- Principled, Formal Analysis that insures detailed attention to key design features
- Extension of Potential Outcomes Analysis (Rubin 1978; Imbens & Rubin, 2015)
- Careful Attention to Assumptions
- Focus on Ideal, Asymptotic Case
  - Statistical Power not an issue
  - Researcher Degrees of Freedom, Garden of Forking Paths Not Issues

- Only Focus: Design Features

- Shifts focus from Statistical Conclusion Validity to Internal Validity
    - Clear General Guidelines for Replication of Experiments, Quasi-Experiments
    - Less Clear:
        - Extension to Studies of Association (Relationships)
            - Trivial Issue: No Manipulation
            - Complex Issue: How to Address Sampling Issues since standardized effect sizes highly dependent on distributions of predictors
        - How do we judge whether features are the same?

Example: Two Perfectly Implemented RCTs

Assumption	Original Study: RCT	Replication I: RCT
A1 Treatment & outcome stability	✓ High fidelity of treatment and control conditions ✓ Outcome measure, instruments & timing ✓ No mode-of-study-selection effects ✓ No peer-, spillover-, or carry-over effects	✓ High fidelity of treatment and control conditions ✓ Outcome measure, instruments & timing ✓ No mode-of-study-selection effects ✓ No peer-, spillover-, or carry-over effects
A2 Equivalence of causal estimands	✓ ATE ✓ effect-generating process ✓ target population $P = Q$ ✓ setting $S_0 = S_1$	✓ ATE ✓ effect-generating process ✓ target population $Q = P$ ✓ setting $S_1 = S_0$
A3 Identification	✓ ATE is identified	✓ ATE is identified
A4 Estimation	✓ Unbiased (mean difference)	✓ Unbiased (mean difference)
A5 Reporting	✓ Correct reporting	✓ Correct reporting

Example: Hull and West (1982).

- Study of “Discounting Principle” in Attribution
- Two Choices: 1 to 3 reasons for each choice

  e.g., stay at university vs. go home for Summer

  Perceived importance of each reason

  Reported Original Experiment and Conceptual Replication showed Effect

  Another experiment, quasi-experiment showed effect (not reported)

  All research conducted at Florida State University (non selective)

Simultaneous Replication Attempts by Hull

- Duke University—Highly Selective Students, More experimental experience
- Experiment 1: Between Subjects, but small N (low power)
- Experiment 2: Within Subjects
- Different Setting: Gothic architecture, individual (lab)
- Different Experimenter

Got complex interaction conditions x latin square order in within subjects design This effect was in same direction, but not statistically significant at FSU. (likely spillover effect. Interesting??? Worth larger study to attempt to show significant? possible smaller effect size at FSU.)

In theory: Two Imperfectly Implemented RCTs
In practice: Implemented about as perfectly as possible.

Assumption	Original Study: FSU	Replication I: Duke
A1 Treatment & outcome stability	✓ High fidelity of treatment and control conditions ✓ Outcome measure, instruments & timing ✓No mode-of-study-selection effects ✓ No peer-, spillover-, or carry-over effects	✓ High fidelity of treatment and control conditions ✓ Outcome measure, instruments & timing ✓ No mode-of-study-selection effects ✗ No peer-, spillover-, or carry-over effects (**within subject**)
A2 Equivalence of causal estimands	✓ ATE ✓ effect-generating process ✓ target population $P = Q$ ✓ setting $S_0 = S_1$	✓ ATE ✓ effect-generating process ✗target population $Q = P$ ✗ setting $S_1 = S_0$
A3 Identification	✓ ATE is identified	✓ ATE is identified
A4 Estimation	✓ Unbiased (mean difference)	✓ Unbiased (mean difference)
A5 Reporting	✓ Correct reporting	✓ Correct reporting

Additional Insights to Supplement: Checklists by Brandt et al., CONSORT (randomized experiments); TREND (quasi-experiments)

Designing the Replication Study

10.Are the original materials for the study available from the author?

a.If not, are the original materials for the study available elsewhere (e.g., previously published scales)?

b.If the original materials are not available from the author or elsewhere, how were the materials created for the replication attempt?

11.I know that assumptions (e.g., about the meaning of the stimuli) in the original study will also hold in my replication because:

12.Location of the experimenter during data collection:

13.Experimenter knowledge of participant experimental condition:

14.Experimenter knowledge of overall hypotheses:

15.My target sample size is:

16.The rationale for my sample size is:

Excerpt from Brandt et al.

Extending Potential Outcomes Analysis to Investigate UTOS

The traditional potential outcomes model is largely mute with regard to construct validity, external validity.

Steiner has extended the rationale of the potential outcomes analysis and provides a principled method of investigating questions. Note. Ideas already a part of researcher's arsenal of approaches, but not always done in principled manner.

Conceptual Replications, "Critical Experiments"

(Steiner highlights: These are primarily informative about direction, not magnitude of effect)

EX 1. Thallium Stress Test

EX 2. Hull vs. Tolman

Cognitive Dissonance vs. Reinforcement—Linder, Cooper, & Jones