USAID
FROM THE AMERICAN PEOPLE

Quasi-experimental Evaluation Design

USAID MENA Advanced MEL Workshop

Session Objectives

By the end of this session, participants will be able to:

- Understand selection bias as a fundamental difficulty in identifying a valid counterfactual
- Recognize randomization as the most effective way to eliminate selection bias
- Identify strategies for generating ‘as-if’ randomization
- Understand the relative strengths and weaknesses of different designs

Level Set

Measuring Social Benefit

We want to know the causal effect of an activity on its beneficiaries

- Job training on earnings and employment
- Teacher qualifications on student outcomes
- Humanitarian assistance on food security

Identifying a Treatment Assignment

- We established the switching equation $Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$ mapping a potential treatment assignment to a potential outcome
- We applied the switching equation to potential outcomes in order to arrive at the difference-in-means estimator
- We decomposed the difference-in-means estimator into the average treatment effect on the treated and selection bias
- We established randomization as the best way to eliminate the selection bias term

Decomposing the Difference in Means

$$E[Y^1 \mid D=1] - E[Y^0 \mid D=0]$$

Average Treatment Effect on the Treated (ATT)

+

$$E[Y^0 \mid D=1] - E[Y^0 \mid D=0]$$

Selection bias

- Under randomization, the selection bias term is zero and the assignment of treatment is independent of the outcome

$$D \perp\!\!\!\perp Y$$

Dealing with Selection Bias

- But what if we cannot randomize the assignment of treatment?
- Without randomization, we don’t have assurance that the selection bias term is zero!
- What else could we do to make the selection bias term zero ?

Approximating ‘as-if’ Randomization

- Adjust for confounders
- Use trends over time to remove bias
- Utilize an intervention’s scoring system
- Use features of the natural world

Adjusting for Confounders

- Let's say we knew that we have non-random assignment of treatment, but we measured EXACTLY what determines treatment (confounder variables X)
- If we control for the confounder variables X then we would achieve independence between the assignment of treatment and the outcome
- This is called unconfoundedness

$$D \perp\!\!\!\perp Y \mid X$$

Controlling For

What does it mean to 'control for' X ?

- Disaggregate across all levels of each variable in X
- Enter variables X into a regression model (regression adjustment)
- Explicitly match treatment and control units on X (future session)

Adjusting for Confounders

- What if we know some variables determine treatment, but we haven't measured them?
- What if there were variables that determine treatment that we aren't aware of?
- These are unobserved confounders, and there's nothing we can do about it

Selection on Observables

- We can adjust for observed variables that help determine treatment
- We could theorize about the variables affecting treatment that are unobserved
- We can’t do anything about variables affecting treatment that we don’t know about

