Experimental & Quasi-Experimental Designs (baseline equivalence, confounding factors, attrition)

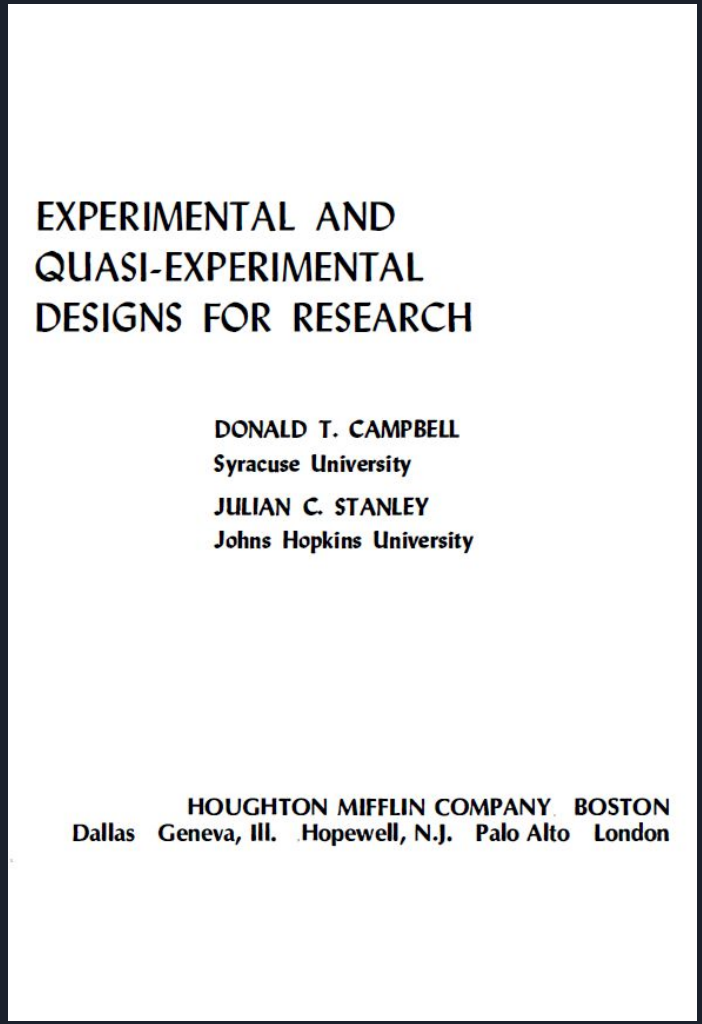
מתודולוגיה במחקר יישומי, לשכת המדען
מפגש שני, 17.1.21

פרופ׳ תום גומפל
ד״ר הדר בהרב

Principles for Inferring Causation

- Research-to-Policy considerations (decision-driven or knowledge-driven) and Research-to-Practice must drive all design decisions
- Comparison group
  - Random assignment / Baseline equivalence
- Time sequencing (causality)
- Manipulation
  - Control
  - Equivalent conditions
- Dealing with
  - Threats to internal and external validity

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.


EXPERIMENTAL AND
QUASI-EXPERIMENTAL
DESIGNS FOR RESEARCH

DONALD T. CAMPBELL
Syracuse University

JULIAN C. STANLEY
Johns Hopkins University

HOUGHTON MIFFLIN COMPANY BOSTON
Dallas Geneva, Ill. Hopewell, N.J. Palo Alto London

Research Designs

- Experimental
  - 3 pre-experimental designs
  - 3 true experimental designs
  - Solomon Four-Group design
  - Posttest Only Control Group design
  - Factorial design
- Basic characteristics
  - Random assignment
  - pre-intervention measures to establish baseline equivalence

Research Designs (cont.)

