

# Technical Assistance Webinar Designing Community-Level Evaluations Tier 1B Grantees & Evaluators Abt Associates

February 4, 2016



# Introductions



Kim Francis, PhD  
Abt Associates  
Project Director



Randall Juras, PhD  
Abt Associates  
TA Lead



Amy Farb, PhD  
OAH  
Evaluation Specialist

- Introductions and Abt Associates' Role
- Grantee-Level Evaluation Requirements
- Selecting Comparison Groups: Key Considerations
- Selecting Comparison Groups in Practice
- Choosing an Appropriate Design
- Q and A

At the conclusion of this webinar,  
participants should be able to:

- Understand evaluation reporting requirements.
- Know the characteristics of appropriate community-level comparison groups.
- Understand community-level evaluation design options.
- Identify potentially problematic designs.

OAH awarded Abt a contract in September 2015 to:

- Document & describe scale-up strategies grantees are using in their communities.
- Design a cross-site evaluation to determine effectiveness of OAH's overall grant strategy.
- Provide evaluation technical assistance for grantee-led evaluations.
  - Periodic group TA opportunities.
  - Individualized TA for grantees invited to participate in evaluation.

## 1. Performance Measures

- a) Collect all performance measures & report to OAH semi-annually
- b) OMB # 0990-0438--expiration date 10/31/2018

## 2. Implementation Evaluation

- a) Document the process of developing and implementing the Tier 1B project- identify key successes, challenges, and lessons learned.
- b) Implementation Study Report to OAH

## 3. Impact Evaluation

- a) Determine the extent to which the outcome goal(s) was met by the end of the grant period and due to the Tier 1B project
- b) Impact Evaluation Report to OAH

Note: No more than 10% of annual grant award should be spent on evaluation activities (FOA, p. 31)



# **SELECTING A COMPARISON GROUP: KEY CONSIDERATIONS**

- In an impact evaluation, the idea is to estimate how the intervention affected an outcome.
- You know what the outcome was with the intervention (because you implemented it).
- You have to guess what the outcome would have been without the intervention.
  - This is called the “counterfactual”



- The goal is to find a comparison group that looks just like the counterfactual.
  - This can be accomplished through **design** and/or **analysis**
- Design: Find a comparison group that looks as similar to the treatment group as possible.
- Analysis: Use statistical methods to correct for differences between the two groups.
- Because statistical methods can only control for observable differences, **design trumps analysis.**

- Selecting a comparison group begins with the specification of research question(s).
- A **well-specified research question** should include the following four elements:
  1. The **intervention name** (i.e., not just the EBPs but the community-wide effort);
  2. The **target population** (e.g. teens ages 14-19 in the community)
  3. The **counterfactual condition** (e.g. business as usual);
  4. The **outcome domain** (e.g. pregnancy)

“Does the Abt Community Project affect the birth rate for girls ages 14-19 in Cambridge, MA, compared with girls in similar communities without community-wide TPP programs?”

“Does the **Abt Community Project** affect the birth rate for girls ages 14-19 in Cambridge, MA, compared with girls in similar communities without community-wide TPP programs?”

### 1. Intervention name

“Does the **Abt Community Project** affect the birth rate for **girls ages 14-19 in Cambridge, MA**, compared with girls in similar communities without community-wide TPP programs?”

1. **Intervention name**
2. **Target Population**

“Does the **Abt Community Project** affect the birth rate for **girls ages 14-19 in Cambridge, MA**, compared with girls in similar communities without community-wide TPP programs?”

1. **Intervention name**
2. **Target Population**
3. **Counterfactual condition**

“Does the **Abt Community Project** affect the **birth rate** for **girls ages 14-19 in Cambridge, MA**, compared with girls in similar communities without community-wide TPP programs?”

