Cyber Seminar Transcript  
Date: 04/08/2015

Series: HEC  
Session: Difference-in-Difference

Presenter: Christine Chee  
 *This is an unedited transcript of this session. As such, it may contain omissions or errors due to sound quality or misinterpretation. For clarification or verification of any points in the transcript, please refer to the audio version posted at www.hsrd.research.va.gov/cyberseminars/catalog-archive.cfm or contact:* ***[christine.chee@va.gov](mailto:christine.chee@va.gov)***

Risha Gidwani: Hello everyone. I am Risha Gidwani, a health economist here at HERC. It is my pleasure to introduce today’s speaker, Christine Pal Pal-Chee. Dr. Chee is also an economist with HERC. She has been here since October of 2012, and came to us after receiving her Ph.D. in economics from Columbia University. Dr. Pal-Chee is a health economist who is interested in the efficiency and effectiveness of healthcare systems. Her research focuses on the delivery and quality of healthcare, on physician and hospital behavior, patient health behavior, and health and economic wellbeing, as well as determinants of health team vitalization and spending. We are very happy to have her present to us today on difference and difference models. Christine, take it away.

Dr. Christine Pal-Chee: Thank you Risha. Thank you Heidi. We will jump right in. Today, I will be discussing natural experiments and using difference-in-differences to estimate causal treatment effects. To start, I will briefly discuss causal effects in randomized controlled trials, to provide some motivation for the topics we will be discussing today. Then we will continue our discussion of natural experiments and difference-in-difference estimators of causal treatment effects.

Before we do that, it would be really helpful to get a sense of the background. To do that, I want to put up two polls. I will ask Heidi to help me there.

Heidi: Yes, the first one is up here. The question is which of the following best describes your familiarity with natural experiments.

Dr. Christine Pal-Chee: Thank you Heidi. You can choose the first option if you are very familiar with the concept of natural experiments. You can select the second option if you have a working understanding of what natural experiments are. The third is if you are new to the concept of natural experiments.

Heidi: Responses are coming in. We will give you all just a few more moments before we close the poll out.

[Silence]

Heidi: It looks like things are slowing down a little bit here. I am going to close this out and we are going to go through the results. We are seeing around 17% saying they are very familiar with the concept of natural experiments. Forty-four percent have a working understanding of what natural experiments are. Thirty-nine percent are new to the concept of natural experiments. Thank you every one.

Dr. Christine Pal-Chee: Thank you Heidi. It looks like there is quite a range of backgrounds when it comes to natural experiments. It looks like a majority of the group is fairly new to the concept of natural experiments. I think this reflects the new interest and energy around natural experiments in health services research. I think this is great for today.

My second question has to do with difference-in-differences. This is our second poll. Is the poll up Heidi?

Heidi: The poll is up.

Dr. Christine Pal-Chee: If you could, choose the option that best describes your familiarity with difference-in-differences. The first option is for you if you are very familiar with difference-in-differences. The second is if you have a working knowledge of difference-in-differences. Perhaps you have seen it in research. The third is if you are new to difference-in-differences.

Heidi: We will give everyone just a few more moments before closing this out.

[Silence]

Heidi: We are waiting for responses to slow down a little bit. It looks like we can close it out here. We are seeing 10% saying they are very familiar with difference-in-differences. Thirty-nine percent say they have a working knowledge of difference-in-differences. Fifty-one percent are new to difference in differences. Thank you everyone.

Dr. Christine Pal-Chee: Thank you Heidi. Again, it looks like there is quite a range, but more people who are new or are fairly new to difference-in-differences. This, I think, reflects the new interest and energy around natural experiments and using them to estimate causal treatment effects. This is great because I think this is in line with the objectives for this lecture.

The first objective for this lecture is to provide an overview of natural experiments. Here, I will provide some motivation. I will talk about what natural experiments are, describe them, and then provide a few examples. The second objective is to provide an overview of the difference-in-differences estimator. Here again, I will start with some motivation. I will define the difference-in-differences estimator. We will walk through an example, and I will also discuss the functions and limitations of difference-in-differences.