Selection on Observables

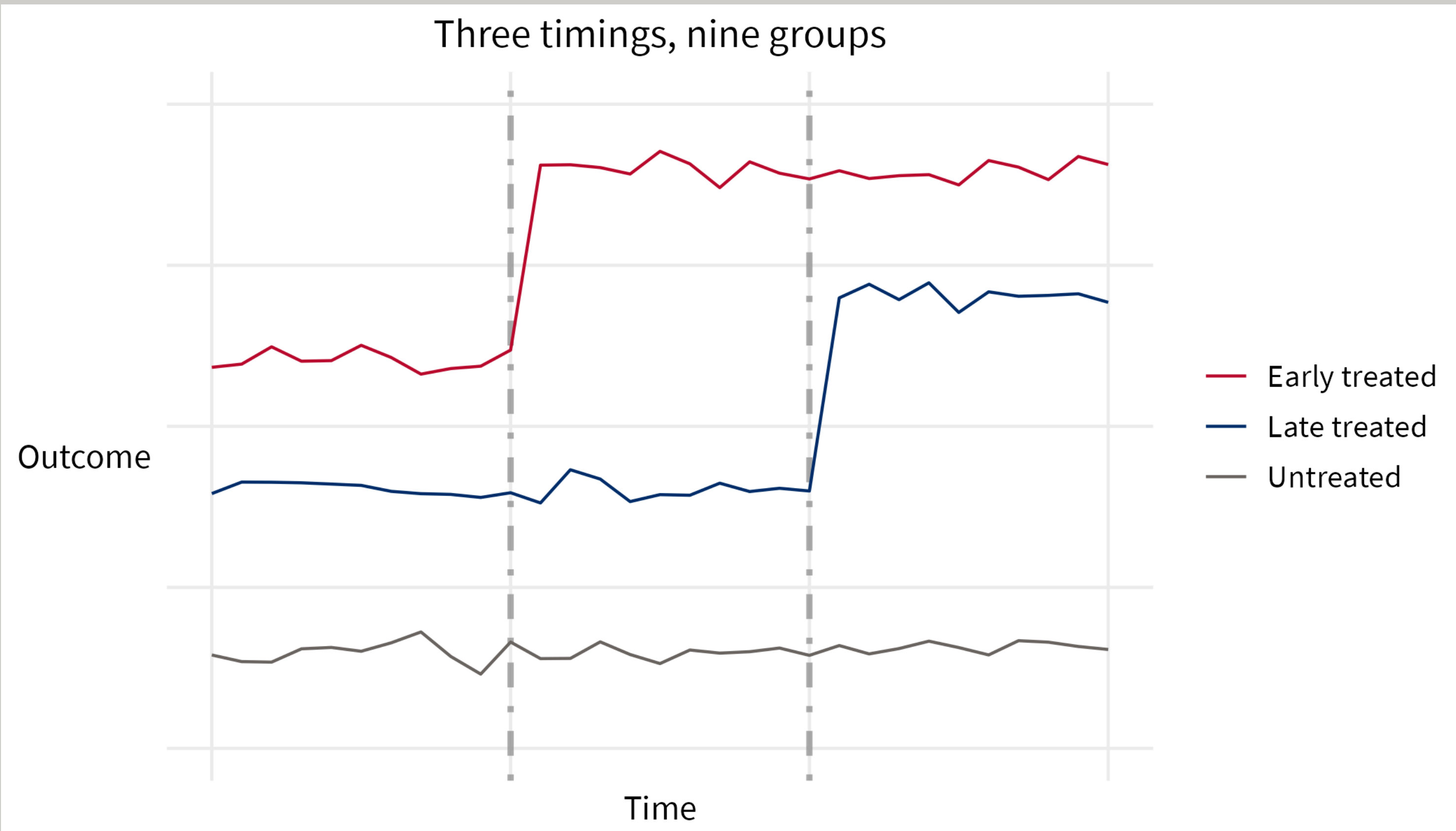
- This highlights the fundamental weakness of adjusting for confounders
- We can control for what we know about and have observed, but there may be other variables we haven't observed
- Controlling for observables is often not a convincing strategy for causal inference

Using Trends Over Time

- The Experimental Evaluation session introduced the randomized difference-in-differences (d-i-d)
- If we have a d-i-d setup WITHOUT randomization, we can use the trend in each group to estimate the counterfactual for the other group
- This only works if we know the pre-treatment trends are the same (parallel) across both groups!
- This is called the parallel trends assumption

Parallel Trends Assumption

- The researcher MUST demonstrate parallel trends, or provide a compelling case for why she believes this assumption holds