See: Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

- Quasi-experimental
  - Time-series experiment design
  - Equivalent Time-Samples design
  - Equivalent Materials design
  - Nonequivalent Control Group design
  - Counterbalanced design
  - Separate-Sample Pretest-Posttest design
  - Multiple Time-Series design
  - Recurrent Institutional Cycle design
  - Regression-Discontinuity design
- Basic characteristics
  - Sequential measures (pre-post) over time
  - pre-intervention measures to establish baseline equivalence

True Experimental Designs

TABLE 1

SOURCES OF INVALIDITY FOR DESIGNS 1 THROUGH 6

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Pre-Experimental Designs:												
1. One-Shot Case Study X O	−	−				−	−			−		
2. One-Group Pretest-Posttest Design O X O	−	−	−	−	?	+	+	−	−	−	?	
3. Static-Group Comparison X O --- O	+	?	+	+	+	−	−	−		−		
True Experimental Designs:												
4. Pretest-Posttest Control Group Design R O X O R O O	+	+	+	+	+	+	+	+	−	?	?	
5. Solomon Four-Group Design R O X O R O O R X O R O	+	+	+	+	+	+	+	+	+	?	?	
6. Posttest-Only Control Group Design R X O R O	+	+	+	+	+	+	+	+	+	?	?	

Note: In the tables, a minus indicates a definite weakness, a plus indicates that the factor is controlled, a question mark indicates a possible source of concern, and a blank indicates that the factor is not relevant.

It is with extreme reluctance that these summary tables are presented because they are apt to be "too helpful," and to be depended upon in place of the more complex and qualified presentation in the text. No + or − indicator should be respected unless the reader comprehends why it is placed there. In particular, it is against the spirit of this presentation to create uncomprehended fears of, or confidence in, specific designs.

Quasi-Experimental Designs

TABLE 2

SOURCES OF INVALIDITY FOR QUASI-EXPERIMENTAL DESIGNS 7 THROUGH 12

	Sources of Invalidity: Internal: History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	External: Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Quasi-Experimental Designs:												
7. Time Series O O O OXO O O O	–	+	+	?	+	+	+	+	–	?	?	
8. Equivalent Time Samples Design X_1O X_2O X_1O X_2O, etc.	+	+	+	+	+	+	+	+	–	?	–	–
9. Equivalent Materials Samples Design M_aX_1O M_bX_2O M_aX_1O M_dX_2O, etc.	+	+	+	+	+	+	+	+	–	?	?	–
10. Nonequivalent Control Group Design O X O O O	+	+	+	+	?	+	+	–	–	?	?	
11. Counterbalanced Designs X_1O X_2O X_1O X_2O X_2O X_1O X_2O X_1O X_1O X_2O X_1O X_2O X_2O X_1O X_2O X_1O	+	+	+	+	+	+	+	?	?	?	?	–
12. Separate-Sample Pretest-Posttest Design R O (X) R X O	–	–	+	?	+	+	–	–	+	+	+	
12a. R O (X) R X O R O (X) R X O	+	–	+	?	+	+	–	+	+	+	+	
12b. R O_1 (X) R O_2 (X) R X O_3	–	+	+	?	+	+	–	?	+	+	+	
12c. R O_1 X O_2 R X O_3	–	–	+	?	+	+	+	–	+	+	+	

Perhaps best also included under *history*, although in some sense akin to *maturation*, would be periodical shifts in the time series related to institutional customs of the group such as the weekly work-cycles, pay-period cycles, examination periods, vacations, and student festivals. The observational series should be arranged so as to hold known cycles constant, or else be long enough to include several such cycles in their entirety.

TABLE 3

SOURCES OF INVALIDITY FOR QUASI-EXPERIMENTAL DESIGNS 13 THROUGH 16

	Sources of Invalidity: Internal: History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	External: Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Quasi-Experimental Designs Continued:												
13. Separate-Sample Pretest-Posttest Control Group Design R O (X) R X O R O R O	+	+	+	+	+	+	+	–	+	+	+	
13a. R O (X) R X O R O (X) R X O R O R O R O R O	+	+	+	+	+	+	+	+	+	+	+	
14. Multiple Time-Series O O OXO O O O O O O O O	+	+	+	+	+	+	+	+	–	–	?	
15. Institutional Cycle Design Class A X O_1 Class B_1 RO_2 X O_3 Class B_2 R X O_4 Class C O_5 X *Gen. Pop. Con. Cl. B O_6 *Gen. Pop. Con. Cl. C O_7												
$O_2 < O_1$ $O_5 < O_4$	+	–	+	+	?	–	?		+	?	+	
$O_2 < O_3$	–	–	–	?	?	+	+		–	?	+	
$O_2 < O_4$	–	–	+	?	?	+	?		+	?	?	
$O_6 = O_7$ $O_{2y} = O_{2z}$		+						–				
16. Regression Discontinuity	+	+	+	?	+	+	?	+	+	–	+	+

*General Population Controls for Class B, etc.