The goal here is to provide a broad overview of what natural experiments are, and what the difference-in-differences estimator is, with the hope of helping the audience develop a broad understanding of this. You can think about it in your own research or be able to think about it differently, or in a new way, when reading other’s research.

In the lecture on research design from about two weeks ago, we highlighted the fact that many questions in health services research aim to estimate causal effects. I would suggest a few examples. Does the adoption of electronic medical records reduce healthcare costs or improve quality of care? In the VA, does the transition to patient aligned care teams in primary care improve quality of care outcomes? What effect will the Affordable Care Act have on the demand for VA healthcare services? These are all questions that get at a causal relationship.

We also discussed in that research design lecture, how these questions are ideally studied through randomized controlled trials. In the context of randomized controlled trials, we can ask what the effect is of receiving some treatment on an outcome or outcomes that we are interested in. we can specify a simple regression model that looks like something like the following, where our outcome variable of interest is our dependent variable, and the main explanatory variable of interest is a binary treatment variable that equals one, the person receives treatment, and zero, the patient did not receive treatment.

We should keep in mind that the regression model is a sort of conceptual model that specifies how the dependent variable is determined. E, our error term, includes all other factors that affect the outcome. These things can include age, gender, preexisting conditions, income, education, and a whole range of other things. In the context of a randomized controlled trial, treatment is randomly assigned. Treatment is exogenous. In that, conditional treatment, conditional in receiving or not receiving treatment, the expected effect of all these other factors is zero. This implies that the error, E, our error term, and treatment are uncorrelated. This allows us to identify just the effect of treatment.

In this case, when this assumption is true, when treatment is exogenous, our OLS estimator, beta-1 or beta-1 \_\_\_\_\_ [00:08:24], which is what we get from linear regression, estimates the average effect of treatment. This is all great when we have a randomized controlled trial. Everything is very clean. We can very easily estimate causal treatment effects. In order to estimate causal treatment effects, we can just randomly assign treatment. Unfortunately, this is not always feasible, ethical, or practical. I believe that thinking about randomization and randomized controlled trials is useful as a conceptual benchmark or gold standard in terms of research design for observational studies. It is also helpful to think about when we are thinking about natural experiments, which basically mimic randomized controlled trials.

What is a natural experiment? A natural experiment occurs when external circumstances produce what appears to be randomization. These factors, these external factors, can include legal institutions, geography, the timing of policies, program implementation, and natural randomness in terms of weather, birth dates, or any other factors that are unrelated to the causal effect of interest. In natural experiments, variation in individual circumstances make it appear as if treatment is randomly assigned. We have exogenous variation in treatment that allows us to estimate causal treatment effects in context, where we would otherwise have endogeneity that would bias our estimates.

We will look at a few examples of natural experiments in the field of health. For the first example, let’s say we are interested in answering the following question. What are the returns to physician human capital? In other words, does seeing a more skilled or more prestigious physician improve quality of care or health outcomes? This causal effect is difficult to estimate just using observational data because patients generally choose their own physicians. They can choose their own physicians based on their own preferences or their own health needs, in ways that would introduce omitted variable bias or selection, that would bias our estimates.

Doyle, Ewer, and Wagner actually identify a natural experiment that gives them very exogenous variation in physician assignment to patients. They can estimate the causal effect. The setting is actually in a VA hospital. This hospital has affiliations with two medical schools. The residency programs at these two medical schools are very substantial, in terms of their ranking. On top of that, the clinical teams from these two different residency programs actually operate independently. They do not actually interact at all. For decades at this facility, patients were assigned to clinical teams or to teams at these two different medical schools, based on the last digits of their Social Security numbers. If a patient had a Social Security number that ended with an odd number, they would be assigned to one set of clinical teams that belonged to one medical school. If it were even, they would be assigned to the other.