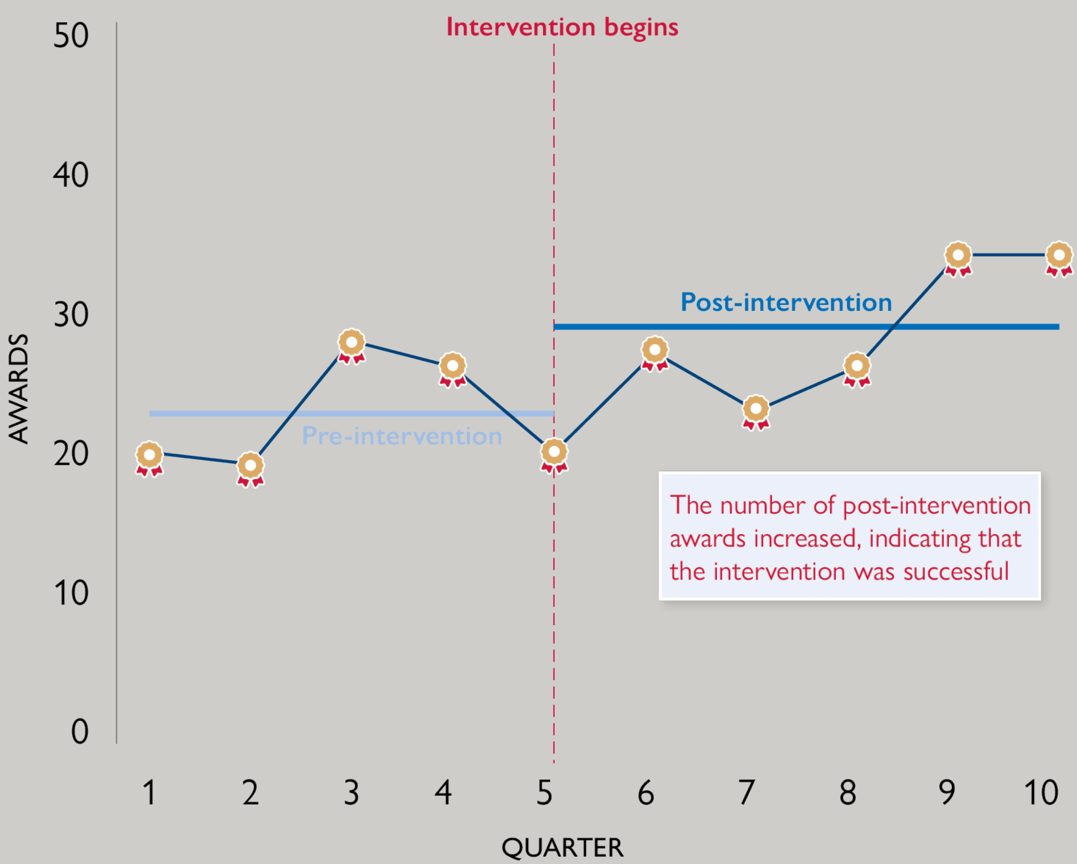


Using Trends Over Time

- Recall that we can't go back in time, and we don't have access to alternate universes
- What if, instead of going back in time, we looked to the past?
- The interrupted time series (ITS) design becomes possible when you have multiple measurements before and after treatment (more is better)

Interrupted Times Series

- Interrupted times series allows a within-comparison (itself) and avoids a between-comparison (something else)