Baseline Equivalence (BE)

- Baseline equivalence (BE) is used in determining if the intervention group and the comparison group had characteristics that were similar enough at the start of the study
- If the two groups are different at baseline on key characteristics that could influence the outcomes, the effect found at the end of the study might be due to the differences that already existed at the beginning
- BE is important
  - in studies that do not randomly assign participants to groups
  - in random assignment studies with high attrition
- BE can only be established on observable characteristics

Determining baseline equivalence - What to measure?

- Characteristics to be tested on equivalence at baseline are determined by the outcomes, e.g.
  - for academic outcomes, BE is established using a pre-intervention test
  - For HS graduation outcomes, BE is examined on related demographic characteristics (e.g., age)

Determining baseline equivalence - How to calculate?

- Calculating an effect size

$$\text{Effect size} = \frac{\text{Difference in means}}{\text{Pooled standard deviation}}$$

- Determining BE

Figure 3: WWC standard for baseline equivalence

Absolute value of effect size ≤ 0.05	0.05 < absolute value of effect size ≤ 0.25	Absolute value of effect size > 0.25
Satisfies baseline equivalence	Statistical adjust-ment required to satisfy baseline equivalence	Does not satisfy baseline equivalence

Confounding Factors

- A confounding factor is an aspect of a study that makes it impossible to tell whether the intervention alone is responsible for the outcomes
  - The confounding factor may be present in the intervention or comparison group

Common types of confounding factors in education studies

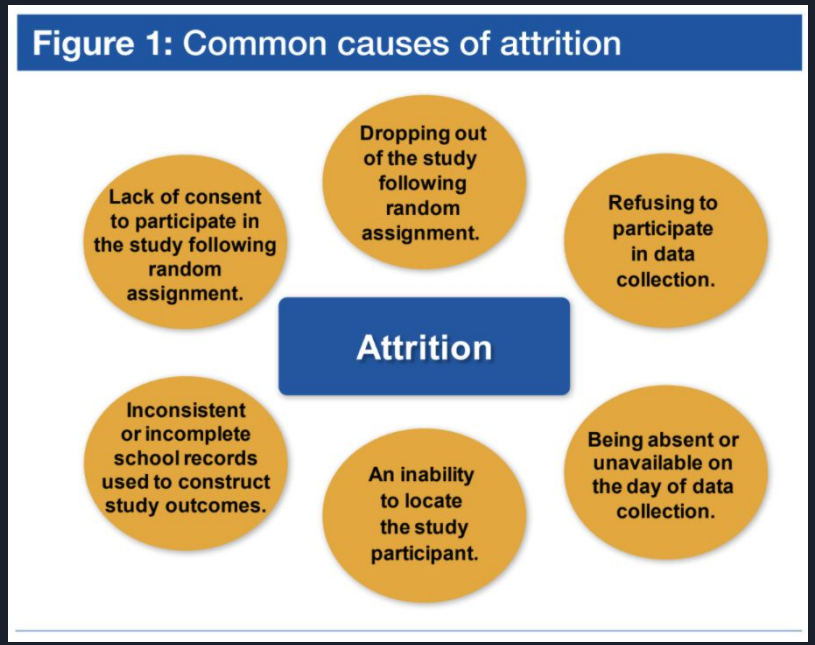
Example	Intervention	Potential confounding factor
In a study of a new reading curriculum, three new teachers volunteer to try the new program, while three teachers with 20 or more years of experience stick with the curriculum they've used for years.	New reading curriculum	Teacher experience (intervention group)
Intervention group students receiving a new math curriculum also receive additional tutoring that is not part of the curriculum. Comparison group students use the standard curriculum and receive no tutoring.	New math curriculum	Tutoring (intervention group)
All of the students in Mrs. Jones's and Mr. Wright's classes use a new software package to work on spelling (the intervention group), while Mrs. Smith's students continue to work only with pencils and paper (the comparison group).	New software package	Single teacher (comparison group)