What this produced was “as if” randomization of patients to clinical teams. It was not the case that there was a clinical trial that was run where patients were truly randomly assigned to different teams for this purpose. It was just the case that the circumstances produced what looked like randomization of patients to clinical teams. Here, we have exogenous variation in physician human capital, because patients were essentially randomized to the two different clinical teams.

In our second example, let’s say we are interested in estimating the effects of increasing Medicaid payments for primary care. We would like to know whether that increases primary care visits and reduces hospital and emergency department use. Gruber, Adams, and Newhouse make use of a natural experiment in Tennessee. In 1986, Tennessee increased its Medicaid payments for primary care services. Georgia, which is a neighboring state, had a very similar Medicaid reimbursement system, but did not increase its payments for primary care services. The authors also argue, at least to their knowledge, that there were no other changes in the structure of payment incentives in either state during the study period. The authors argue that there is an exogenous increase in Medicaid payments for primary care in this setting.

Finally, for our third example, let’s think about evaluating the effect of intensive treatment for heart attacks in the elderly. We are interested in whether these treatments actually reduce mortality. The challenge in identifying the causal effect here is that treatment depends on a range of things, which include patient preferences, physician preferences, and patient health. This last one is very important. In general, patients who are sicker, who need the treatment, will receive it. McClellan, McNeil, and Newhouse have an insight. Patients who live closer to hospitals that have the capacity for these intensive treatments are more likely to receive these treatments. It is because, generally when patients have a heart attack, they are taken to the nearest hospital. The distance that a patient lives from a given hospital should be independent of his or her health status. In this case, distance, which is arguably exogenous, affects the probability of receiving intensive treatment for heart attacks.

In each of these three cases, the authors argue that they have exogenous variation in treatment, that induced by external or natural factors. If that is the case, the OLS estimate, beta-1 \_\_\_\_\_ [00:15:57], and this is our estimate of the causal treatment effect, will be unbiased. However, if the “as if” randomization fails to actually produce random assignment of treatment, that is, if we actually do not have a true natural experiment, then the OLS estimator, beta-1 \_\_\_\_\_ [00:16:19], will be biased. This “as if” random assumption is critical to identifying causal treatment effects.

To evaluate the validity of the “as if” random assumption, we can do a few things. First, we can check for differences between the treatment and control groups. If there were randomization, then we would expect to find no differences between the control and treatment group. However, finding no observable differences is unfortunately not sufficient to establish “as if” randomization. What is arguably most important here is that we use contextual knowledge and judgement to assess whether there actually was “as if” randomization, whether the assumption is reasonable.

As similar point came up in the research design lecture, when we discussed how to assess whether there is exogenous variation in the explanatory variable we are interested in, where there is endogeneity or exogeneity. There, there are a few things we can check empirically, but the most important thing is that we use contextual knowledge to be able to reason or justify the assumption we are making.

There is one last point I wanted to make about natural experiments before we move on to estimating causal effects. It is that in general, there are two types of natural experiments. In the first, variation in individual circumstances causes treatment to be “as if” randomly assigned. Having a Social Security number that is odd or even causes a person to be assigned to one team or another. Living in the state of Tennessee instead of Georgia in 1986 or 1987 causes someone to have higher Medicaid payments for primary care. In these cases, we can use OLS to estimate the causal effect, and estimate the average causal treatment effect.

In the second type, variation in individual circumstances only partially determines treatment. An example of this is the third example that we discussed. There, distance affected the probability of receiving intensive heart attack treatment. It did not exactly determine treatment. For patients who lived far, it was not the case that the patients who lived very far from high intensity facilities never received high intensity treatment, or that patients who lived very close always received high intensity treatment. It was just that patients who lived closer had a higher probability of receiving high intensity treatment.

In these cases, we use instrumental variables regression to estimate the causal effect. We will talk more about this in the instrumental variables regression lecture in two weeks, on April 22nd. Today, we are going to focus on the first case, where variation in circumstances determined treatment. We will specifically talk about how we can use difference-in-differences to estimate the average treatment effect.