1. **Intervention name**
2. **Target Population**
3. **Counterfactual condition**
4. **Outcome Domain**

**“Does the Abt Community Project affect the birth rate for girls ages 14-19 in Cambridge, MA, compared with girls ages 14-19 in similar communities without community-wide TPP programs?”**

*vs.*

**“Does the Abt Community Project affect the birth rate for girls ages 14-19 who participate in the intervention in Cambridge, MA, compared with girls in similar communities without community-wide TPP programs who would have participated if given the chance?”**



- A good comparison group is **local** and **focal**.
  - Local = Close to same locale as treatment units
  - Focal = Similar characteristics as treatment group
- Outcomes should be measured at the **same level of aggregation** across the two groups.
- Ideally, outcomes should also be measured **in the same way** across the two groups.

- Research Question: *“Does applying a new kind of fertilizer to apple trees improve the seed count in apples compared with using no fertilizer?”*
  - Intervention: new fertilizer
  - Target population: apple trees
  - Counterfactual: no fertilizer use
  - Outcome domain: seed count in apples
- How could you answer this question?

## Option 1: Randomized Controlled Trial

Select several McIntosh apple trees planted in the same soil, near to each other

Randomly select some to treat with the new fertilizer and leave the others untreated

Calculate the average number of seeds per apple in each group of trees and compare

- This is a convincing comparison.
- On average, the comparison group should be just like the treatment group except for the fertilizer (i.e., local and focal).
- Note that you need more than  $n=2$  observations: What if one tree were hit by lightning?

## Option 2: Quasi-Experimental Design

Treat Orchard 1 (mostly McIntosh trees) with the new fertilizer

Leave nearby Orchard 2 (mostly Golden Delicious trees) untreated

Calculate the average number of seeds per apple in each orchard and compare

- Less convincing: local (neighboring orchard) but not focal (different mix of trees).
- Partial analytic solution: Adjust for difference in mix of McIntosh and Golden Delicious trees between orchards.

## Option 3: Benchmark to National Average

Treat a sample of McIntosh apple trees with the new fertilizer

Measure the average number of seeds per apple in this sample

Find the national average seed count per apple for apple trees using national data

Compare across groups

- Less convincing: not focal (national avg. includes other trees) and not local (different growing conditions).
- More difficult to adjust: Perhaps measure the difference in seed count between your treated trees and the national average before you apply the treatment and then again after, and see if your trees improve *relative to* the others.
- This assumes changes over time would be the same in both types of apple if untreated.

## Option 4: QED with Inappropriate Comparison

Treat McIntosh trees with the new fertilizer

Treat nearby McIntosh trees with an existing fertilizer

Calculate the average number of seeds per apple in each group of trees and compare

- This comparison group is local and focal.
- But it answers a different research question! It cannot tell you the impact compared with *no fertilizer*.
- If this is your only option, may need to revise research question.

## Option 5: Apples-to-Oranges Comparison

Treat McIntosh trees with the new fertilizer

Leave faraway Orange trees untreated

Calculate the average number of seeds per fruit in each group of trees and compare

- The comparison group is neither local nor focal.
- No amount of statistical adjustment will solve this problem: Cannot assume that changes over time would be the same in apple and orange trees, since they benefit from different climates.

**Putting more effort into the design (i.e., selecting a good comparison group) makes the analysis easier *and* more convincing.**



- Notice that in the previous examples, the unit of comparison was the *tree*, not the *apple*.
  - We selected comparison trees, and then averaged across apples in each group of trees to measure the outcome.
  - Comparing “similar” apples (e.g., apples from the bottom branches of each tree) is not a convincing comparison if the trees are different types.
- This is analogous to community-level TPP interventions – instead of comparing similar people, compare people *in similar communities*.

- The key challenge in selecting a comparison group is to remove as many confounding factors as possible.
- A **confounding factor** or **confound** is a component completely aligned with only one study condition.
  - Impossible to separate effect of intervention from effect of confound.
  - Thus, cannot attribute impact solely to intervention.

