CTE | Career & Technical Education RESEARCH NETWORK

Comparative Interrupted Time Series: Theory

August 22, 2020

Dr. Scott Davis

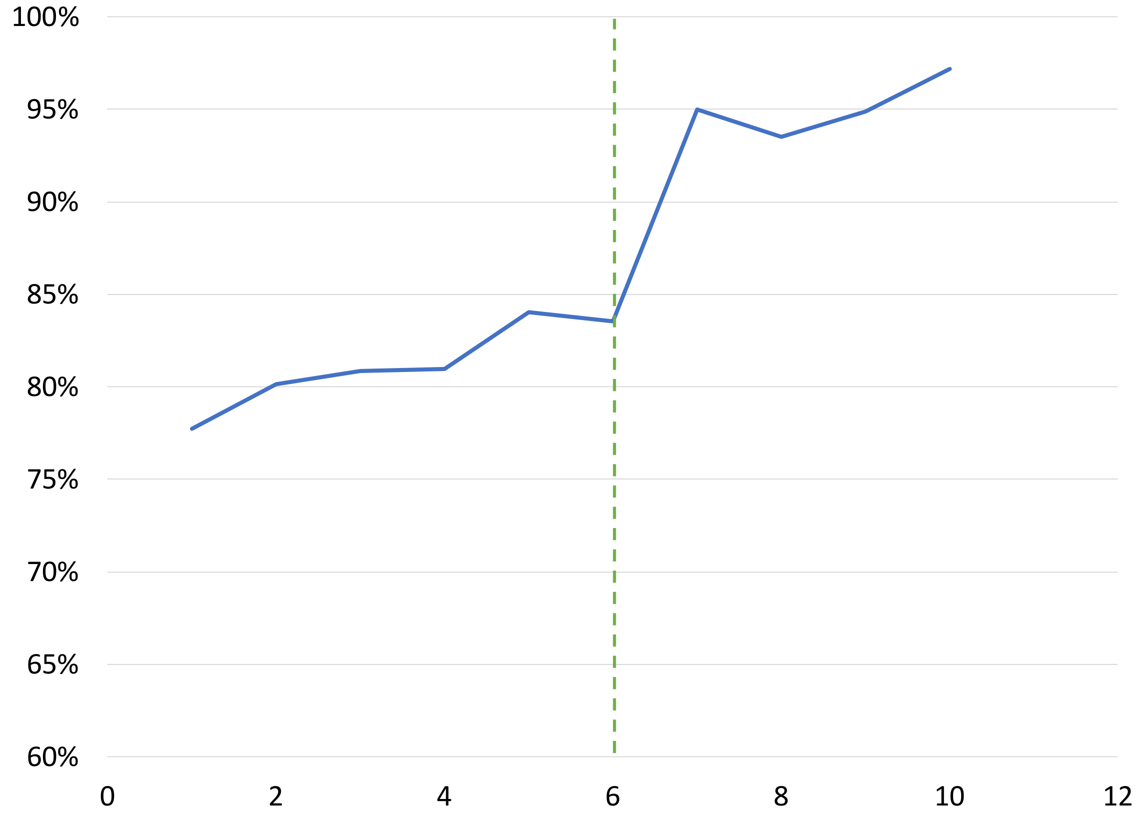
Principal Researcher

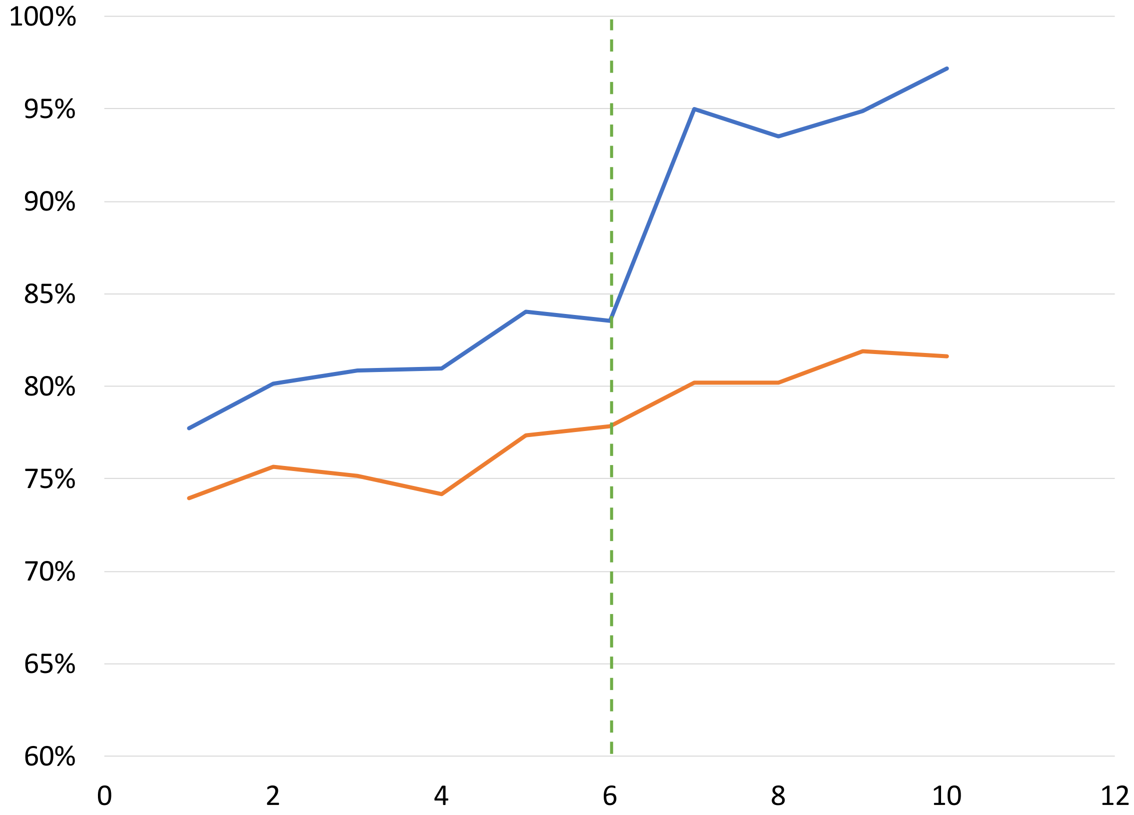
The work of the CTE Research Network Lead is supported by the Institute of Education Sciences at the U.S. Department of Education with funds provided under the *Carl D. Perkins Career and Technical Education Act* through Grant R305N180005 to the American Institutes for Research (AIR). The work of the 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

