

Session IV

Practical Issues

Thomas J. Leeper

Government Department
London School of Economics and Political Science

- 1 Beyond One-Shot Designs
- 2 Handling “Broken” Experiments
- 3 Research Ethics
- 4 Conclusion

- 1 Beyond One-Shot Designs
- 2 Handling “Broken” Experiments
- 3 Research Ethics
- 4 Conclusion

Beyond One-shot Designs

- Surveys can be used as a measurement instrument for a field treatment or a manipulation applied in a different survey panel wave
 - 1 Measure effect duration in two-wave panel
 - 2 Solicit pre-treatment outcome measures in a two-wave panel
 - 3 Measure effects of field treatment in post-test only design
 - 4 Randomly encourage field treatment in pre-test and measure effects in post-test

Beyond One-shot Designs

- Surveys can be used as a measurement instrument for a field treatment or a manipulation applied in a different survey panel wave
 - 1 Measure effect duration in two-wave panel
 - 2 Solicit pre-treatment outcome measures in a two-wave panel
 - 3 Measure effects of field treatment in post-test only design
 - 4 Randomly encourage field treatment in pre-test and measure effects in post-test
- Problems? Compliance & nonresponse

I. Effect Duration

- Use a two- (or more-) wave panel to measure duration of effects
 - T1: Treatment and outcome measurement
 - T2+: Outcome measurement
- Two main concerns
 - Attrition
 - Panel conditioning

II. Within-Subjects Designs

- Estimate treatment effects as a difference-in-differences
- Instead of using the post-treatment mean-difference in Y to estimate the causal effect, use the difference in pre-post differences for the two groups:

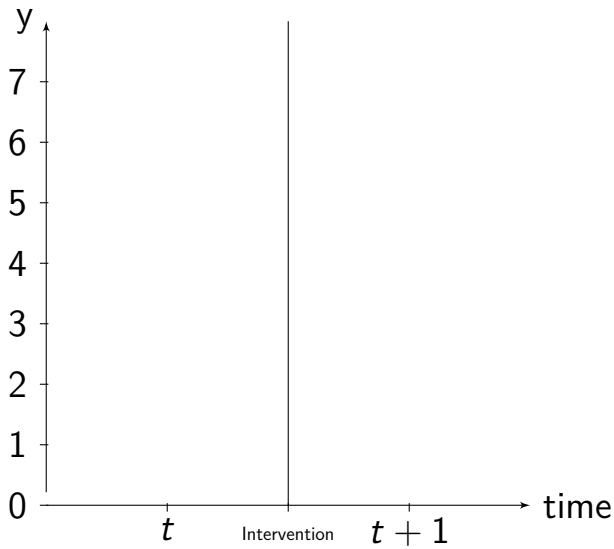
$$(\hat{Y}_{0,t+1} - \hat{Y}_{0,t}) - (\hat{Y}_{j,t+1} - \hat{Y}_{j,t})$$