- Common confounding factors include:
  - Different data used to measure outcomes across groups (e.g. surveys for treatment group vs. admin data for comparison group).
  - Groups in dissimilar geographic locations (e.g., treatment group in Northeast and comparison in West).
  - Different demographic characteristics across populations (e.g. low income Hispanics in treatment group and high income Hispanics in comparison group).

- To avoid confounds, comparison groups should be:
  - ✓ **Local**
  - ✓ **Focal**
  - ✓ **Aggregated at the Same Level**
  - ✓ **Measured Using the Same Data Source**
- Once you have found a comparison group that meets these criteria, you should check for baseline equivalence.

- The two groups should be as similar as possible prior to the intervention.
  - Any pre-existing differences could carry through to the outcomes.
- Demonstrate that there were no differences in key characteristics at baseline.
  - Especially the baseline measure of the outcome.
  - Your communities probably have high teen birth rates – that’s why you chose them. So the comparison group should have high teen birth rates too!
  - Baseline differences are usually expressed as effect sizes.
  - If the effect size difference is too big, you might reject the comparison.

- You can only answer a research question if you can measure the outcome.
  - Preferably from both *before* and *after* the intervention.
  - Using extant data (FOIA states not to survey youth).
- You will be most interested in data aggregated at the **community level**.
  - Not at the individual level; e.g., student-level records from schools.
- We expect to have a future TA webinar on data sources.



# **SELECTING A COMPARISON GROUP IN PRACTICE**

- In a small sample and with few observable characteristics, it would be possible to select a comparison group by hand.
  - E.g., in a state-level intervention there are only 50 states to choose from.
- In such a small sample, it is easy to keep track of the key factors by hand.
  - Local and focal.



## Treatment Observation

State	Region	Outcome Rate	Poverty Rate
A	NW	5%	20%

What is the best comparison for State A?

## Comparison Group Candidates

State	Region	Outcome Rate	Poverty Rate
Z	NE	8%	10%
Y	SW	6%	5%
X	NW	5%	20%
W	SE	5%	20%
V	NW	20%	5%
U	NE	7%	10%

## Treatment Observation

State	Region	Outcome Rate	Poverty Rate
A	NW	5%	20%

What is the best comparison for State A?

✓ Local

## Comparison Group Candidates

State	Region	Outcome Rate	Poverty Rate
Z	NE	8%	10%
Y	SW	6%	5%
X	NW	5%	20%
W	SE	5%	20%
V	NW	20%	5%
U	NE	7%	10%

## Treatment Observation

State	Region	Outcome Rate	Poverty Rate
A	NW	5%	20%

What is the best comparison for State A?

- ✓ Local
- ✓ Focal

## Comparison Group Candidates

State	Region	Outcome Rate	Poverty Rate
Z	NE	8%	10%
Y	SW	6%	5%
X	NW	5%	20%
W	SE	5%	20%
V	NW	20%	5%
U	NE	7%	10%

## Treatment Observation

State	Region	Outcome Rate	Poverty Rate
A	NW	5%	20%

## Comparison Group Candidates

State	Region	Outcome Rate	Poverty Rate
Z	NE	8%	10%
Y	SW	6%	5%
X	NW	5%	20%
W	SE	5%	20%
V	NW	20%	5%
U	NE	7%	10%

What is the best comparison for State A?

- ✓ Local
- ✓ Focal

## Treatment Observation

State	Region	Outcome Rate	Poverty Rate
A	NW	5%	20%

## Comparison Group Candidates

State	Region	Outcome Rate	Poverty Rate
Z	NE	8%	10%
Y	SW	6%	5%
X	NW	5%	20%
W	SE	5%	20%
V	NW	20%	5%
U	NE	7%	10%

What is the best comparison for State A?