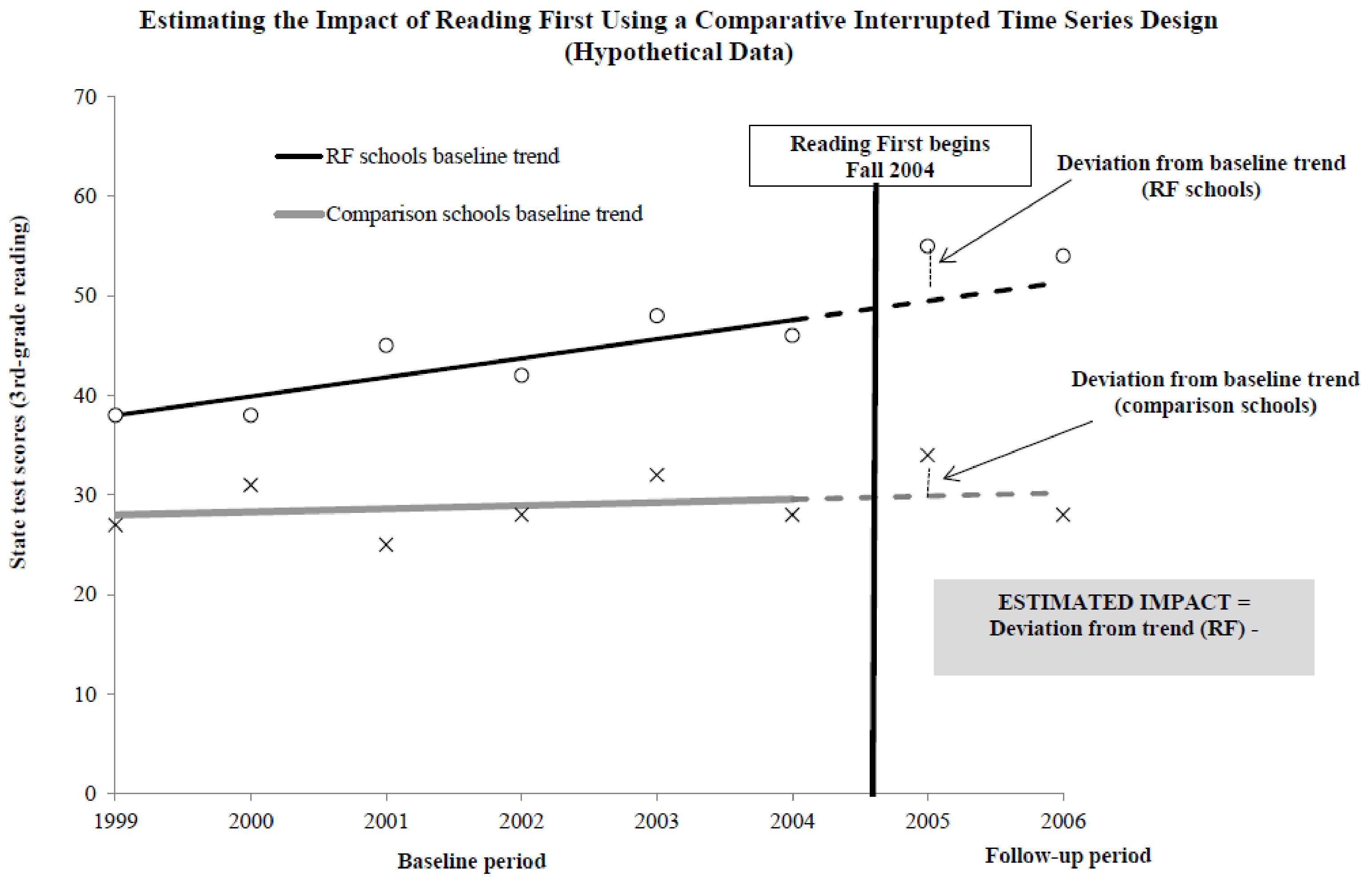
- The purpose is to introduce comparative interrupted time series (CITS) and some practical considerations when using them.
- The session will focus on CITS as applied to aggregate-level outcomes, but a similar approach can be used with individual-level outcomes.
- Structure and content follow:
  - Hallberg et al. (2018). *Short Comparative Interrupted Time Series Using Aggregate School-level Data in Education Research*
  - Reichardt (2019). *Quasi-Experimentation: A Guide to Design and Analysis*