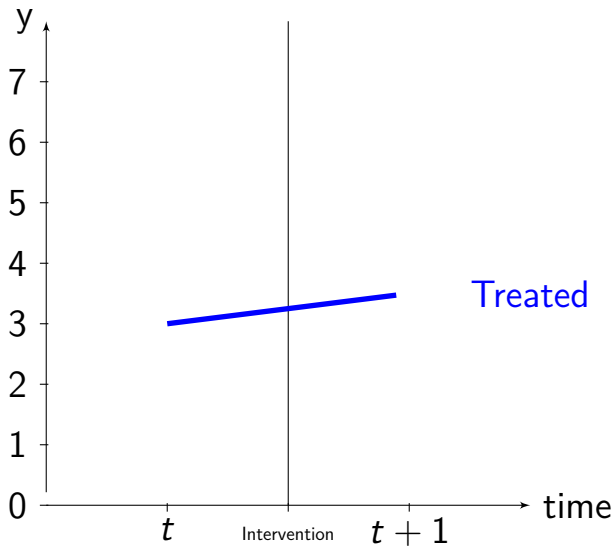
II. Within-Subjects Designs

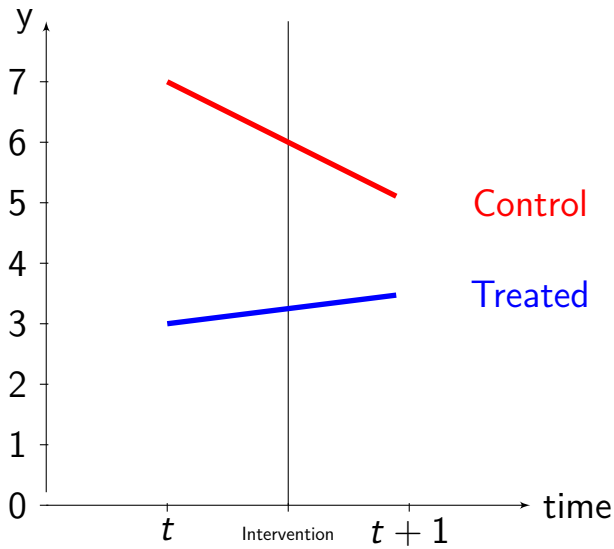
- Estimate treatment effects as a difference-in-differences
- Instead of using the post-treatment mean-difference in Y to estimate the causal effect, use the difference in pre-post differences for the two groups:

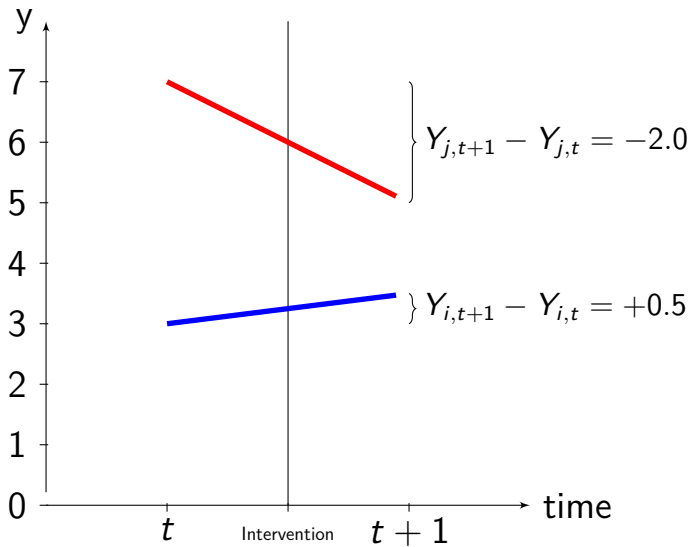
$$(\hat{Y}_{0,t+1} - \hat{Y}_{0,t}) - (\hat{Y}_{j,t+1} - \hat{Y}_{j,t})$$

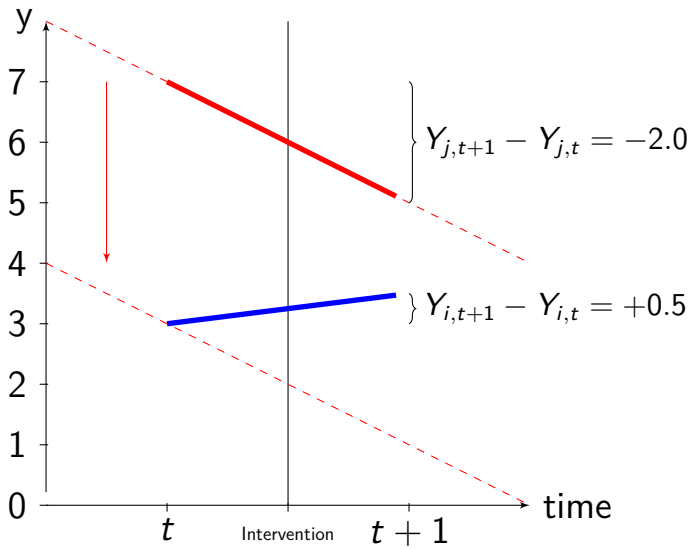
- Advantageous because variance for paired samples decreases as correlation between t_0 and t_1 observations increases

