Comparative Effectiveness Research Methods Training

Module 2: Research Designs

J. Michael Oakes, PhD
Associate Professor
Division of Epidemiology
University of Minnesota
oakes007@umn.edu

A little (more) about me.

Image from cartoonbank.com removed.

Image description: Depicts cat looking at mouse in a maze. Mouse is looking up at cat with caption "Well, you don't look like an experimental psychologist to me."
Module #2 Outline

1. Review of Core Ideas
2. Experimental Designs
3. Observational Designs
4. Issues & Assumptions
5. Review
6. Questions
Module #2 Outline

1. Review of Core Ideas
2. Experimental Designs
3. Observational Designs
4. Issues & Assumptions
5. Review
6. Questions

---

On being asked to talk on the principles of research, my first thought was to arise… and say, “Be careful”, and to sit down.


---

1. Review of Core Ideas
Counterfactuals
& Potential Outcomes
Potential Outcomes

<table>
<thead>
<tr>
<th>Condition Assigned</th>
<th>Outcome if Treated, $Y^1$</th>
<th>Outcome if not Treated, $Y^0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Observed</td>
<td>Unobservable Counterfactual</td>
</tr>
<tr>
<td>Control</td>
<td>Unobservable Counterfactual</td>
<td>Observed</td>
</tr>
</tbody>
</table>

More formally, $T$ has a **causal effect** on $Y$ for person $i$ if

$$Y_i^{T=0} \neq Y_i^{T=1}$$

But we can only observe one of these outcomes for any $i$.

At the population level, we use probabilities and assuming exchangeability,

$$\text{Prob}[Y_{T=0} = 1] \neq \text{Prob}[Y_{T=1} = 1]$$

Causal inference is about finding the best **counterfactual substitutes** for the unobservable counterfactuals.
Effect Identification

An effect is identifiable if it is theoretically possible to learn the true value of the parameter when the sample size approaches infinity.

Imagining that your sample size is infinitely large eliminates all problems of statistical uncertainty, confidence intervals and p-values.

Identification is about ruling out competing hypotheses or explanations.

Apart from statistical imprecision, if more than one explanation for your effect (e.g., difference b/w treatment and control) exists, then you have an identification problem. You cannot say X₁ caused Y because the cause might have been X₂, or X₃, or Xₙ.
It's all always about data and assumptions

Got quality data?

Got plausible assumptions?

You can draw causal conclusions from a cross-sectional study with a non-random sample!

You just need heroic assumptions!
Designs vs. Models

Image from cartoonbank.com removed.

Image description: Depict two men in suits standing at the base of a large main frame computer. One man is gesturing at input feed to computer. Caption reads: “The machine then selects the likely equations from a complicated pattern of theoretical probables. It calculates these, and the correct answer is printed on a card. Then our Miss Swenson files them God knows where, and we can never find the damn things again.”

Most people learn some statistics, a few learn fancy statistical models, yet fewer still learn research design.

Research Design is more important than fancy statistical models.

Simple Design. Intense Content.
The quality and strength of evidence provided by a research study is determined largely by its design. Excellent methods of analysis will not salvage a poorly designed study... Aspects of design include the framing of the questions and assessment of measures, bias, and precision... An observational study that begins by examining an outcome variable is a formless, undisciplined investigation that lacks design... Design anticipates analysis.

Adapted from Rosenbaum 2010, p. ix

Three Types of Validity

• Internal Validity
• External Validity
• Statistical Conclusion Validity

Threats to Internal Validity

• Temporal Precedence
• Selection
• History (exogenous events)
• Maturation (endogenous events)
• Regression to mean
• Attrition
• Testing and instrumentation
Threats to External Validity

• Interactions by units and/or contexts

2. Experimental Designs

Image from despair.com removed.

Image description: Sign reads “Dear science: I beg you not to prove that any more of my pleasures will harm me.”

History of Experiments

• Emergence of scientific method, 16th century (esp. Bacon)
• Galileo’s telescope… and prosecution
• Hill’s randomized clinical trial for TB
• Income maintenance experiments
Key Epistemological Point

Experiments are used to test our knowledge of the world and correct it as necessary.

What is an Experiment?

1) Exogenous treatment/intervention
2) Randomly assigned to subjects

Find effects of (known) causes
Exogenous?

Random?

Random vs. Random

- Random sampling (external validity)
- Random assignment (internal validity)
Random Assignment

- Balances both measured and unmeasured confounders (ie, removes bias)

- “Works” in *expectation* only (ie, large numbers)

Manipulation of Treatments?

- Better to have it...
  
  “Twist the lion’s tale” to reveal his secrets

- Can learn w/out it, especially when phenomena don’t react to measurement (eg, planetary orbits)

Observational Study Differs how?

- Exogenous treatment/intervention

- NOT randomly assigned to subjects

  Find causes of (known) effects
Experiments vs. Obs Designs

Experimental Designs | Observational Designs
--- | ---
- Find effects of causes | - Find causes of effects
- Manipulation of Tx | - No manipulation of Tx
- Random assignment | - No random assignment
- Rule out competing explanations by design | - Rule out competing explanations by analysis

Caution!

If treatment is not exogenous but endogenous then estimation of treatment effects is virtually impossible.

(eg, social capital)

Research Design Diagrams

“O” represents an observation/measurement
“X” represents an intervention/treatment
“R” indicates randomization
“NR” indicates no randomization
Posttest only

R X O
R O

Counterfactual Substitutes?

Same outcome values but for the treatment?
Pretest-Posttest, with same pre-post subjects

Counterfactual Substitutes?

Pretest-Posttest, with same pre-post subjects

Counterfactual Substitutes?