Intuition Behind ITS and CITS





Visualizing CITS



Source: Somers, M., Zhu, P., Jacob, R., and Bloom, H. (2013). *The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation*. New York, NY: MDRC. Reprinted with permission of MDRC.

Conditions for Implementing CITS

- It is not feasible to implement an RCT.
- The treatment is implemented at an aggregate level.
- You have data covering multiple time periods.
- There is a good comparison group, meaning:
  - It is similar to the treatment group across time other than through the treatment.
  - It is exposed to the same history and instrumentation threats as the treatment.
- Data are measured consistently for both treatment and control groups.

Motivating CITS: Threats to Validity in ITS

- CITS addresses a number of potential treats to internal validity in ITS design:
  - **History**: Events that happen simultaneously with the treatment
  - **Selection**: A change in the composition of the treatment group from before to after the intervention
  - **Instrumentation**: A change in how the outcome is measured
- If these threats affect both the treatment and comparison groups the same way, then adding the comparison group eliminates bias.

Length of Time Series and Functional Form

- To implement CITS, you need data from multiple time periods.
  - Outcome data
  - Relevant characteristics of both treatment and comparison groups
- The minimum number of time periods depend on the specific modeling approach.
  - Need to observe at least two preintervention time periods
  - Need at least three preintervention time periods if using more flexible functional forms

Identifying a Comparison Group

- There are five common approaches to identify the comparison group:
  - **Using all available nontreatment schools**: include data from all available nontreatments in the same district, state, or country
  - **Matching**: selection based on preintervention characteristics, including the outcome measure
  - **Focal selection**: schools that match in terms of observable characteristics
  - **Local**: schools that are geographically close to treatment schools
  - **Hybrid**: mix of both focal and local matching
- There is some evidence that hybrid matching can reduce bias but no clear support for any single matching approach.

Estimating Treatment Effects: Baseline Mean Model

- Effects are estimated via CITS using regression analysis.
- The **baseline mean model** is the simplest approach and resembles the difference-in-difference model:

$$Y_{jt} = \beta_0 + \beta_1 Z_{jt} + \beta_2 \text{Post}_t + \beta_3 T_j + \beta_4 (\text{Post}_t \cdot T_j) + v_j + u_{jt}$$

  - Schools are indexed by *j*, and time is indexed by *t.*
  - Z_{jt} is a set of school characteristics, Post_t is a vector of indicators for each post-treatment period, T_j is an indicator for whether the school is in the treatment group, v_j is a school-level error term, and u_{jt} is a mean-zero error term.
  - The parameters of interest are the elements of the vector β_4.