How do we estimate the treatment effect in natural experiments, where individual circumstances determine treatment? One option is to compare pre and post-treatment outcomes in the treatment groups. We can use panel or repeated cross-sectional data for this. We can specify a regression model that looks something like the following.

We are interested in an outcome variable for Person I and Time T. Our main explanatory variable is the post-binary variable, that equals one, in the time period we are observing for that individual after treatment, post-treatment. Zero is that time period of pre-treatment. Beta-1 here, our beta-1 \_\_\_\_\_ [00:20:52] will estimate our treatment effect. The issue here is that if there are other factors that affect the outcome or treatment, that change during our study period, our estimate of the treatment effect, beta-1 \_\_\_\_\_ [00:21:05], will be biased. That is because beta-1 \_\_\_\_\_ [00:21:09] may capture the effect of these other factors. It will not capture only the effect of treatment.

Another option is to compare post-treatment outcomes between treatment and control groups. We can do that by specifying the following regression model. Here, we are interested in our outcomes for Person I, and our main explanatory variable is a binary treatment variable. That equals one if the person receives treatment, and zero if the person did not receive treatment. Again here, beta-1 \_\_\_\_\_ [00:21:51] will estimate the treatment effect. The issue here is that if there are any differences, in particular unobserved differences, between the two groups, our estimate of the treatment effect, beta-1 \_\_\_\_\_ [00:22:03] will be biased.

This is where difference-in-differences comes in. in difference-in-differences; we compare the change in the pre and post-treatment outcomes across the treatment and control groups. We can do that by specifying the following regression model.

Here, we have a dependent variable, an outcome for Person I at Time T. We have three main explanatory variables. The first is a treatment, a binary variable. The second is a post-treatment binary variable. The third is an interaction of these two binary variables. We interact treatment x post. We will talk a little bit more about this and what this does. Here, beta-3, our estimate beta-3 \_\_\_\_\_ [00:23:02] will estimate the average change in outcomes for those in the treatment group, minus the average change in outcomes for those in the control group. This will give us the average treatment effect in the population that we are studying. We will see why beta-3 \_\_\_\_\_ [00:23:19] is called the difference-in-differences estimator.

This is the same regression model or equation that was on the previous slide, but to simplify notation, Y here stands for outcome and PX here stands for treatment. Again, we have the exact same model. We have our outcome variables that are dependent variables. We have a treatment binary variable. We have a post-binary variable. We have the interaction between the treatment and the post-binary variable.

Let’s think about what this regression model gives us. Let’s first calculate or compute the expected outcome for those in the control group in the pre period. Here, treatment would equal zero and post would equal zero, because we are talking about the control group in the pre-treatment period. We can set those equal to zero and take the expected value of this. This gives us beta \_\_\_\_\_ [00:24:24]. Beta \_\_\_\_\_ [00:24:25] is the expected outcome in the control group in the pre-treatment period.

Now we can do the same for the control group in the post-treatment period. Here, treatment would be zero and post would equal one. This gives us beta \_\_\_\_\_ [00:24:42] plus beta-2. We can do the same for the treatment group in the pre period. Here, we can set treatment equal to one and post equal to zero. This gives us beta \_\_\_\_\_ [00:24:57] plus beta-1, because post is equal to zero. Finally, we can compute the expected outcome for the treatment group in the post period. Here, treatment would equal one and post would also equal one. This gives us beta \_\_\_\_\_ [00:25:17] plus beta-1, plus beta-2 plus beta-3.

Now, let’s focus on the control group. This corresponds to the first two rows. The first row gives us the expected outcome in the control group in the pre period. The second row gives us the expected outcome in the control group in the post period. We can actually compute the difference between outcomes for the control group from the pre to the post period. We can take the second row and subtract the first row. This gives us beta-2. In the control group, there is a change of beta-2 in the outcome variables between the pre and the post period.