Confounding Factors, Special Cases (and WWC's approach)

- A single-unit design
  - for example, a study may be interested in analyzing the effect of attending a specific charter school
  - these studies are still eligible to Meet WWC Design Standards and are considered on a case-by-case basis
- A quasi-experimental design (e.g., the intervention group includes those who volunteered, while the comparison group includes those who did not volunteer)
  - unmeasured differences between groups are often present in quasi-experimental design studies.
  - The WWC accounts for these issues by not allowing such design studies to receive the highest rating

Attrition

- Attrition is the loss of sample during the course of a study
- It occurs for many reasons,

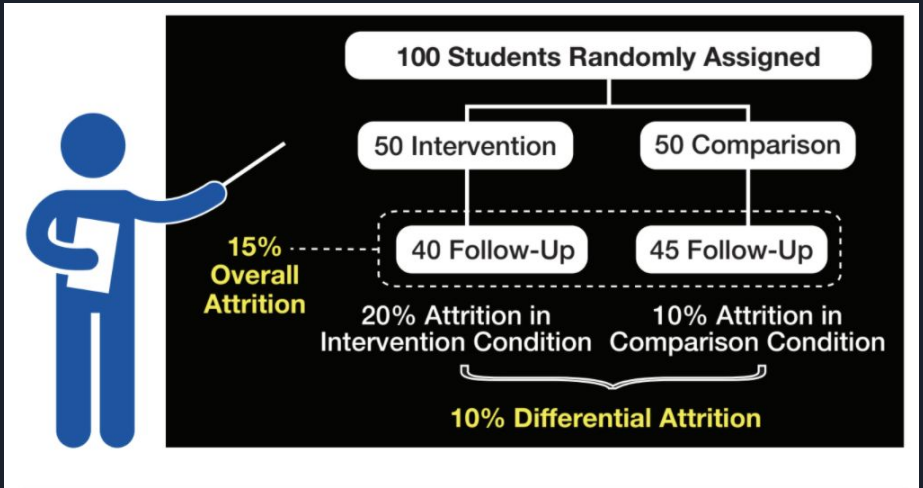


Impact of Attrition

- If attrition occurs, the members of the intervention and comparison groups may not have had similar characteristics at baseline, preventing us from being able to attribute any differences in outcomes solely to the intervention
- Then, any observed effect of the intervention may be biased
- A study with low attrition is expected to have low levels of bias

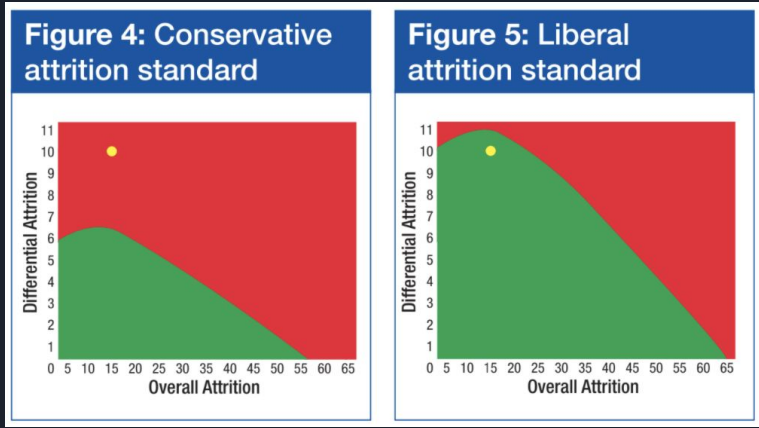
Calculating Attrition

- Distinguishing between two kinds of attrition:
  - Overall - attrition for all study participants
  - Differential - differences in attrition between the groups



Determining potential bias due to attrition

- WWC uses two attrition standards :
  - coservative - used when attrition is likely to be related to the intervention
  - liberal - used when an intervention is unlikely to affect attrition



References

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

What Works Clearinghouse, Standards Briefs. https://ies.ed.gov/ncee/wwc/StandardsBriefs