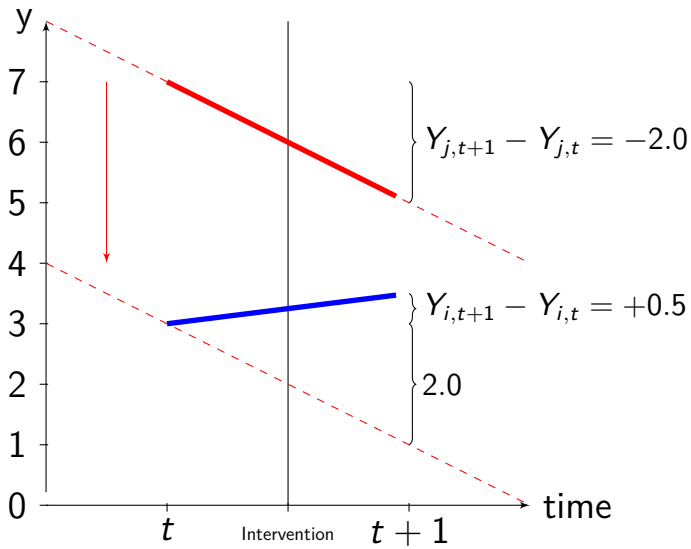


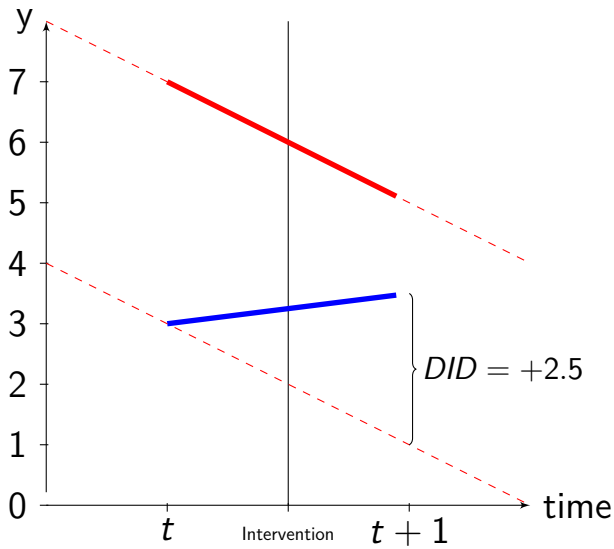












Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.¹

¹Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.¹

- 1 History (simultaneous cause)

¹Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.¹

- 1 History (simultaneous cause)
- 2 Maturation (time trends)

¹Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.¹

- 1 History (simultaneous cause)
- 2 Maturation (time trends)
- 3 Testing (observation changes respondents)

¹Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.¹

- 1 History (simultaneous cause)
- 2 Maturation (time trends)
- 3 Testing (observation changes respondents)
- 4 Instrumentation (changing operationalization)

¹Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.¹

- 1 History (simultaneous cause)
- 2 Maturation (time trends)
- 3 Testing (observation changes respondents)
- 4 Instrumentation (changing operationalization)
- 5 Instability (measurement error)

¹Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.¹

- 1 History (simultaneous cause)
- 2 Maturation (time trends)
- 3 Testing (observation changes respondents)
- 4 Instrumentation (changing operationalization)
- 5 Instability (measurement error)
- 6 Attrition

¹Shadish, Cook, and Campbell (2002)

III. Randomized Field Treatment

- Examples:

III. Randomized Field Treatment

- Examples:
 - 1 Citizens randomly sent a letter by post encouraging them to reduce water usage

III. Randomized Field Treatment

- Examples:
 - 1 Citizens randomly sent a letter by post encouraging them to reduce water usage
 - 2 Different local media markets randomly assigned to receive different advertising

III. Randomized Field Treatment

- Examples:
 - 1 Citizens randomly sent a letter by post encouraging them to reduce water usage
 - 2 Different local media markets randomly assigned to receive different advertising
- Survey is used to measure outcomes, when treatment assignment is already known

III. Randomized Field Treatment

- Examples:
 - 1 Citizens randomly sent a letter by post encouraging them to reduce water usage
 - 2 Different local media markets randomly assigned to receive different advertising
- Survey is used to measure outcomes, when treatment assignment is already known
- Issues

III. Randomized Field Treatment

- Examples:
 - 1 Citizens randomly sent a letter by post encouraging them to reduce water usage
 - 2 Different local media markets randomly assigned to receive different advertising
- Survey is used to measure outcomes, when treatment assignment is already known
- Issues
 - Nonresponse
 - Noncompliance

Noncompliance

- Compliance is when individuals receive and accept the treatment to which they are assigned
- Noncompliance:
“when subjects who were assigned to receive the treatment go untreated or when subjects assigned to the control group are treated” ²
- This causes problems for our analysis because factors other than randomization explain why individuals receive their treatment
- Lots of methods for dealing with this, but the consequence is generally reduced power

²Gerber & Green. 2012. *Field Experiments*, p.132.