- ✓ Local
- ✓ Focal
- ✓ Baseline Equivalence

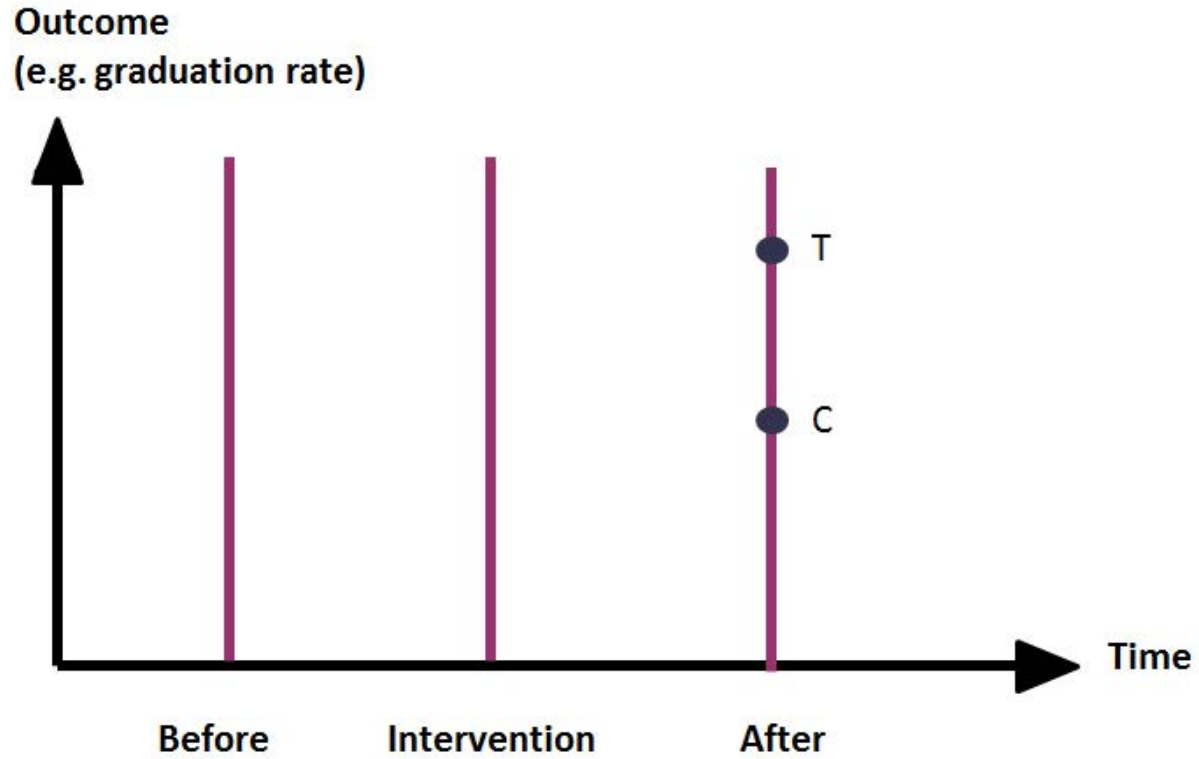
- With a large number of variables or a lot of observations, this can get out of hand quickly.
- Fortunately, you can let the computer do it for you!
- Methods such as Propensity Score Matching (PSM) and Mahalanobis Matching have been developed for this purpose.
- These methods do the same thing as hand-matching, but in a rigorous way.
  - i.e., they answer the question: For each community in the treatment group, which communities in the comparison group look most similar?

