CTE | Career & Technical Education RESEARCH NETWORK

# Randomized Controlled Trials

**August 18, 2020**


Dr. Trey Miller

Principal Researcher



The work of the CTE Research Network Lead is supported by the Institute of Education Sciences at the U.S. Department of Education with funds provided under the *Carl D. Perkins Career and Technical Education Act* through Grant R305N180005 to the American Institutes for Research (AIR). The work of the Network member projects is supported by the Institute. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education

# Introduction

- The purpose is to introduce randomized controlled trials (RCTs) and some practical considerations when using them.
- The RCT is considered the gold standard of impact evaluation.
  - The advantage of an RCT study design is that it allows us to estimate the counterfactual based on the outcomes of nonparticipants.
  - Comparing the outcomes of program participants and nonparticipants shows the impact of the program.

Source: Placeholder for sources and permissions (if needed).

# Basics of RCT

- In a basic RCT, units of analysis/observation are assigned to the treatment group or the control group (the experimental conditions).
- The assignment to experimental conditions must be solely by chance.
- RCTs can have different designs:
  - The design could have more than two experimental conditions.
  - The probability of assignment does not have to be 50/50.
- The **analytic sample** is the set of observations used to estimate the impact of an intervention being studied and should only include those randomly assigned to the experimental conditions.

# Conditions Amendable to RCT

- RCTs can be challenging to implement.
  - Program administrators can be reluctant to deny services to clients or resist using a lottery.
  - Randomization requires close collaboration to be done correctly.
- RCTs are best suited for situations with excess demand.
  - Filling limited program slots by lottery seems fair.
  - Administrators may be more open to RCT if the probability of assignment to treatment is > 50%.

# Levels of Randomization

- Individual-level versus cluster
  - **Individual-level** random assignment involves randomization at the level of program participants.
  - **Cluster** random assignment involves randomization at the level of groups of individuals, called clusters (e.g., schools).
- Sometimes individual-level randomization is not feasible.
  - But the outcome(s) of interest may still be individual level.
  - In a cluster RCT, you need to account for the clustering at the analysis stage.

# Randomization and Blocking Strategies

- One simple approach would be to randomize each participant as they enter the study.
  - There is a risk of unbalanced group assignments simply by chance.
- Using a **blocking** strategy (also called **stratified random assignment**) can help ensure that the groups are balanced.
  - The sample is first split into groups by one or more characteristics.
  - Units are randomly assigned within each group.
- If you use blocking, you need to account for the blocks when estimating treatment effects.

# Statistical Power

- When designing an RCT, you want to make sure that the study has a good chance of detecting an impact if one exists.
- Beforehand, you can identify the sample size necessary to give the design sufficient **power**.
- Generally, larger sample sizes provide greater power.
  - Example: Outcome is graduation rate, control group rate is 85%.
    - With a sample size of 100, you can only detect an impact as small as 15 pp.
    - With a sample size of 1,000, you can detect an impact as small as 6 pp.
- Most statistical computing packages have commands for doing power analyses (`power` in Stata and `pwr` in R).

# Assessing Baseline Equivalence

- If random assignment was executed well, the baseline characteristics of the treatment and control groups should be balanced.
  - The similarity between treatment and control groups prior to the start of the intervention is called **baseline equivalence.**
- To confirm baseline equivalence, you test for statistically significant differences in baseline characteristics between the two groups.
  - Data often come from a baseline survey or administrative data.
- There should be no (or few) differences that are statistically significant.

# Potential Threats: Attrition and Missing Data

- Two other issues can arise when implementing an RCT: attrition and missing data.
- **Attrition** is when outcome data are unavailable for some sample members – it can bias the estimated treatment effects.
  - Two types of attrition matter: (1) **overall attrition** and (2) **differential attrition** between the treatment and comparison group.
- **Missing data** arise when some data are missing for study participants.
  - There are various ways to address this issue, such as by dropping observations with incomplete data or imputing the missing data.

# Estimating Treatment Effects: Differences in Means

- If the RCT was implemented correctly, the simplest estimate of the treatment effect is the **difference in means** of the outcome variable between the treatment and comparison groups:

$$\text{Treatment effect} = \overline{Y_T} - \overline{Y_C}$$

- Formal statistical hypothesis testing procedures are used to judge statistical significance.
- Hypothesis testing is easily done in major statistical computing packages.

# Estimating Treatment Effects: Regression Adjustment

- Often researchers will supplement the comparison of means with **regression-adjusted** estimates.
  - This approach increases precision by controlling for variation in the outcome that is correlated with observable baseline characteristics.
- To do this, you estimate a regression model:

$$Y_i = \alpha + \beta X_i + \delta T_i + \varepsilon_i$$

  - $X_i$ represents the set of baseline characteristics, and $T_i$ is an indicator variable for assignment to the treatment.
  - $\delta$ represents the treatment effect.
- You then test whether the estimated treatment effect, $\hat{\delta}$, is statistically significant.

# Estimation With Noncompliance

- In practice, some treatment group members may not enroll in the program or vice versa, leading to **noncompliance**.
- You can ignore noncompliance and estimate the **intent-to-treat** (ITT) effect.
- Or you can estimate the average treatment effect on individuals who would comply with their treatment assignment, called the **complier average causal effect** (**CACE**).

$$\widehat{\text{CACE}} = \widehat{\text{ITT}} / (\overline{D_T} - \overline{D_C})$$

  - $D_T$ is an indicator variable for being assigned to and receiving treatment, and $D_C$ is an indicator variable for being assigned to control and receiving treatment.

# Estimation With Noncompliance

- You can also estimate the CACE using an instrument variables approach:

$$D_i = \alpha_1 + \beta T_i + \epsilon_i$$
$$Y_i = \alpha_2 + \delta D_i + \mu_i$$

  - $D_i$ is an indicator variable for receiving the treatment.
  - The estimate $\hat{\delta}$ is the estimate of the CACE.

# WWC Standards for RCTs

- RCTs are the gold standard for estimating program impacts but must be executed well to produce reliable results.
- The WWC has two criteria for well-executed randomization: (1) assignment of units entirely by chance and (2) every unit must have a chance to be assigned to each group.
- WWC reviewers consider two types of attrition thresholds used by WWC reviewers:
- For individual-level RCTs, attrition must be sufficiently low to be eligible for the highest WWC standard rating. There are two types of attrition thresholds:
  - The **optimistic attrition threshold** applies to interventions that are unlikely to affect attrition.
  - The **cautious attrition threshold** applies to interventions that are likely to affect attrition.

# WWC Standards for RCTs

- To be eligible to receive a rating of **Meets WWC Group Design Standards Without Reservations,** a cluster-level RCT study must have low cluster-level attrition, limit the risk of bias because of joiners, and have low individual-level nonresponse.
- RCT studies that do not meet the requirements for the highest rating can be eligible to receive a rating of **Meets WWC Group Design Standards With Reservations** under certain conditions:
  - An individual-level RCT with high attrition must demonstrate baseline equivalence (BLE).
  - A cluster-level RCT must establish BLE if any of the requirements noted above are not met.
  - There are some other requirements for cluster-level RCTs.
- Studies that do not satisfy the more lenient standards will receive a rating of **Does Not Meet WWC Group Design Standards.**

CTEResearchNetwork.org

CTEResearchNetwork@air.org