Asymmetric Noncompliance

- Noncompliance *asymmetric* if only in one group
- We can ignore non-compliance and analyze the “intention to treat” effect, which will underestimate our effects because some people were not treated as assigned
$$ITT = \overline{Y}_1 - \overline{Y}_0$$
- We can use “instrumental variables” to estimate the “local average treatment effect” (LATE) for those that complied with treatment:
$$LATE = \frac{ITT}{PercentCompliant}$$
- We can ignore randomization and analyze data “as-treated”, but this makes our study no longer an experiment

Local Average Treatment Effect

- IV estimate is *local* to the variation in X that is due to variation in D
- LATE is effect for those who *comply*
- Four subpopulations:
 - Compliers: $X = 1$ only if $D = 1$
 - Always-takers: $X = 1$ regardless of D
 - Never-takers: $X = 0$ regardless of D
 - Defiers: $X = 1$ only if $D = 0$
- Exclusion restriction! Monotonicity!

Two-Sided Noncompliance

- Two-sided noncompliance is more complex analytically
- Stronger assumptions are required to analyze it and we won't discuss them here
- Best to try to develop a better design to avoid this rather than try to deal with the complexities of analyzing a broken design

IV. Treatment Encouragement

- Design:
 - T1: Encourage treatment
 - T2: Measure effects
- Examples:
 - 1 Albertson and Lawrence³

³Albertson & Lawrence. 2009. "After the Credits Roll." *American Politics Research* 37(2): 275–300.
10.1177/1532673X08328600.

IV. Treatment Encouragement

- Design:
 - T1: Encourage treatment
 - T2: Measure effects
- Examples:
 - 1** Albertson and Lawrence³
- Issues

³Albertson & Lawrence. 2009. "After the Credits Roll." *American Politics Research* 37(2): 275–300.
10.1177/1532673X08328600.

IV. Treatment Encouragement

- Design:
 - T1: Encourage treatment
 - T2: Measure effects
- Examples:
 - 1 Albertson and Lawrence³
- Issues
 - Nonresponse
 - Noncompliance

³Albertson & Lawrence. 2009. "After the Credits Roll." *American Politics Research* 37(2): 275–300.
10.1177/1532673X08328600.

Treatment Noncompliance

Treatment Noncompliance

- Several strategies
 - “As treated” analysis
 - “Intention to treat” analysis
 - Estimate a LATE

Questions?

1 Beyond One-Shot Designs

2 Handling “Broken” Experiments

3 Research Ethics

4 Conclusion

Quiz time!

Compliance

- 1 What is compliance?

Compliance

- 1 What is compliance?
- 2 How can we analyze experimental data when there is noncompliance?

Balance testing

- 1 What does randomization ensure about the composition of treatment groups?

Balance testing

- 1 What does randomization ensure about the composition of treatment groups?
- 2 What can we do if we find a covariate imbalance between groups?

Balance testing

- 1 What does randomization ensure about the composition of treatment groups?
- 2 What can we do if we find a covariate imbalance between groups?
- 3 How can we avoid this problem entirely?

Nonresponse and Attrition

- 1 Do we care about outcome nonresponse in experiments?

Nonresponse and Attrition

- 1 Do we care about outcome nonresponse in experiments?
- 2 How can we analyze experimental data when there is outcome nonresponse or post-treatment attrition?

Manipulation checks

- 1 What is a manipulation check? What can we do with it?

Manipulation checks

- 1 What is a manipulation check? What can we do with it?
- 2 What do we do if some respondents “fail” a manipulation check?

Null effects

- 1 What should we do if we find our estimated $\widehat{SATE} = 0$?

Null effects

- 1 What should we do if we find our estimated $\widehat{SATE} = 0$?
- 2 What does it mean for an experiment to be *underpowered*?