Interrupted Time Series

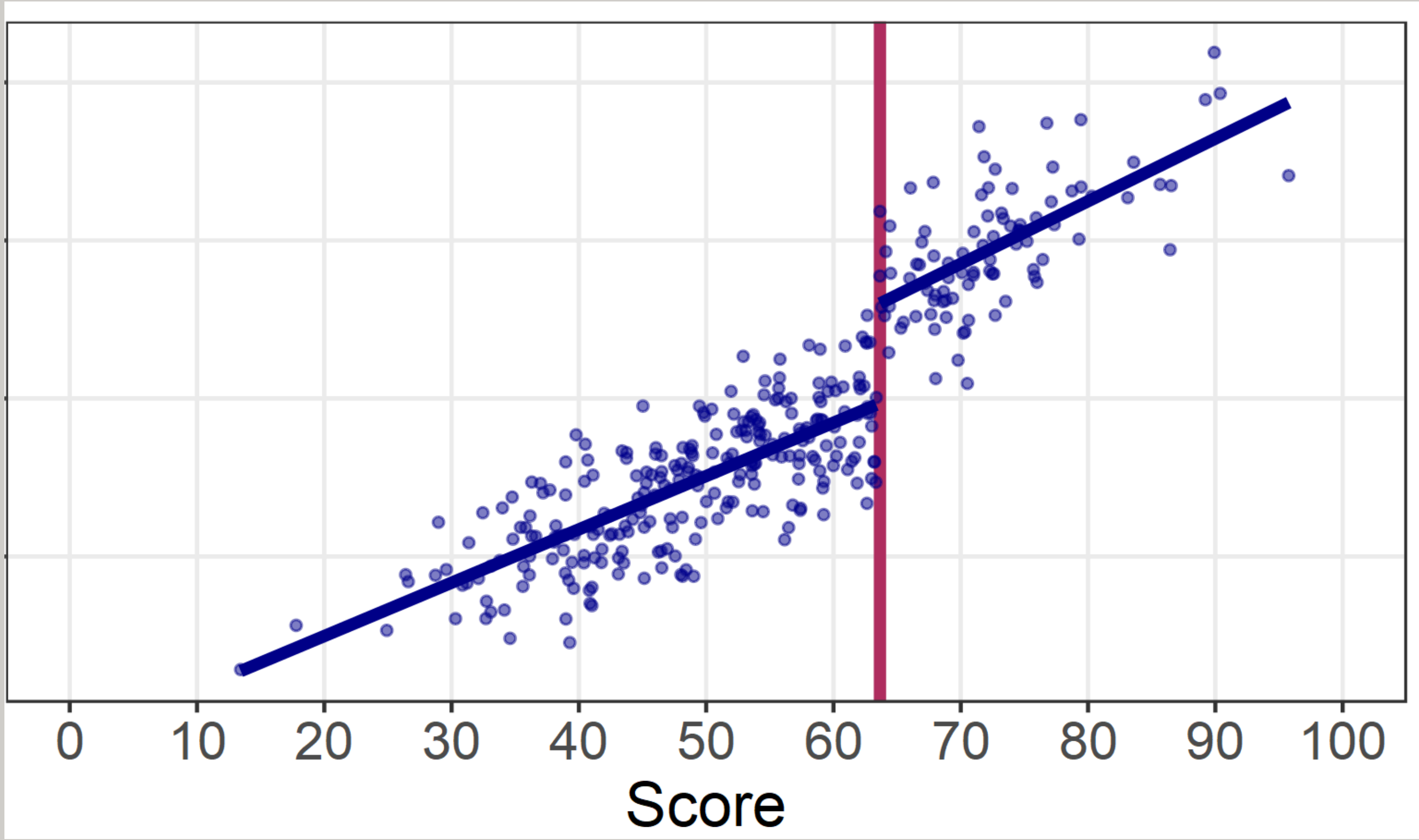
- Interrupted time series is especially applicable when you have abundant data but entire systems or populations are treated
- Because ITS looks at a longer historical period, there may be other historical trends affecting your data
- The historical trend could be within your own treated units (maturation)

Using an Intervention's Scoring System

- Recall that regression adjustment balances observed characteristics across treatment and control groups
- What if there were a variable that divided treatment from the controls around an arbitrary cut-off?
- This is the regression discontinuity (RD) design

Regression Discontinuity

- The most common application of the RD design is when there is a score to determine treatment



Real-world RD

RD applies in any situation where there is a running (score) variable with an arbitrary cutoff that creates ‘as-if’ randomization