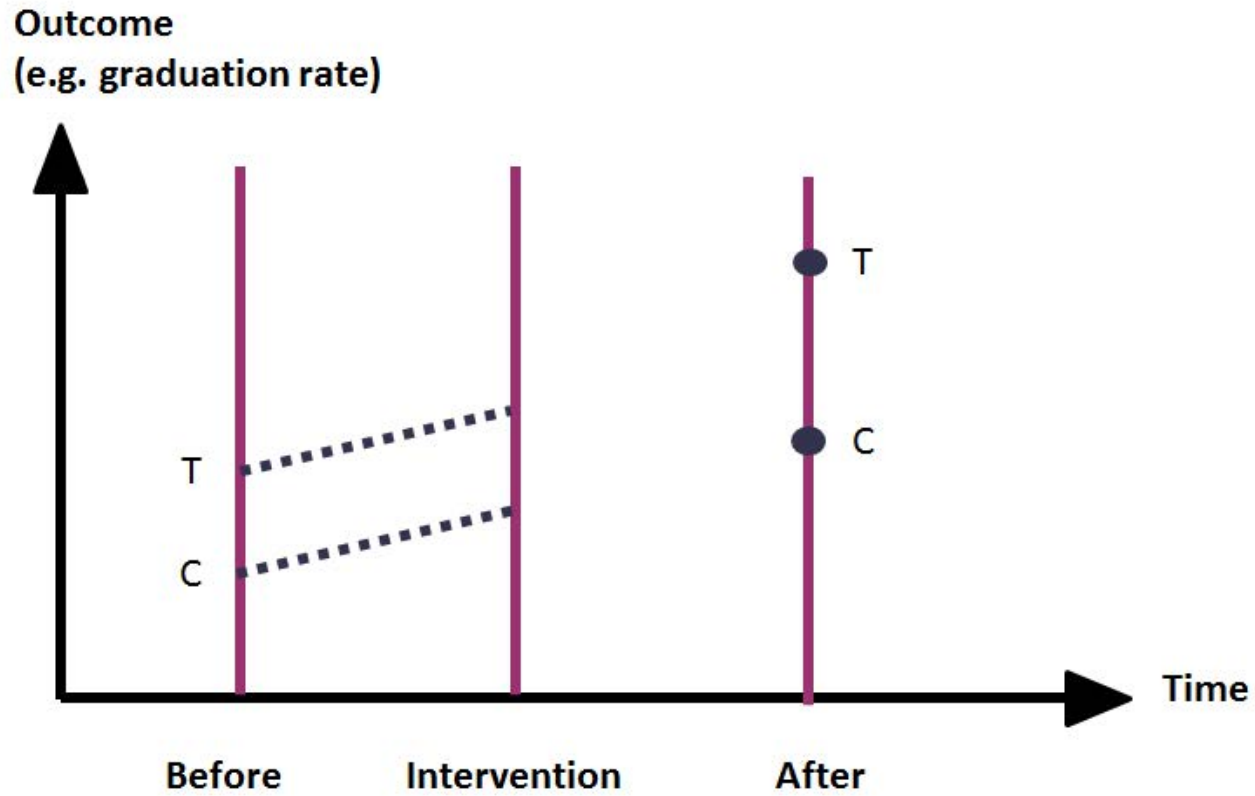
- Software-based methods work best when the pool of potential comparisons is **already as similar as possible to the treatment group**.
  - Matching can only account for observed differences, not unobservable ones.
- Matching should be done at the level of selection into the treatment.
  - Since you are treating whole communities, match to other communities.
  - Remember comparing apple trees, not apples.