Pretest-Posttest, with same pre-post subjects

Counterfactual Substitutes?
Double-Blind RCT

- Neither researchers nor subjects know if Tx or Cx is administered (placebos)
- Gold standard for FDA (pharm. studies)
- Not feasible for most "social" research

Non-Blind RCT

- Researchers and/or subjects know if Tx or Cx is administered
- Typical experiment in "social" research
- Sometimes subjects can be blinded while investigators are not

Cross-Over Experiments

- Subjects exposed to all conditions, flip-flopping or switching between them
- Requires "wash out" period
- Rarely feasible in "social" research
Cross-Over Experiment

\[
\begin{array}{cccc}
R & O & X_a & O \\
R & O & X_b & O \\
\end{array}
\]

Factorial Experiments

- Estimate independent effects of one or more treatments
- NOT factor analysis!
- Usually 2 levels per factor: \(2^n\) conditions
- 4 factors each of 2 levels = \(2^4 = 16\) conditions

Consider two treatments: A and B

<table>
<thead>
<tr>
<th>Factor B</th>
<th>Level 1</th>
<th>Level 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1+B1</td>
<td>A1+B2</td>
</tr>
<tr>
<td>Factor A</td>
<td>Level 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A2+B1</td>
<td>A2+B2</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td></td>
</tr>
</tbody>
</table>
Group-Randomized Trial

- Randomize in-tact social groups to Tx or Cx
- Use when Tx is for whole groups
- Use when changing "local" infrastructure
- Typically low-powered: $df = g-1$

Natural Experiments

- Mother Nature (eg, earth quake)
- Human Lotteries (eg, military draft)
- Policy change???

Regression Discontinuity Designs
RDD

- Offer treatment at prespecified cut-off of assignment variable
- Developed in the late 1950s
- Nearly as powerful as RCT

Figure 2. Regression discontinuity experiment with an effective treatment

Shadish et al. 2002

Figure 3. Randomized experiment with an ineffective treatment

Shadish et al. 2002
3. Observational Designs

Quasi-Experimental Designs

without Control Group

One-group Posttest

\[ X \quad O_1 \]

One-group Pretest-Posttest

\[ O_1 \quad X \quad O_2 \]

One-group Pretest-Posttest, with Double Pretest

\[ O_1 \quad O_2 \quad X \quad O_2 \]
Quasi-Experimental Designs

with

Control Group

Is Control Group

*Exchangeable*

with

Treatment Group?

Posttest only

\[
\begin{array}{c|c}
\text{NR} & \text{X} & \text{O}_1 \\
\hline
\text{NR} & \text{O}_2 \\
\end{array}
\]

Pretest-Posttest, with different pre-post subjects

\[
\begin{array}{c|c|c}
\text{NR} & \text{O}_1 & \text{X} & \text{O}_2 \\
\hline
\text{NR} & \text{O}_1 & \text{O}_2 \\
\end{array}
\]
Pretest-Posttest, with same pre-post subjects

\[ \begin{array}{c|c|c}
NR & O_1 & X \ O_2 \\
\hline
NR & O_1 & O_2 \\
\end{array} \]

Pretest-Posttest, same pre-post subjects and 2 posttests

\[ \begin{array}{c|c|c|c}
NR & O_1 & X & O_3 \\
\hline
NR & O_1 & O_2 & O_3 \\
\end{array} \]

Policy Change

\[ \Delta = 6 \]

\[ \Delta = 0 \]
Pre-Post, same subjects, 2 pretests and 2 posttests

NR O₁ O₂ X O₃ O₄
NR O₁ O₂ X O₃

Pre-Post, same subjects, 2 pretests, 2 posttests, switching

NR O₁ X O₂ O₃
NR O₁ O₂ X O₄

Pre-Post, same subjects, 2 pretests, 2 posttests, switching

NR O₁ X O₂ O₃
NR O₁ O₂ X O₄
Interrupted Time Series Designs

Simple ITS

\[ O_1, O_2, O_3, O_4, O_5, O_6, O_7, O_8, O_9 \]

\[ \uparrow \]

**FIGURE 6.1** The effects of charging for directory assistance in Cincinnati
4. Issues & Assumptions

Experiments...

- Usually tell us that some cause (known to us) yields an effect
- Rarely tells us how the effect came about (no insight into mechanism)
- A full explanation requires deep understanding of the mechanisms
Assumptions in Experiments

- Full compliance
- No attrition
- External validity?

...in Observational Studies

Internal validity?

- Exchangeability
- Positivity
- Consistency

Oh, what shall I do?

Image from despair.com removed.

Image description: Picture shows a scenic mountain sun rise in the background with river in the foreground. Caption of picture reads “It's always darkest just before it goes pitch black.”
1) Seek causal effects

2) Always think experiment

3) If you cannot randomize, imagine the experiment you wish you could conduct

4) Conduct analyses that mimic the desired experiment and rule out competing explanations

5. Review

I. Background Concepts
   a. Experiments correct knowledge
   b. Counterfactuals
   c. Effect identification
   d. Data and assumptions
   e. Design trumps analysis

II. Experiments
   a. Randomization
   b. Double-blind RCT
   c. Factorial Experiments
   d. Group-randomized trial
   e. Natural experiments
   f. RCT

III. Observational Designs
   a. Exchangeability
   b. Positivity
   c. No control group
   d. Control group(s)
   e. Interrupted Time Series

IV. Assumptions & Issues
   a. Black-box mechanisms
   b. Experimental Fidelity
   c. Exchangeability
6. Questions

1. What are the two elements that make up an experiment?
2. What is a factorial experiment?
3. Why is selection such a problem for observational designs?
4. What does exogenous mean?
5. When should one conduct a group-randomized trial?

References & Resources

4. Shadish, WR, Cook, TD, Campbell, DT. Experimental & Quasi – Experimental Designs for