We can do the same for the treatment group. We can compare the third and the fourth rows. The fourth row gives us the post-treatment outcomes in the treatment group. We can subtract the pre-treatment outcomes from that same group. That gives us beta-1. That difference there is beta-2 plus beta-3. In the control group, there was a change of beta-2 plus beta-3 in the outcome variable that we are interested in, between the pre and post-treatment period. Each of these is the differences between the post and the pre-treatment period for each of these groups.

Now we can take the difference between these, beta-1 minus beta \_\_\_\_\_ [00:26:52]. This gives us beta-3. Beta-3 is the difference-in-differences. It gives us the difference between treatment and control groups, in the difference between pre and post-treatment period outcomes. Beta-3 is our difference-in-differences. Beta-3 \_\_\_\_\_ [00:27:20] is the difference-in-differences estimator.

Now, let’s see this in practice. To do that, we will return to the second natural experiment example that we discussed. Here, we would like to evaluate the effect of increasing Medicaid payments for primary care in Tennessee in 1986, using Georgia during the same years as a control.

Risha Gidwani: Christine, we have one question here. It is about what types of data you normally use for difference-in-differences. Is it normally panel data or any data where you have two or more observations in time?

Dr. Christine Pal-Chee: Yes. For difference-in-differences, you do need that time dimension. You need observations from multiple people at multiple time periods. You can use panel data, which follows the same people over time, or repeated cross-sectional data, which is basically a different set of people at multiple time periods.

Risha Gidwani: Great. As long as you have two time periods, you are able to employ the difference-in-differences estimator.

Dr. Christine Pal-Chee: Yes, you need at least two time periods. You need that time dimension, because we would like to compare pre and post outcomes. We can use more than two time periods. You can have multiple pre or multiple post time periods also.

Risha Gidwani: Great, thank you.

Dr. Christine Pal-Chee: Returning to this example of the increase in Medicaid payments, we will look at the Gruber, Adams, and Newhouse paper. In that paper, they used difference-in-differences to estimate the treatment effect of increasing Medicaid payments for primary care. There, we would use the same regression model that we just saw. His or her treatment would equal one if a person were a Tennessee enrollee, and zero if the person were a Georgia enrollee. Only those in Tennessee would have been affected by treatment. The post variable would equal one for years after 1986 when this increase was implemented, and zero for years before the increase was implemented. This would be the same in Georgia and Tennessee.

Let’s look at their results. The dependent variables they are interested in are the share of patients or share of enrollees for whom the dominant site of care, where the person received most of their care, was a physician’s office. What share of patients had the physician’s office be their dominant site of care? Let’s look at the difference-in-differences result. We can focus on just the first row, the physician’s offices.

What does this table tell us? Let’s look at the first column. In Tennessee in the pre-treatment period, about 26% of patients or enrollees have the physician’s be their dominant site of care. In the second column, we see that in the post-treatment period, that share increased to 29%. Twenty-nine percent of enrollees have the physician’s office be their dominant site of care. In the third column, we see that this is a difference of 3.5 percentage points. There was an increase of 3.5 percentage points in the share of patients who had physician’s office be their dominant site of care.

Now, let’s look at Georgia. In Georgia during the pre period, 35% of patients had the physician’s office be their dominant site of care. In the post period, only 33% of patients had the physician’s office be their dominant site of care. Actually, in Georgia, there was a decrease of two percentage points in the share of patients who had the physician’s office be their dominant site of care.

Here, the purpose of Georgia as a control group is that Georgia tells us what would have happened in Tennessee in the absence of treatment. Georgia received no treatment and they actually had a decrease in the share of patients who had physician’s offices be their dominant site of care. We can actually factor that in and take the difference-in-differences. We can take Column 3 and subtract Column 6. This gives us the last column, Column 7. This is our difference-in-differences. We see that the difference in differences estimate of the effect of increasing Medicaid payments for primary care services in Tennessee was 5.5-percentage point increase in the share of patients who had physician’s offices as their dominant site of care.

