression)
  - We can't completely rule out the possibility that **Z confounds** the process and that, therefore, there is not really a causal effect of X on Y.
  - "**Self-selection**": the Z confounds that lead individuals or units to select themselves into the treatment may also influence changes in Y over time.

Types of Research Design

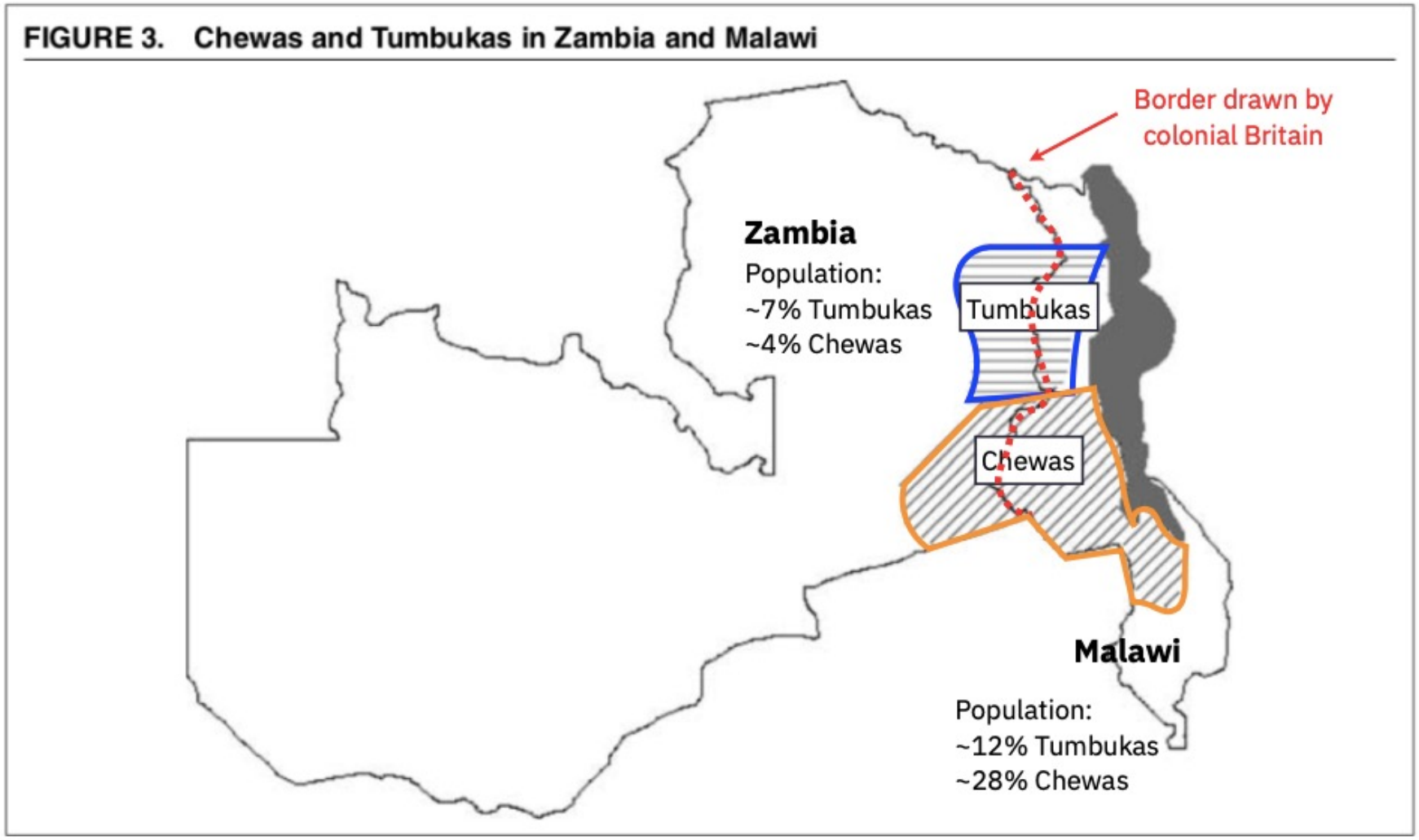


The Classic Natural Design

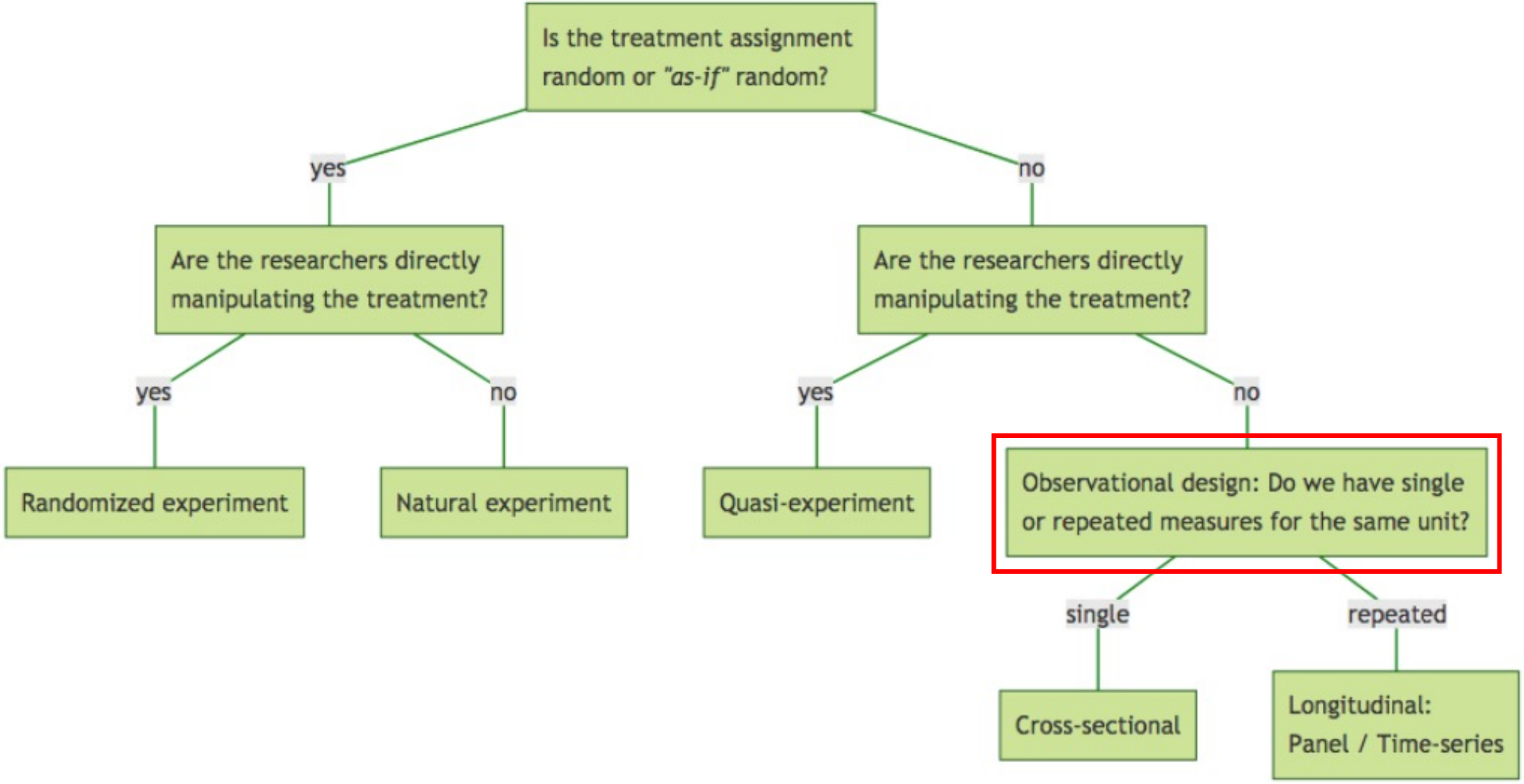
		Pre-Test		Post-Test	Difference
Treatment Group	N	M_{1t}	X	M_{2t}	$M_{2t} - M_{1t}$
Control Group		M_{1c}		M_{2c}	$M_{2c} - M_{1c}$

1. Two groups are created by "as if" random assignment
2. Measure the dependent variable
3. Treatment is given by "nature"
4. Measure the dependent variable again

Journal Article: Posner (2004)



Types of Research Design



"Passive Observation" or "Ex Post Facto" Design

		Post-Test
Units with Value (A) on X ("Treatment Group")	X	M_t
Units with Value (B) on X ("Control Group")		M_c

1. Treatment is given in some ways
2. Measure the dependent variable

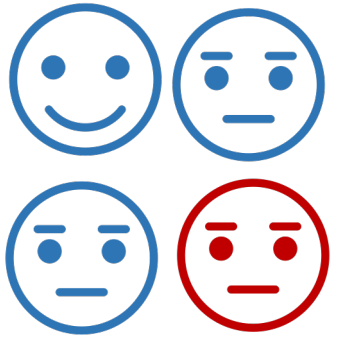
The Observational Design (Example)

Outcome Measurement

Self-report Question?

Watched

Not Watched



Strengths and Limitations of "Passive Observation"

- Strengths
  - **External validity**
  - No manipulation of X

- Limitations
  - **Spuriousness**: No randomization & No pre-test
  - No control of X? **Endogeneity problem** (Y causes X, not X causes Y)

- How to overcome these threats?
  - **Statistical controls** of Zs (multivariate analysis)
  - More time periods and More observations
  - A Panel & Time-series & Multi-level design

Article 1.
"Why Demonstrating Is Good for Kids" (NYT)

- Following the researcher's argument, what does lead young adults who voted, volunteered or engaged in activism to go further in school and have higher incomes?

- Does their evidence necessarily guarantee the causal linkage between activism and better academic and financial outcomes? Why, or why not?

Article 2.
"Do 'Fast and Furious' Movies Cause a Rise in Speeding?" (NYT)

- What is the suggested causal linkage between watching bad behaviors in movies and engaging in risky behaviors in real life? Is this problematic?

- What research design do studies use to overcome such kinds of potential problems in making a causal claim?
  - What is "as-if" random assignment?
  - Do you think this is random enough?

Article 3. “Coronavirus Vaccines Work. But This Statistical Illusion Makes People Think They Don’t” (Washington Post)

- What is Simpson’s paradox? How is it related to the high hospitalization figures in Israel and the high death rate of White people in the US?

- What does make people read data incorrectly? Can you explain this with the concept of “parallel trend”?