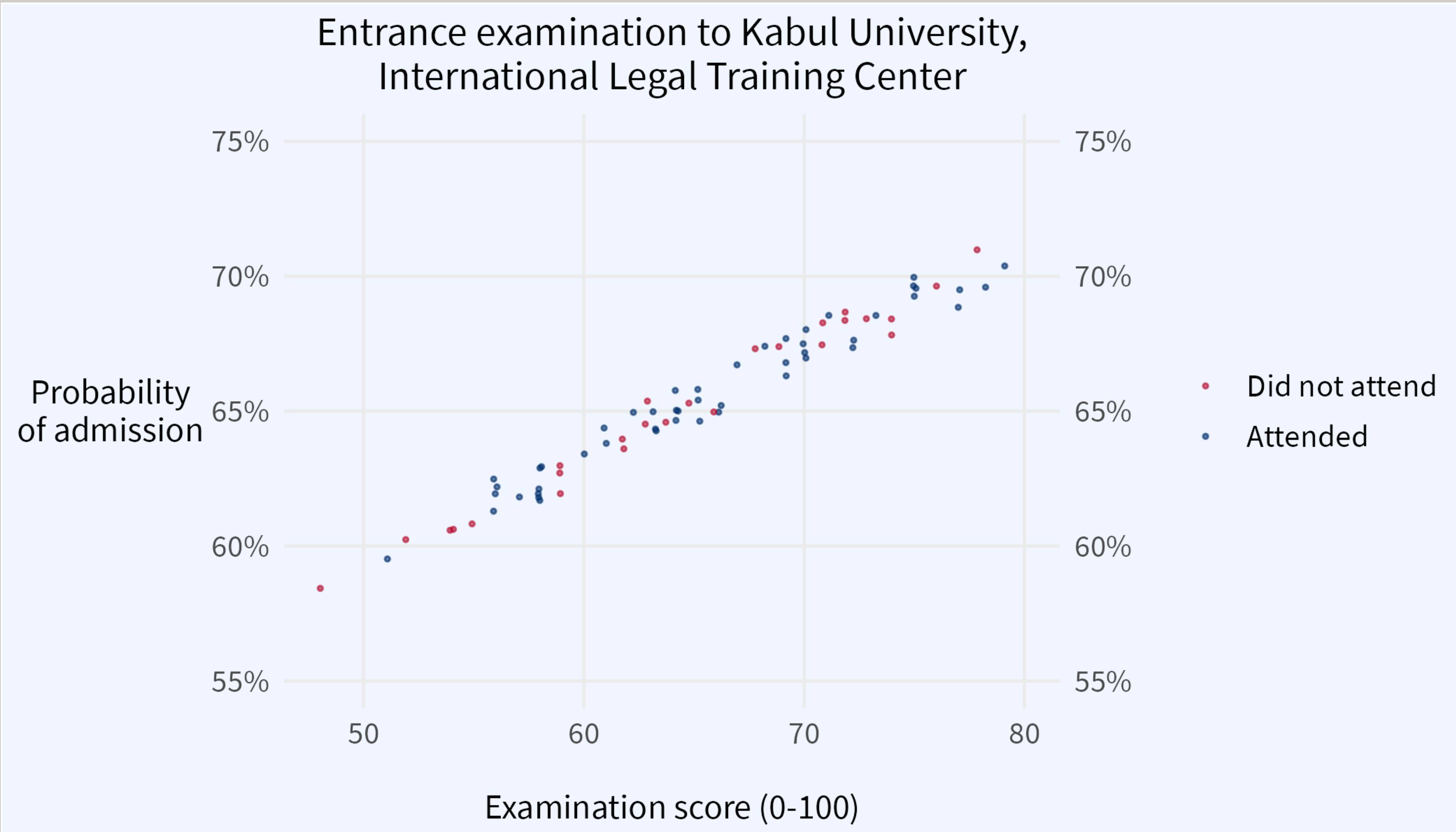
RD Strengths and Limitations

For a valid RD design:

- The cutoff score must be arbitrary
- The score must COMPLETELY determine treatment
- The score cannot be manipulated

RD is Difficult in Development Settings

Where is the cutoff score .. ?