We can focus just on the last column and move down. We can look at the second set of results for outpatient clinics. We see that the increase in Medicaid payments for primary care actually resulted in a decrease of the share of patients who had outpatient clinics be their dominant site of care. That decrease was a 4.1-percentage point decrease. If we look at the third and fourth sets of rows, we see that there was no effect on hospital outpatient department use or emergency room use. It did not look like people were shifting from the hospital setting to the primary care setting.

Risha Gidwani: Christine, are these estimates that you are presenting here predictive probabilities, or are these beta-coefficients?

Dr. Christine Pal-Chee: These are beta-coefficients.

Risha Gidwani: Great, thank you.

Risha Gidwani: We do have a couple of other questions here. One of the questions is about the outcome variable, and whether it needs to be continuous or if we can employ categorical or binary data as well.

Dr. Christine Pal-Chee: The variable does not need to be continuous in order to use difference-in-differences, although the nature of the dependent variable needs to be taken into account when we actually choose what specific sort of method we are going to use to estimate our regression estimates. The difference-in-differences framework remains the same, regardless of how our dependent variable is structured.

Risha Gidwani: Great.

Dr. Christine Pal-Chee: I had mentioned earlier that in this case, in the difference-in-differences framework here, we are using Georgia as a control. The assumption here is that in the absence of treatment, Tennessee would have experienced the same change that Georgia experienced. We will talk a little bit more about this. To do this, I have constructed a very simple figure. I am plotting average outcomes on the Y-axis by Time on the X-axis. Here we have two time periods for simplicity. T1 is a pre-treatment time period and T2 is a post-treatment time period.

Now, let’s look at the different points. Point A plots the average outcomes for the control group in the pre-treatment period. Point B plots the average outcomes for the control group in the post-treatment period. We see that the average outcome actually increased from A to B in the control group. Now, let’s look at Point C. Point C plots the average outcomes in the pre-treatment period for the treatment group. Point D plots the post-treatment outcomes for the treatment group. This is not the actual post-treatment outcome. This is what we observed for the treatment group.

Now, let’s look at Point E. Point E is the post-treatment outcome that we think we would observe in the absence of treatment. This is for the treatment group. We estimate this or guess this based on what we observe in the control group. The assumption here is that in the absence of treatment, we would observe the same change in the treatment group as we do in the control group. Our difference-in-differences here, beta-3, is actually the difference in outcomes at Point B and Point E. It gives us the causal treatment effect, because we believe that treatment causes the treatment group to be at Point B instead of Point E.

There is an assumption here that we make, in order to say this. There are common trends in the two groups, in the control and treatment groups, in the absence of treatment. In the absence of treatment, what happened to the control group would have happened to the treatment group. This underlying assumption is actually very important. We are assuming here that trends in the outcomes would be the same in both treatment and control groups, in the absence of treatment. Difference-in-differences estimates the deviation, which we attribute to treatment, from that common trend.

The difficulty here is that we actually cannot check this. We do not observe what would have happened in the control group in the absence of treatment. That is a counter-fact that we do not observe in real life. What do we do there? One thing we can do is check the pre-treatment trends. We can compare the trends and the outcome variables in the pre-treatment period, across the control and treatment groups. If those trends were the same, it would be more reasonable to make the assumption that the trends would have been the same in the post-treatment period too. Here, we can use data to do this. Again, we would also have to rely on contextual knowledge and judgement to assess whether this assumption is reasonable.

Risha Gidwani: Can you go back one slide please? We have a question. The question is regarding the B and the E. The person is asking whether B and E are then considered the same value.

Dr. Christine Pal-Chee: No, B and E can be different values. The Y-axis here is the average outcomes. E would be the average outcomes in the treatment group, in the post period, had the treatment group followed the same trend or experienced the same change that the control group did. We can see that dotted line between C and E is actually parallel to the solid line between A and B. There, we are saying that the same change that happened in the control group would have happened in the treatment group.

Risha Gidwani: This is also known as the parallel line assumption. Is that correct?

Dr. Christine Pal-Chee: Yes, or the common trends assumption.