Null effects

- 1 What should we do if we find our estimated $\widehat{SATE} = 0$?
- 2 What does it mean for an experiment to be *underpowered*?
- 3 What can we do to reduce the probability of obtaining an (unwanted) “null effect”?

Effect heterogeneity

- 1 What should we do if, post-hoc, we find evidence of effect heterogeneity?

Effect heterogeneity

- 1 What should we do if, post-hoc, we find evidence of effect heterogeneity?
- 2 What can we do pre-implementation to address possible heterogeneity?

Representativeness

- 1 Under what conditions is a design-based, probability sample necessary for experimental inference?

Representativeness

- 1 Under what conditions is a design-based, probability sample necessary for experimental inference?
- 2 What kind of causal inferences can we draw from an experiment on a descriptively unrepresentative sample?

Peer Review

- 1 What should we do if a peer reviewer asks us to “control” for covariates in the analysis?

Peer Review

- 1 What should we do if a peer reviewer asks us to “control” for covariates in the analysis?
- 2 What should we do if a peer reviewer asks us to include or exclude particular respondents from the analysis?

Questions?

- 1 Beyond One-Shot Designs
- 2 Handling “Broken” Experiments
- 3 Research Ethics**
- 4 Conclusion

History: Key Moments

- 1 Tuskegee (1932-1972) and Guatemala (1946-1948) Studies
- 2 Nuremberg Code (1947)
- 3 Helsinki Declaration (1964)
- 4 U.S. 45 CFR 46 (1974) and “Common Rule” (1991)
- 5 The Belmont Report (1979)
- 6 EU Data Protection Directive (1995; 2012)
 - UK Data Protection Act (1998)

Helsinki Declaration

- Adopted by the World Medical Association in 1964⁴
- Narrowly focused on medical research
- Expanded the Nuremberg Code
 - Relaxed consent requirements
 - Risks should not exceed benefits
 - Institutionalization of ethics oversight

⁴<http://www.bmj.com/content/2/5402/177>

Helsinki Declaration

- Adopted by the World Medical Association in 1964⁴
- Narrowly focused on medical research
- Expanded the Nuremberg Code
 - Relaxed consent requirements
 - Risks should not exceed benefits
 - Institutionalization of ethics oversight
- Do these rules apply to non-medical research?

⁴<http://www.bmj.com/content/2/5402/177>

The Belmont Report

- Commissioned by the U.S. Government in 1979⁵
- Three overarching principles:
 - 1 Respect for persons
 - 2 Beneficence
 - 3 Justice
- Three policy implications:
 - Informed consent
 - Assessment of risks/benefits
 - Care for vulnerable populations

⁵<http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>

Benefits and Harm

- What is a “benefit”?
- What is a “harm”?
- How do we balance the two?

Ethical Considerations

- Most ethical issues are not unique to *experimental* social science
- Some especially important issues:
 - 1 Randomization
 - 2 Informed consent
 - 3 Privacy
 - 4 Deception
 - 5 Publication bias

I. Randomization

- Is it ethical to randomize?

II. Informed Consent

- Persons must consent to being a research subject

II. Informed Consent

- Persons must consent to being a research subject
- What this means in practice is complicated
 - What is consent?
 - What is “informed” consent?
 - What exactly do they have to consent to?

II. Informed Consent

- Persons must consent to being a research subject
- What this means in practice is complicated
 - What is consent?
 - What is “informed” consent?
 - What exactly do they have to consent to?
- Cross-national variations
 - Consent forms required in U.S.
 - Not required in UK

III. Privacy

- Under EU Data Protection Directive (1995), data can be processed when:
 - Consent is given
 - Data are used for a “legitimate” purpose
 - Anonymous or confidential
- Data cannot leave the EU except under conditions

III. Privacy

- Experimental might be additionally sensitive

III. Privacy

- Experimental might be additionally sensitive
- Answers reflect “manipulated” attitudes, behaviors, perceptions, etc. that respondents may not have given in another setting

IV. Deception

- Major distinction between psychology tradition and economics tradition⁶
 - Purpose of the study
 - Purpose of specific items or tasks
 - Order or length of questionnaire

⁶Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

IV. Deception

- Major distinction between psychology tradition and economics tradition⁶
 - Purpose of the study
 - Purpose of specific items or tasks
 - Order or length of questionnaire
- Psychologists focus on *debriefing*