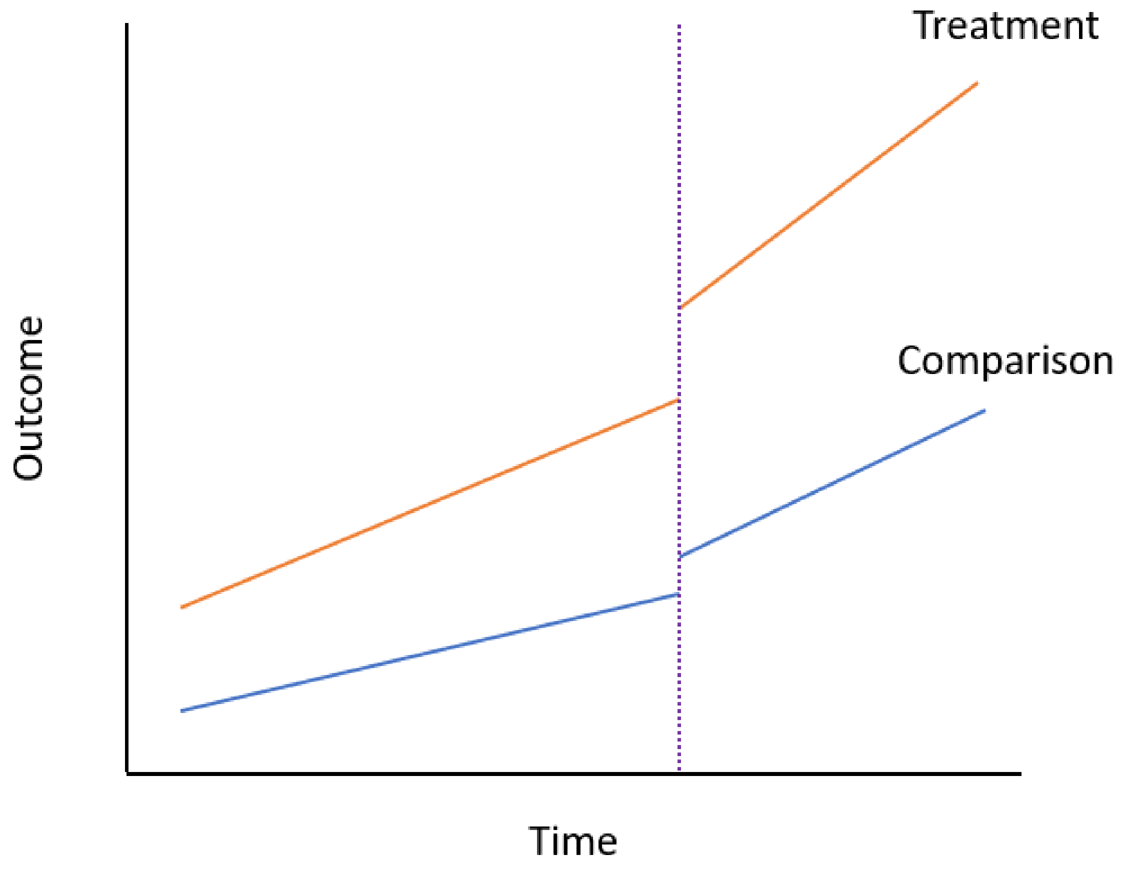
Estimating Treatment Effects: Baseline Linear Trend

- The **baseline linear trend model** extends the baseline model by adding a linear term for time:

  $$Y_{jt} = \beta_0 + \beta_1 \text{Time}_t + \beta_2 Z_{jt} + \beta_3 \text{Post}_t + \beta_4 T_j + \beta_5 (\text{Time}_t \cdot T_j) + \beta_6 (\text{Post}_t \cdot T_j) + v_j + u_{jt}$$

  - Schools are indexed by *j*, and time is indexed by *t.*
  - β_1 is the pretreatment slope of the outcome trend for the comparison group, β_5 is the slope for the treatment group.
  - β_6 represents the treatment effect.

