Research Design

Beyond Randomized Control Trials

Jody Worley, Ph.D.
College of Arts & Sciences
Human Relations

Introduction to the series

- Day 1: Nonrandomized Designs
- Day 2: Sampling Strategies
- Day 3: Matching Techniques for Balanced Designs

Nonrandomized Designs

1. What are Randomized Control Trials and why are they considered the "Gold Standard" in research design?
2. If the primary goal in quantitative research is to establish some degree of certainty about causal relationships, are there any viable alternatives to RCT?
3. Non-randomized designs *are* the only option when random assignment is not possible, or not ethically appropriate.
    - What are our options, then, among the various non-randomized design protocols?
    - What are the critical considerations for research design protocol?
    - What are the advantages and disadvantages for each approach?

What are Randomized Control Trials (RCT) and why are they considered the “Gold Standard”?

- Strong control over extraneous variables
- Primary goal is understanding the “average” overall benefit and risk
- Efficacious and Effective
  - **Efficacy** refers to the beneficial effects under optimal conditions
  - **Effectiveness** refers to effects under more real-world conditions

Potential Reasons for Differences Between the Efficacy and Effectiveness of a Treatment

Study Design

- Setting: Standardized, highly controlled –v– Naturalistic, unsystematic
- Outcomes include less meaningful end points than desired in naturalistic setting

Participant Selection

- Strict inclusion/exclusion criteria –v– Eligible; available participants
- Randomization addresses confounding –v– Unknown, unmeasured factors

Treatment / Program Implementation

- Complex and multifaceted treatments or programs are challenging to implement
- Differences in procedural experience of researchers/providers influences outcomes

Are there any viable alternatives to RCT?

Or, is any other approach to establishing causality just a dubious search for "fool's gold?"

- Models of Causality
- Definition of Effect
- Theory of Cause
- Causal Generalization
- Implications for Design and Analysis

Models of Causality

- Donald T. Campbell's Model (Campbell Causal Model, CCM)
- Donald B. Rubin's Model (Rubin Causal Model, RCM)
- Judea Pearl's Model (Pearl Causal Model, PCM)

Campbell's Model

- Validity Typology
- Threats to Validity
- Implications for Research Design

Campbell's Validity Typology

- Statistical Conclusion Validity
- Internal Validity
- Construct Validity
- External Validity

Threats to Validity

- Plausible rival hypotheses, or alternative explanations

  For example:

  One may infer that results from a nonrandomized experiment support a hypothesis that the treatment worked.

  It is possible to be wrong in many ways!

  - History
  - Maturation
  - Regression
  - Selection-treatment interaction

Definition of Effect

- **Knowledge of the counterfactual** (Campbell).
  - A 'counterfactual' is a condition that would occur if some part of the world were different than it really is.
  - The effect of a cause is the difference between what did happen to the person who received the cause (the fact) and what would have happened to that person had they not received the cause (the counterfactual).

$$ACE = E[\delta] = E[Y_1] - E[Y_0].$$

The *Average Causal Effect* (ACE) is a difference at the population level: it's the outcome when the treatment is received minus the outcome when the treatment is not administered.

Definition of Effect

- **Potential outcome** (Rubin)
  - Rubin notes that all potential outcomes could, in principle, be observed until treatment is assigned, that some can be observed after that.
  - By definition, counterfactuals can never be observed.

    It is not possible to observe a fact that that does not exist.
  - RCM developed a propensity score logic for observational studies grounded in what researchers know about regular (RCT) designs.

    See: Rubin (2004, 2005); Morgan & Winship (2007); Winship & Morgan (1999)

Theory of Cause

- Clear operationalization of the treatment
- Full implementation of the treatment
- Causes must be manipulable, and
- Cause includes whatever was manipulated

- CCM aspires to be a theory of generalized inference
- RCM has a more narrow purpose, to define an effect clearly and precisely in single experiments

Bandwidth *versus* Fidelity

Causal Generalization

- External validity and Theory of construct validity
- Meta-analysis plays a key role
- Neither RCM nor CCM has been particularly successful in generating applications of their ideas on causal generalization.

Priority is given to internal validity.

Implications for Design and Analysis

- Internal validity is the priority concern
- During design stage (Before data collection)

  Minimize the number of rival hypotheses and plausible alternative explanations
- After a study is completed

  Assess remaining threats to validity

This is what led to the development of some of the more complex and rigorous nonrandomized designs for use when RCTs are not possible or feasible.

Non-randomized designs

What to do when random assignment is not possible, or not ethically appropriate

- What are our options, then, among the various non-randomized design protocols?
- What are the critical considerations for research design protocol?
- What are the advantages and disadvantages for each approach?

Non-randomized designs

What to do when random assignment is not possible, or not ethically appropriate

- Quasi-Experimental
  - Nonequivalent control group
  - Regression Discontinuity Design
- Interrupted Time-Series Design with control series
- Complex Pattern Matching and Propensity Scores

Quasi-Experimental Designs

- A quasi-experimental design is one that looks a bit like an experimental design but lacks the key ingredient -- random assignment.
- **Nonequivalent Control Group** design requires a pretest and posttest for a treated and comparison group.
- **Regression discontinuity design** (RDD) is a quasi-experimental pretest-posttest design that elicits the causal effects of interventions by assigning a cutoff or threshold above or below which an intervention is assigned.

Interrupted Time-Series with a control series

Schema of a quasi-experiment using ITS-CG Design

Treatment Group	O_1	O_2	O_3	O_4	X	O_5	O_6	O_7	O_8
Comparison Group	O_1	O_2	O_3	O_4	-	O_5	O_6	O_7	O_8

Matching Procedures

- Metric approaches: Mahalanobis Distance Matching
- Nearest Neighbor: Propensity Score Matching
- Coarsened exact matching and caliper-based approaches

Strengths and Weaknesses Relative to RCT

- Strengths
  - Large and diverse samples in real-world settings
  - Opportunity for insight where multiple or complex phenomena occur simultaneously
  - Relatively inexpensive
- Weaknesses
  - Confounding or selection bias of participants
  - Difficult to compare nonequivalent groups
  - Multiple methodological approaches are often available but may be inconsistently applied or reported

Considerations

When planning or interpreting results of nonrandomized studies:

1. Does the study sample fit the hypothesis and the target population of interest?
2. Is the size of the sample adequate to answer the research question?
3. Is the study design conducive to addressing or answering the research question?
4. Is the treatment exposure determined accurately? Was exposure assessed before the outcome occurred? Can duration of exposure be quantified and treatment-response be explained? (sufficient exposure? e.g., large enough)
5. Is the outcome measured accurately and is it relevant for practice?
6. Are confounding factors measured accurately to control for confounding? Any potentially known confounding variables that are not measured?
7. Is there any cost or loss to follow-up?
8. Are the statistical methods and their assumptions suitable for the research question?
9. (if using secondary data) Does the data fit or address the research question?

So…,If RCT is not an option, then what?

- Representative sample, and
- *Matched* comparisons
- Strive for internal validity

Any Questions?