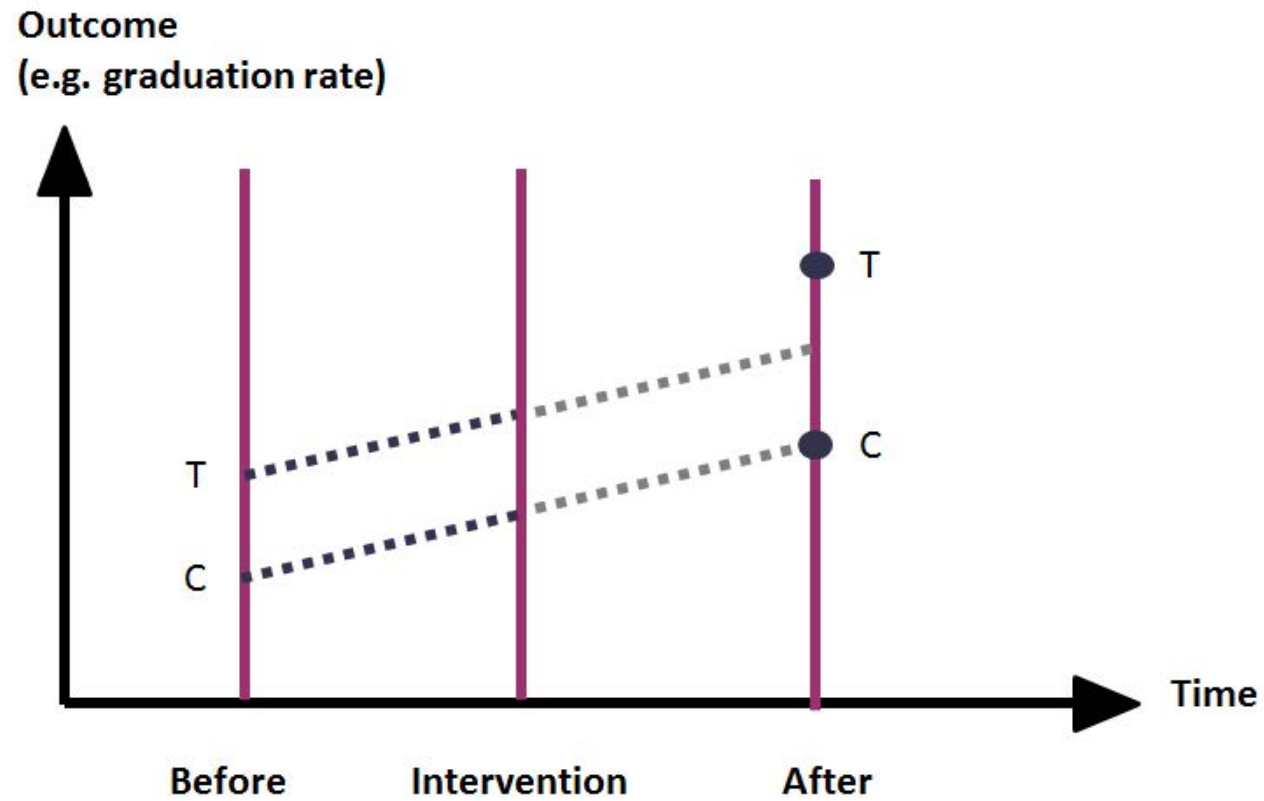
- Most statistical software will readily perform matching.
- We expect that grantees chosen for the federal evaluation will implement software-based matching.
  - We hope that many others will try matching.
  - Abt can provide additional resources on how to use this technique but not extensive TA.
- After matching, remember to check whether matched groups are balanced.

