CPP 524: Foundations of Program Evaluation II

Evaluation Design Project – Step #4
Diagram Your Study Design

This exercise is designed to help you formalize and communicate the research design for your proposed evaluation.

This exercise looks specifically at how you are constructing your counterfactual and which estimator you will use to measure program impact.

The three basic counterfactuals and their complimentary estimators are:

- Reflexive design: T2 – T1
- Post-test only design: T2 – C2
- Pre-post with comparison design: (T2-T1) – (C2-C1)

There are additional variants of these designs (the interrupted time series combines elements of the first and third, for example), but most research design maps onto one of these three approaches to estimation.

There are two main things to pay attention to:

1. What are the assumptions needed for each estimator to be unbiased?
2. Is your counterfactual a good representation of what the world would look like in the absence of the treatment?

Part of this exercise is reasoning through what a robust comparison would look like. If you have full control over the world how would you construct a comparison group so it is a robust counterfactual?

And partly it is making sure that your estimator captures the intended counterfactual.

Recall that you CANNOT use a Randomized Control Trial (RCT) as the means of constructing your counterfactual analysis (it is too simple). You must use one of the other quasi-experimental approaches. This might include randomization is a process exists independent of your study (a lottery for charter school seats, the draft for Vietnam, etc.).

Define Your Counterfactual

Explain your research design in terms of the following:

1. Who is included in the treatment group? Who qualifies for the program and what does the "typical" participant look like?
2. What would your ideal counterfactual then look like? Can you point to another group that would be more or less identical to the treatment group prior to the treatment? This needs to be a group that can be observed (outcomes measured for the group) and you can guarantee will not receive the treatment.

The difficulty of the second question will depend upon the degree of selection in the program. If participation is more or less random then participants and non-participants will look a lot alike. It should be easy to find cases that could represent your counterfactual.

If there are strong selection mechanisms into the program, and the drivers of program participation are highly correlated with outcomes, then counterfactual reasoning becomes a lot more challenging.

For example, a mandatory summer school requirement for students that fail their high school math standardized exams creates a problem. Selection into the program is basically a measure of their mathematical ability, so we know that program participants will be fundamentally different from non-participants. Comparing math performance of summer school students to the rest of the student population is definitely an apples to oranges comparison that will make summer school look very ineffective, even if students in the program progressed from their baseline performance over the summer.

Conversely, if you want to know whether competing in an Iron Man competition increases risk of a heart attack, your counterfactual group should be a group with equal fitness levels but that does not participate. How do you find a group that is physically fit enough to compete in an Iron Man race but chooses not to?

The goal is to define a group that is arguably equivalent enough to your treatment group that it can be used for counterfactual analysis. You can manually construct this group by applying observable criteria that allows you to isolate a part of the data that can serve as a strong counterfactual.

For example, in the summer school case can you look for students that recently moved from neighboring school district that did not require summer school? Find students that had scores on the state-wide standardized math exam that would have forced them into summer school, but they did not attend because it was not available in their district. But now they have moved to the same district as the summer school kids, so you can look at math scores the following year for these two groups. Do summer school kids perform better than the comparison group of kids that scored the same on the standardized tests, but did not attend summer school because they lived in different districts at the time? One can find weaknesses with this comparison, but since their pre-treatment performance is measured by a state-wide standardized exam one could make a strong case that they would serve as a good counterfactual to show what performance would look like for the summer school kids if they had not attended summer school.

You are invite to be creative in how you construct the group, as long as it seems feasible to get data on the counterfactual group in your proposed study.

Diagram Your Groups

Create a hypothetical diagram that shows the group means for your treatment and comparison groups before and after the treatment occurs. In constructing this diagram answer the following questions:

(1) Do the treatment and comparison group have identical outcomes prior to the treatment?
(2) Do you anticipate any secular trends in the study (gains independent of the treatment over the study period)?
(3) Does program success mean that the outcome increases (e.g. lifetime earnings) or decreases (e.g. crime rates)?

Make sure your diagram reflects your answers to these three things, and explain your assumptions about each. Your assumptions are important because when they are met you can use simpler (less data-intensive, thus cheaper to implement) estimators.

If they are not met then your estimator might be biased.

Note, your treatment might be levels and not groups. For example, perhaps your question is: does Medicaid spending reduce COVID fatalities in states? Your treatment would be measured as millions of US dollars of Medicare funding received by a state (in per capita units). Since all states receive Medicare there is no treatment and control group. In a case like this define one group as "high level of aid" and one group as "low level of aid" and proceed with these two simplified cases for the exercise.

Describe Your Estimator:

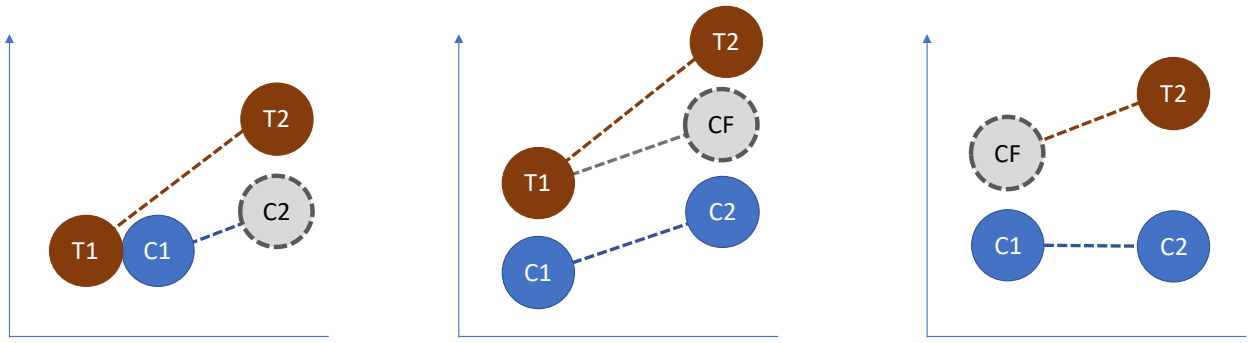
Which estimator will you use to measure program impact? These match the order of the diagram below.

- T2 – C2 (post-test only design)
- (T2-T1) – (C2-C1) (difference-in-difference)
- T2 – T1 (reflexive design)

Does your proposed estimator match the counterfactual you proposed for your study?

STEP 4: WHERE IS YOUR COUNTERFACTUAL?

What would your treatment group look like if it had not received the treatment?



C1=T1

C2 is the counterfactual

C1≠T1

T1 + (C2-C1) is the counterfactual

C1=C2

T1 is the counterfactual

Regression Model (bonus points)

Write down the regression model that captures your proposed counterfactual.

- Include a list that describes each regression coefficient, especially dummy variables you are using in your model.
- Explain which regression coefficient will capture program impact.
- Is the significance test of the coefficient the same as the significance test that the program has achieved impact?

Deliverables:

Submit 1-2 pages with your hypothetical diagram of expected study results, and an explanation of the construction of your counterfactual group and how it fits with your proposed estimator.

This step is instrumental in understanding whether an evaluator is estimating something meaningful in the study, or is just putting data into a model without having a clear sense of what program impact would look like.

As you have learned while reviewing Chapters 20 and 21 on the Milwaukee School Voucher Program, two evaluation teams can use the exact same data and come to different conclusions about program effectiveness. This results from the teams constructing the counterfactuals in different ways. There is often more than one defensible approach, and trade-offs between them. You should at the very least be cognizant of the counterfactual you are building into your model, and in the ideal be able to look at a question from several angles using different counterfactuals to find the one that best fits the situation.