Estimating Treatment Effects: Extensions



Estimating Treatment Effects: Extensions

- The **baseline nonlinear-trend model** extends the baseline linear-trend model by adding a function of time.
  - This approach requires more data for the preintervention period.
- The **school and year fixed-effects model** uses year and school indicators to control for the relationships between schools, time, and the outcome:

$$Y_{jt} = \sum_{t=0}^{T} \beta_t \text{Year}_t + \beta_T \text{T\_Post}_{jt} + \sum_{k=0}^{N} \beta_{sk} s_k + u_{jt}$$

  - β_T measures the difference in average performance.

Estimating Treatment Effects

- There is no clear guidance on which approach to use in a specific circumstance.
- The best approach is to plot the data to help guide the choice.
  - The **baseline mean model** is appropriate when the pre-intervention data look flat and parallel between treatment and control groups.
  - The **baseline linear trend model** is appropriate when the pre-intervention data show different slopes for the two groups.
  - The **baseline nonlinear trend model** or the **school and year fixed effects model** may perform best when patterns are not obvious or are difficult to interpret.
- Once you choose the model, you estimate the regression model (typically with OLS).

Knowledge Check

CITS Modeling Approaches

Diagnostics and Robustness Checks

- It is a good practice to assess sensitivity of your result to your design choices.
- Do this by re-estimating the treatment effect using other alternatives.
- In practice, there are two key aspects of the design to change:
  1. How the comparison group is chosen
     - Can change both the parameters and method
  2. Change the modeling approach
- The goal is to see whether the alternative results are similar to your preferred model.

Additional Design Elements for Augmented CITS

- There are other ways to augment the basic CITS model.
- In a **multiple baseline** design, the same treatment is implemented in two or more different places at different times.
  - Estimate the treatment effect for each instance of the treatment.
  - This approach helps reduce history threats.
- You may also analyze **nonequivalent dependent variables.**
  - The idea is to estimate the treatment effect on an outcome related to the outcome of interest, but that should not be affected by the treatment.
  - This approach can reinforce the credibility of the estimated treatment effect.

WWC Standards for CITS

- To be eligible for WWC review as a group design study,
  - Participants must be assigned to only one group (treatment or comparison).
  - Pre- and post-intervention periods must be accounted for.
  - The model must include an interaction between the post-intervention period and the treatment indicator.
  - With multiple preintervention periods, baseline equivalence must be assessed from a single period of preintervention data.
- Treatment effects from different postintervention time points will be reviewed separately.
- Direction of the impact estimate at a specific time point must be reported.
- Highest possible rating for a CITS study is **Meets WWC Group Design Standards With Reservations.**

Comparative Interrupted Time Series: Practice

Overview

- The goal is to let you practice carrying out CITS analysis using real data.
  - First, I will demonstrate some examples.
  - Second, you will have time to work in small groups through a series of activities.
- The first part is meant to show you how to implement some of the techniques we discussed.
  - All commands are from **Stata** but have **R** equivalents
- After your independent work time, you will meet with trainers in virtual office hours.

Example Analysis Scenario

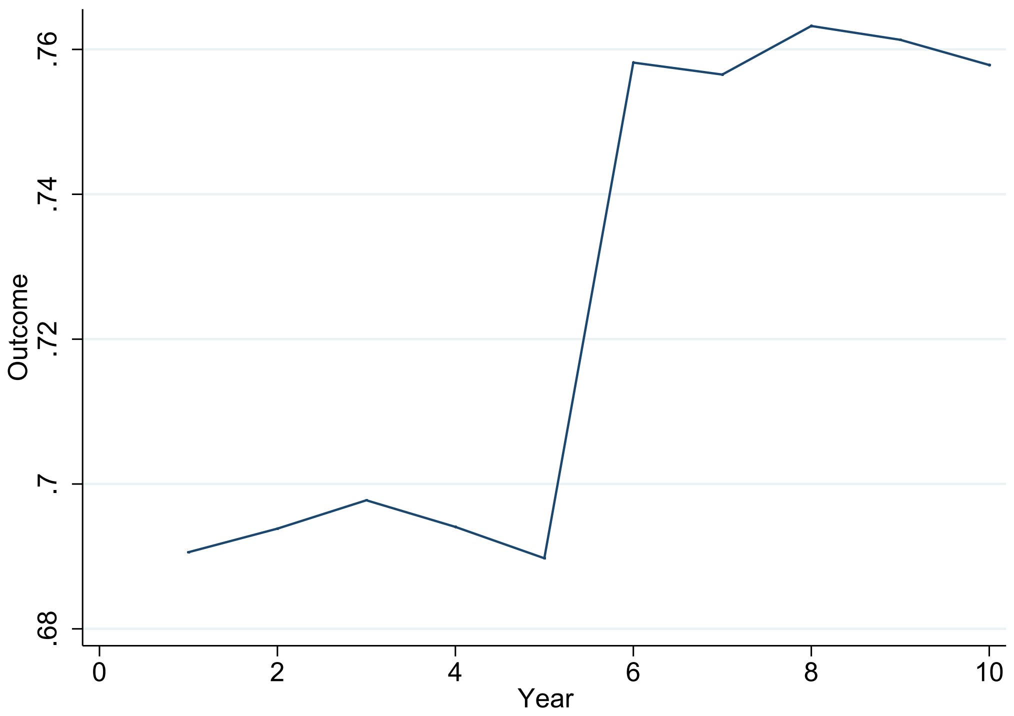
- Suppose a state implemented an innovative work-based learning program in select high schools, and we want to know whether the program improved high school graduation rates.
- Imagine we have data on schools covering several years before (Years 1–5) and several years after (Years 6–10) the program was implemented.
- We can use a **CITS** approach to estimate the effect of the program on high school graduation rates.
- For this example, the data are based on a simulated data set available as supplementary material from Hallberg et al. (2018).