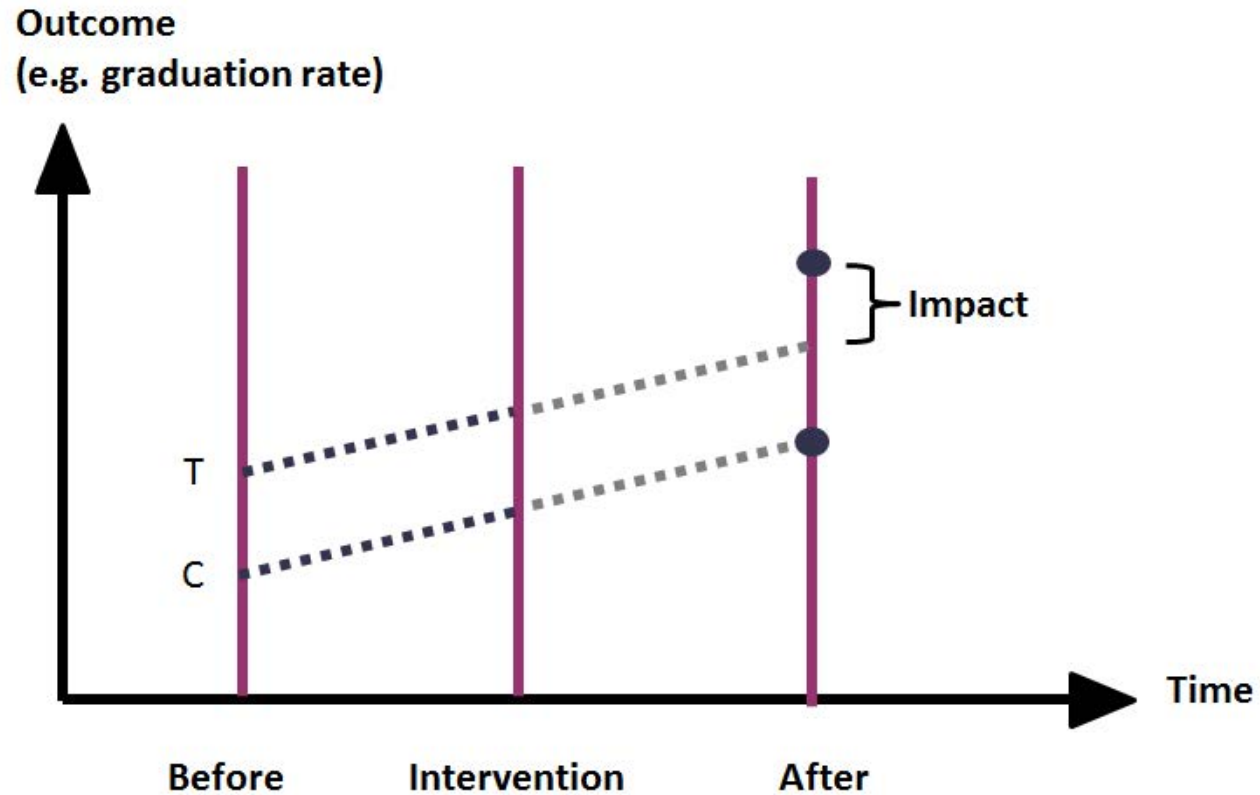


- You can use measurements over time to improve on the comparison-group design.
  - Administrative data often go back many years.
  - Using these data can strengthen analysis.
- The best of these is called a Comparative Short Interrupted Time Series (C-SITS) analysis.
  - Allows you to find a comparison group with similar pre-intervention *trends*, not just similar baseline characteristics.









- Several common designs are potentially problematic in this context. They include:
  - Using an individual-level comparison group.
  - Pre/Post or Interrupted Times Series (ITS).
  - Benchmarking to state/national trends.

## **Using a Within-Community Individual-Level Comparison Group**

Select individuals in your community who look like participants in EBPs as comparison group.

Compare outcomes between your participants and these comparison individuals.

- Looks a lot like what many Tier 1 grantees were encouraged to do in last round.
- Problematic because current implementation saturates communities.
- This makes it likely that individuals not directly receiving EBPs will still be affected by program.

## Interrupted Time Series

Uses a single group over time to measure impacts.

When there are many observations, you can establish a clear trend in the outcome and see if post-program observations deviate from the trend.

- ITS requires outcomes to be measured at *many points* over a *long time span*.
- Shadish, Cook, and Campbell assert that at least 100 observations are required.
- With only a handful of observations and no comparison group, this method is not very convincing.