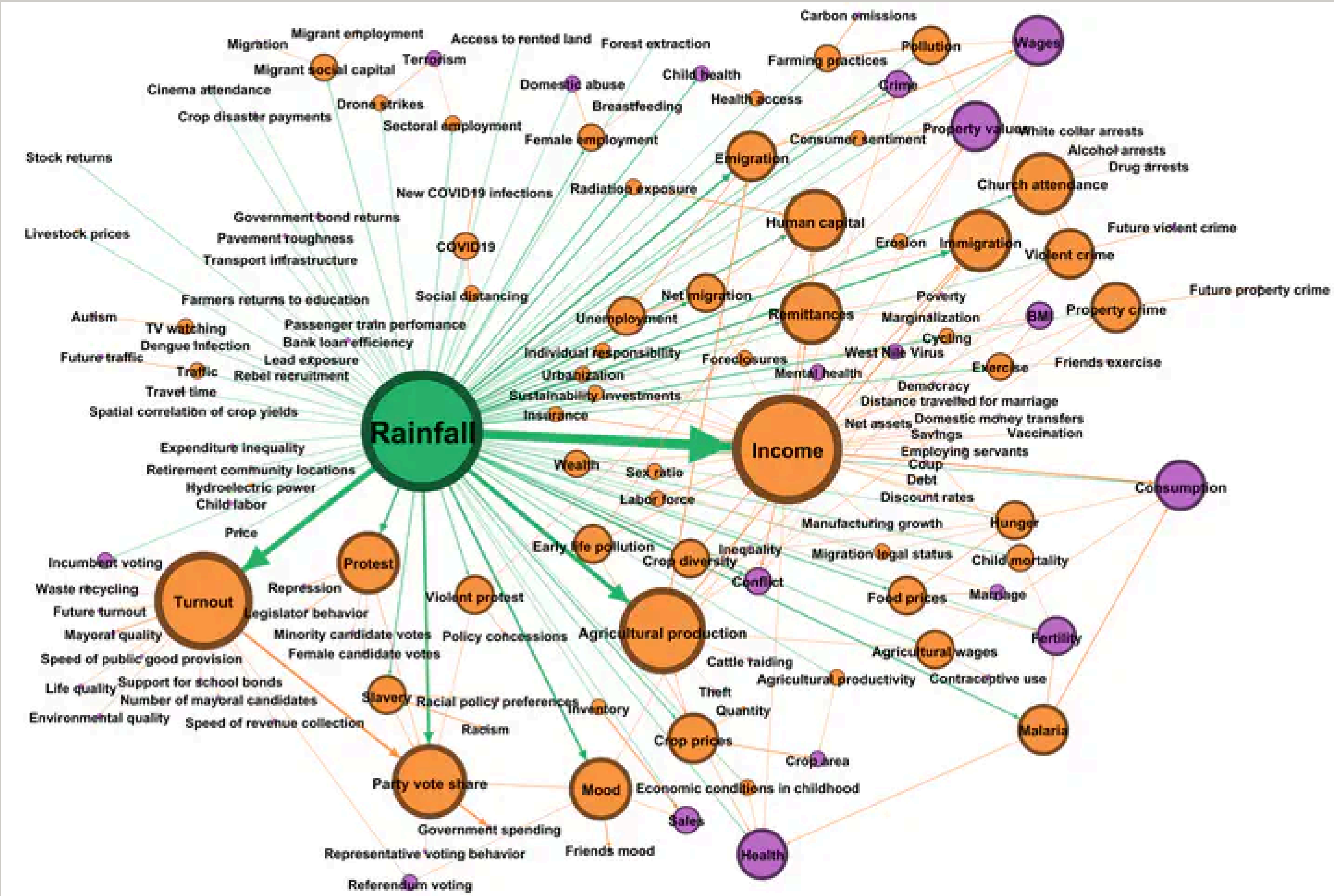
Using Features of the Natural World

- What if we could find some real-world process that behaved randomly with respect to what you are studying?
- We could look at how our outcome responds to treatment, where treatment is ONLY applied based on the real-world random process
- This is instrumental variable estimation

Watch Your Instruments

- The instrument must affect your outcome ONLY through its effect on treatment (exclusion restriction)
- Is rainfall a random process with regard to development interventions?
- But rainfall has now been used as an instrument to generate treatment effect estimates across hundreds of studies

194 Possible Exclusion Restrictions



Ramadan, Economic Activity, and Happiness

Surah Al-Baqarah (2:183-185):

“O you who have believed, decreed upon you is fasting as it was decreed upon those before you that you may become righteous”

“So whoever sights the new moon of the month, let him fast it”

Is the Lunar Calendar a Valid Instrument?

- The lunar calendar moves Ramadan 11 days per year
- This also affects the number of hours of fasting per day
- Cross-reference fasting hours per day with economic activity or subjective well-being
- Exclusion restriction: Does the moon affect economic activity or subjective well-being in any other ways?

Review

- We have reviewed evaluation designs that seek to generate 'as-if' randomization
- This 'as-if' randomization must eliminate selection bias
- Experimental designs: $D \perp\!\!\!\perp Y$
- Quasi-experimental designs: $D \perp\!\!\!\perp Y \mid X$

What are the Strengths and Limitations?

- Regression adjustment / matching
- Difference-in-differences
- Interrupted time series
- Regression discontinuity
- Instrumental variables

Quasi-experimental Evaluation Designs

- Researchers must fully engage with the limitations of their design, and address how they mitigated threats to validity
- Stay tuned for a deeper dive on matching, and a case study on the use of advanced d-i-d methods in a real-world evaluation

Thank you!