Example Analysis Variables and Approach

- For our example scenario, the unit of observation is a school.
- The key variables in the data set are:
  - **Y**, the outcome of interest (high school graduation rate)
  - **X**, which measures time from Year 1 to Year 10
  - **T**, an indicator for observations in the treatment group
  - **S**, a set of indicators identifying schools (N = 80)
  - **G**, a group variable (think of this as identifying school districts)
  - **P**, an indicator for post-intervention time periods (Years 6–10)
  - **Z1**, a school characteristic (e.g., percentage of students eligible for free or reduced-price lunch)

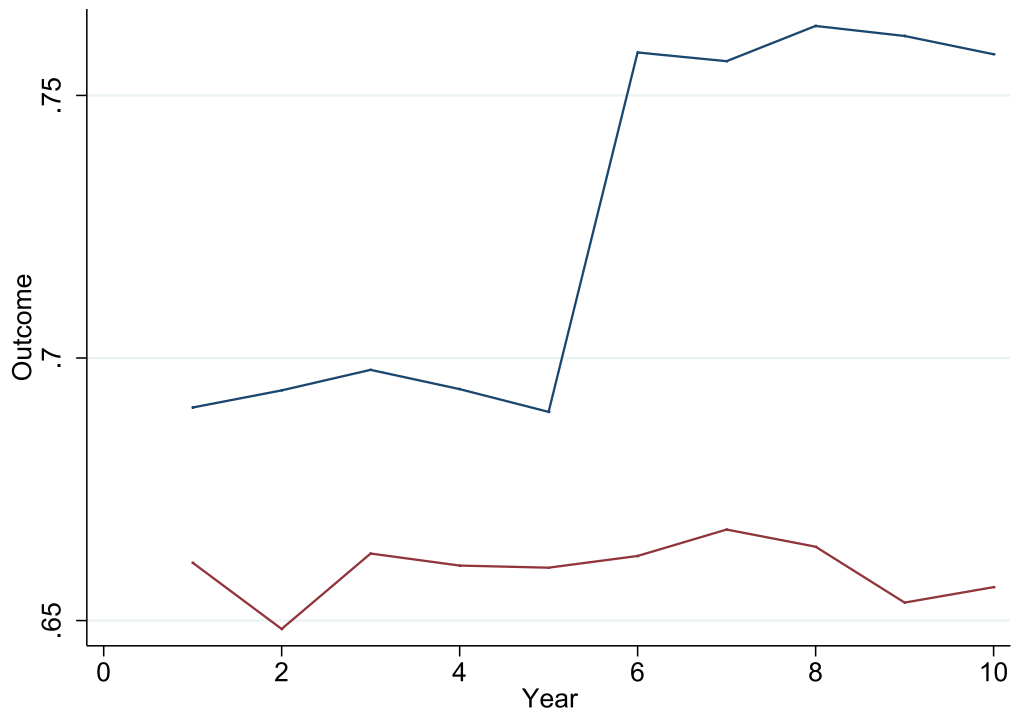
Step 1: Plot the Data (Treatment)

```
capture drop Y_bar

bysort X T: egen Y_bar =
mean(Y)

twoway (line Y_bar X if T),
graphregion(color(white))
ytitle("Outcome") legend(off)
```