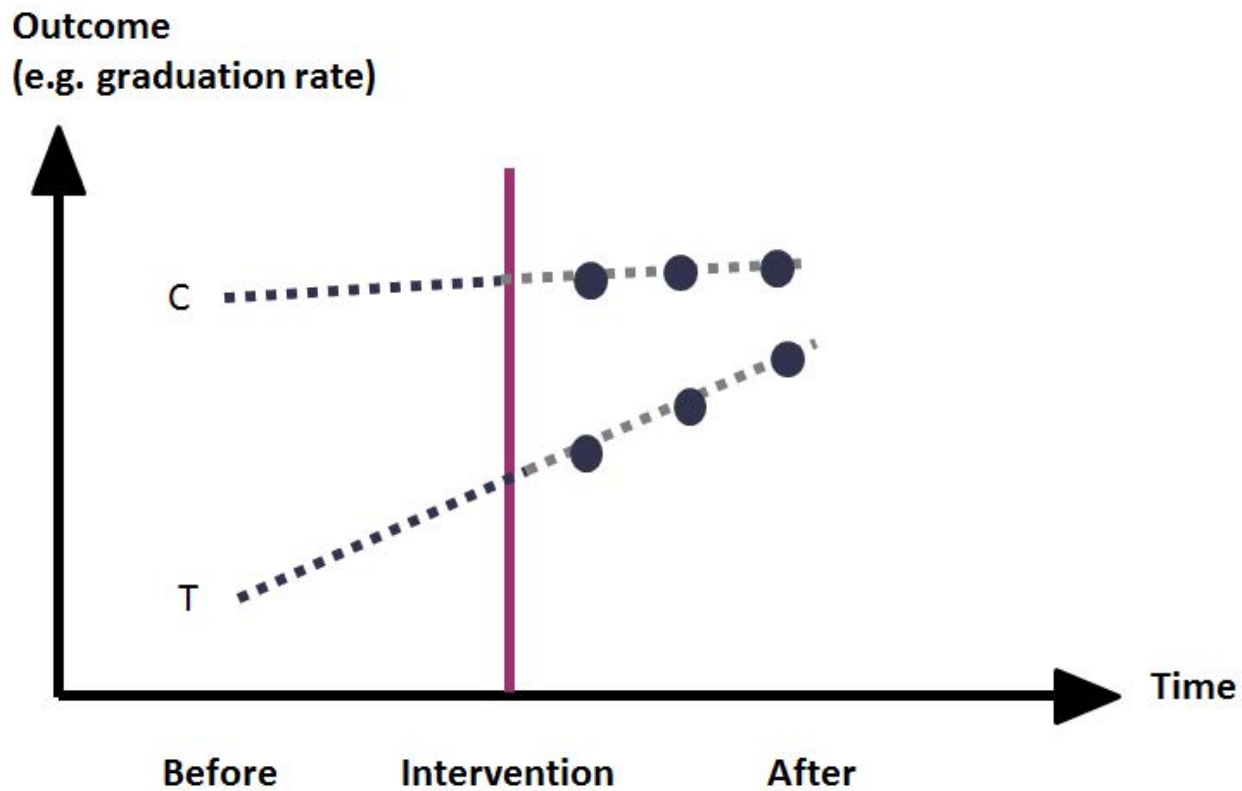


## Benchmarking to State or National Data

Several grantees have proposed various kinds of benchmarking; i.e., comparing the outcome in their communities to a state or national average for that outcome.

- Be careful! The trends in your carefully-selected treatment communities may be much different than the state or national trends.
- But it may be possible to implement a Difference-in-Differences or C-SITS using these data.

# C-SITS Analysis with Benchmarking



- Begin with a well-specified research question.
  - Intervention name, target population, counterfactual condition, and outcome domain.
- Select a comparison group that avoids confounds.
  - Local and focal.
  - Same level of aggregation and same data source.
- Match treatment and comparison groups.
- Test for baseline equivalence.
- Apply statistical adjustments if necessary.

**Randall Juras, Ph.D**

Randall\_Juras@abtassoc.com

**Kim Francis, Ph.D.**

Kimberly\_Francis@abtassoc.com

# Connect With Us!



Use OAH's Award Winning Website

[www.hhs.gov/ash/oah/](http://www.hhs.gov/ash/oah/)



Follow Us on Twitter

[@TeenHealthGov](https://twitter.com/TeenHealthGov)



Sign up for E-Updates



Watch us on YouTube

[www.youtube.com/teenhealthgov](http://www.youtube.com/teenhealthgov)