⁶Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

IV. Deception

- Major distinction between psychology tradition and economics tradition⁶
 - Purpose of the study
 - Purpose of specific items or tasks
 - Order or length of questionnaire
- Psychologists focus on *debriefing*
- Within economics, norms about *acts of omission* versus *acts of commission*

⁶Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

IV. Deception

- Major distinction between psychology tradition and economics tradition⁶
 - Purpose of the study
 - Purpose of specific items or tasks
 - Order or length of questionnaire
- Psychologists focus on *debriefing*
- Within economics, norms about *acts of omission* versus *acts of commission*
 - Omission: In a multi-round trust game, an additional round is added

⁶Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

IV. Deception

- Major distinction between psychology tradition and economics tradition⁶
 - Purpose of the study
 - Purpose of specific items or tasks
 - Order or length of questionnaire
- Psychologists focus on *debriefing*
- Within economics, norms about *acts of omission* versus *acts of commission*
 - Omission: In a multi-round trust game, an additional round is added
 - Commission: Telling respondents it is a dictator game, but it is actually a trust game

⁶Dickson, E. 2011. "Economics versus Psychology Experiments." *Cambridge Handbook of Experimental Political Science*.

V. Publication Bias

- Publication bias not typically discussed as an ethical question

V. Publication Bias

- Publication bias not typically discussed as an ethical question
- If studies are meant to policy or practical implications, then we care about PATE or a set of CATEs, including whether their effects are positive, negative, or zero.

V. Publication Bias

- Publication bias not typically discussed as an ethical question
- If studies are meant to policy or practical implications, then we care about PATE or a set of CATEs, including whether their effects are positive, negative, or zero.
- Publication bias (toward “significant” results) invites wasting resources on treatments that actually don’t work

Lots of Other Ethical Questions

Lots of Other Ethical Questions

1 Funding

Lots of Other Ethical Questions

- 1 Funding
- 2 Independence and Politicization

Lots of Other Ethical Questions

- 1 Funding
- 2 Independence and Politicization
- 3 Vulnerable populations (e.g. children, sick)

Lots of Other Ethical Questions

- 1 Funding
- 2 Independence and Politicization
- 3 Vulnerable populations (e.g. children, sick)
- 4 Incentives

Lots of Other Ethical Questions

- 1 Funding
- 2 Independence and Politicization
- 3 Vulnerable populations (e.g. children, sick)
- 4 Incentives
- 5 Cross-national research

Lots of Other Ethical Questions

- 1 Funding
- 2 Independence and Politicization
- 3 Vulnerable populations (e.g. children, sick)
- 4 Incentives
- 5 Cross-national research
- 6 End uses/users of research

Lots of Other Ethical Questions

- 1 Funding
- 2 Independence and Politicization
- 3 Vulnerable populations (e.g. children, sick)
- 4 Incentives
- 5 Cross-national research
- 6 End uses/users of research
- 7 Others. . .

Questions?

- 1 Beyond One-Shot Designs
- 2 Handling “Broken” Experiments
- 3 Research Ethics
- 4 Conclusion**

Learning Outcomes

By the end of the week, you should be able to...

Learning Outcomes

By the end of the week, you should be able to...

- 1 Explain how to analyze experiments quantitatively.

Learning Outcomes

By the end of the week, you should be able to...

- 1 Explain how to analyze experiments quantitatively.
- 2 Explain how to design experiments that speak to relevant research questions and theories.

Learning Outcomes

By the end of the week, you should be able to...

- 1 Explain how to analyze experiments quantitatively.
- 2 Explain how to design experiments that speak to relevant research questions and theories.
- 3 Evaluate the uses and limitations of several common survey experimental paradigms.

Learning Outcomes

By the end of the week, you should be able to...

- 1 Explain how to analyze experiments quantitatively.
- 2 Explain how to design experiments that speak to relevant research questions and theories.
- 3 Evaluate the uses and limitations of several common survey experimental paradigms.
- 4 Identify practical issues that arise in the implementation of experiments and evaluate how to anticipate and respond to them.

Wrap-up

- Thanks to all of you!
- Stay in touch (t.leeper@lse.ac.uk)
- Good luck with your research!