Step 1: Plot the Data (Treatment + Comparison)

```
twoway (line Y_bar X if T)
(line Y_bar X if T == 0),
graphregion(color(white))
ytitle("Outcome") legend(off)
```



Step 2: Choosing a Comparison Group (Local)

- One option for choosing the comparison group is to use those schools that are close to the treatment schools, such as those in the same districts.
- The variable G in our data set identifies districts.
  - The treatment schools are all from districts 1 and 2.
  - The comparison schools are split between districts 1–4.
- If we want to use local matching to choose the comparison group, we limit the comparison group to schools in districts 1 and 2.
- We will create an indicator for this comparison group called C1.

```
gen C1 = (G == 1 | G == 2) & T == 0
```

Step 2: Choosing a Comparison Group (Focal)

- Another option for choosing a comparison group is to select schools that match the treatment schools in terms of observable characteristics.
- Identifying these is a multi-step process.
  1. Estimate the propensity score – the probability of being in the treatment group.
  2. For each treatment group school, identify the best match (or multiple matches) from among the comparison group schools by comparing propensity scores.
  3. Use the set of all the satisfactory matches as the comparison group.
- We will create an indicator for this comparison group called C2.

Step 2: Choosing a Comparison Group (Hybrid)

- A third option for selecting a comparison group is to blend the first two approaches.
  - Use propensity-score matching to identify the best matches for each treatment group school.
  - Among the best matches, pick the one(s) that are from the same district.
- In our example, that means picking the best match from among the schools in districts 1 and 2.
- We will create an indicator for this comparison group called C3.

Step 3: Estimate the Treatment Effect

Baseline mean model:

$$Y_{jt} = \beta_0 + \beta_1 Z_{jt} + \beta_2 P_t + \beta_3 T_t + \beta_5 PT_t + v_j + u_{jt}$$

```
mixed Y Z1 P T PT if T | C1 || S:
```

- The estimated effect is 0.066, and it is statistically significant ($p < 0.05$).

```
Mixed-effects ML regression                     Number of obs     =        600
Group variable: S                               Number of groups  =         60

                                                Obs per group:
                                                              min =         10
                                                              avg =       10.0
                                                              max =         10

                                                Wald chi2(4)      =    2384.61
Log likelihood =  1186.6415                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          Z1 |   .6681269   .0167823    39.81   0.000     .6352343    .7010196
           P |   .0005959   .0047354     0.13   0.900    -.0086853    .0098772
           T |    .000708   .0041491     0.17   0.864     -.007424    .0088401
          PT |   .0656329   .0057997    11.32   0.000     .0542657        .077
       _cons |   .3450565    .008762    39.38   0.000     .3278834    .3622296
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
S: Identity                  |
                  var(_cons) |   7.48e-23   4.21e-22      1.21e-27    4.63e-18
-----------------------------+------------------------------------------------
               var(Residual) |   .0011212   .0000647      .0010012    .0012555
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 0.00          Prob >= chibar2 = 1.0000
```

Step 4: Robustness Checks (Baseline Linear Trend)

Baseline linear trend model:

$$Y_{jt} = \beta_0 + \beta_1 X_t + \beta_2 Z_{jt} + \beta_3 P_t + \beta_4 T_t + \beta_5 XT_t + \beta_6 PT_t + \upsilon_j + u_{jt}$$

```
mixed Y X Z1 P T XT PT if T | C1 || S:
```

- The estimated effect is 0.065, and it is statistically significant.

```
Mixed-effects ML regression                     Number of obs     =        600
Group variable: S                               Number of groups  =         60

                                                Obs per group:
                                                              min =         10
                                                              avg =       10.0
                                                              max =         10

                                                Wald chi2(6)      =    2384.69
Log likelihood =  1186.6498                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |   .0000956   .0016742     0.06   0.954    -.0031857     .003377
          Z1 |   .6681269    .016782    39.81   0.000     .6352347    .7010191
           P |   .0001177   .0096175     0.01   0.990    -.0187323    .0189677
           T |   .0005872   .0074199     0.08   0.937    -.0139554    .0151299
          XT |   .0000403   .0020505     0.02   0.984    -.0039786    .0040591
          PT |   .0654315    .011779     5.55   0.000      .042345     .088518
       _cons |   .3447696   .0100993    34.14   0.000     .3249753    .3645639
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
S: Identity                  |
                  var(_cons) |   7.71e-23   4.31e-22      1.33e-27    4.47e-18
-----------------------------+------------------------------------------------
               var(Residual) |   .0011212   .0000647      .0010012    .0012555
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 1.8e-12       Prob >= chibar2 = 1.0000
```

Step 4: Robustness Checks (School + Year Fixed Effects)

School and year fixed effects model:

$$Y_{jt} = \sum_{t=0}^{T} \beta_t X_t + \beta_{PT} PT_{jt} + \sum_{k=0}^{N} \beta_{sk} S_k + u_{jt}$$

```
regress Y PT i.X i.S if T | C1
```

- The estimated effect is 0.066 and is statistically significant.

```
. regress $Y PT i.$X i.$S if $T | $C1

      Source |       SS           df       MS      Number of obs   =       600
-------------+----------------------------------   F(69, 530)      =     33.08
       Model |  2.71572828        69  .039358381   Prob > F        =    0.0000
    Residual |  .630641271       530  .001189889   R-squared       =    0.8115
-------------+----------------------------------   Adj R-squared   =    0.7870
       Total |  3.34636955       599  .005586594   Root MSE        =    .03449

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          PT |   .0656329   .0059747    10.99   0.000     .0538959    .0773698
             |
           X |
          2  |  -.0009349   .0062979    -0.15   0.882    -.0133067     .011437
          3  |   .0072851   .0062979     1.16   0.248    -.0050867     .019657
          4  |    .004563   .0062979     0.72   0.469    -.0078088    .0169348
          5  |   .0011531   .0062979     0.18   0.855    -.0112187    .0135249
          6  |   .0023749   .0074517     0.32   0.750    -.0122636    .0170134
          7  |    .003869   .0074517     0.52   0.604    -.0107695    .0185075
          8  |   .0060583   .0074517     0.81   0.417    -.0085803    .0206968
          9  |   .0022233   .0074517     0.30   0.766    -.0124152    .0168619
         10  |   .0005206   .0074517     0.07   0.944    -.0141179    .0151592
             |
           S |
          2  |   .1028354   .0154265     6.67   0.000     .0725308    .1331401
          3  |   .0943448   .0154265     6.12   0.000     .0640401    .1246494
          4  |   .0953538   .0154265     6.18   0.000     .0650491    .1256584
          5  |   .0861139   .0154265     5.58   0.000     .0558092    .1164185
          6  |   .0719275   .0154265     4.66   0.000     .0416228    .1022321
          7  |   .0704953   .0154265     4.57   0.000     .0401906    .1007999
```

Step 4: Robustness Checks (Summary)

The estimated effects are consistent regardless of specification or choice of comparison group.

Model	C1 (Local)	C2 (Focal)	C3 (Hybrid)
Baseline mean	0.066***	0.060***	0.061***
Baseline linear trend	0.065***	0.059***	0.064***
School + year fixed effects	0.066***	0.060***	0.061***

Step 5: Drawing a Conclusion

- Our analysis shows that the work-based learning program led to a 6–7 percentage point increase in the high school graduation rate.
- Suppose our preferred specification is the baseline linear trend model using comparison group C3.
  - The estimated effect is an increase of 6.4 percentage points.
  - The estimate represents an increase of 9 percent over the average graduation rate (67.6 percent) in the years before implementation.

Comparative Interrupted Time Series: Small Group Work

Overview

- We'll send you a CITS Activity Guide just after we end this morning.
  - The guide presents a series of tasks for your group to complete, using the data sets we've provided.
- You have the rest of the afternoon to work on the tasks.
- Each group will have some time to meet with our trainers to talk about your progress and ask questions.
  - The agenda tells you when each group will meet and with whom.
- At 4:30, we will reconvene as a group to recap today's session.

References

Somers, M., Zhu, P., Jacob, R., and Bloom, H. (2013). *The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation*. New York, NY: MDRC. Reprinted with permission of MDRC.

CTE Research Network Lead

The work of the CTE Research Network Lead is supported by the Institute of Education Sciences (IES), U.S. Department of Education, through grant R305N180005 to the American Institutes for Research. The opinions expressed are those of the authors and do not represent views of IES or the U.S. Department of Education.

CTEResearchNetwork.org

CTEResearchNetwork@air.org