Risha Gidwani: Okay. There is also another question about difference-in-differences, and whether difference-in-differences is the same thing as putting in dummy variables for time.

Dr. Christine Pal-Chee: Oftentimes in difference-in-differences, people do include fixed effect, or dummy variables, for time. What is key here for difference-in-differences is that we need the interaction between treatment and time. We need the interaction between treatment and post. Sometimes, people will estimate the treatment effect at multiple time periods, or different time periods. You could interact various time dummy variables with the treatment variables, to see the treatment effect at that time period.

Risha Gidwani: How does the difference-in-differences estimation relate to interrupted times series analysis?

Dr. Christine Pal-Chee: I have to say I am not super or completely familiar with the terminology in interrupted time period analysis. From what I understand, an interrupted time series analysis has some change. There is an interruption that we believe happens. It is the same, and that is similar to what happens in difference-in-differences. We believe that there is some start point in treatment. There is a pre and a post period.

In the basic difference-in-differences model, we are basically calculating the average change between the pre and post-treatment group. It is possible to extend that to estimate other parameters, for example, the change in float, which I believe is a parameter that is often estimated in interrupted time periods. I think that fundamentally, they are doing a very similar task.

Risha Gidwani: When you are evaluating these common trends, what sort of rule of thumb do you use to evaluate whether that assumption has been met or not?

Dr. Christine Pal-Chee: Yes. I had mentioned that it is actually impossible to evaluate this specific assumption, because we never observe what would have otherwise happened in the treatment group. We do not observe the world where the treatment did not happen. What researchers do is evaluate trends in the pre-treatment period. In practice, I think this involves a combination of things. Oftentimes people will plot the average outcome for each control and treatment group at different time periods. There, you would need multiple time periods in order to do this. You would need more than just two time periods. Ideally, you would have multiple pre-treatment time periods.

You could plot the average outcome for each time period. You can visually inspect it, to see if the lines look parallel. You can also empirically test it. You can test the float. The difference in the float is not statistically significant. There are a few tests you can do to check for this.

Risha Gidwani: Great, thank you.

Dr. Christine Pal-Chee: We have talked about the common trends. Now I want to mention a few other things about difference-in-differences. One is the limitations. There are a few main limitations and criticisms of estimating causal effects in natural experiments. One is that the generalizability of the results to contexts other than the one studied may be limited. As an example, we looked at how the authors evaluated the effect of a very specific change in reimbursements in Tennessee in 1986. It is hard to say whether we can generalize those effects to a different context.

Say the Medicaid program in California is thinking about implementing a change that increases primary care payments, perhaps by a different amount, now in 2015. Can we use those results and generalize them to what would happen in California? We might be limited in our ability to generalize there.

The second is that the mechanism for treatment is often unknown. With difference-in-differences, what we estimate is the effect. We estimate the effect and the outcome variables. All we know is what happened. The difference-in-differences estimate does not tell us why it happened. It does not tell us anything about that. There, we would need to rely on \_\_\_\_\_ [00:44:22] our different models, which tests or checks these various behavioral parameters.

Finally, I wanted to make one last note. It is that when using repeated cross-sectional panel data, which is actually often very common in difference-in-differences, oftentimes people do not just use two time periods. We often have multiple time periods available. When using this larger number of time periods, estimated standard errors must account for serial correlation. Because we are observing the same entity, people within the same space or maybe even the same person over time, there is going to be serial correlation in the outcome variables. It is outside the scope of this lecture to go into more details, but I will refer those who are interested to the paper by Bertrand and her co-authors, which is also listed at the end, in the references and resources.

To summarize, natural experiments are situations where external circumstances produce what appears to be randomization. It is as if treatment is randomly assigned. We have exogenous treatment, which allows us to estimate the causal treatment effect. The difference-in-differences is one method of estimating causal treatment effects in natural experiments. In order to estimate the causal effect of treatment, we need exogenous, or “as if” random variation in treatment. We need common underlying trends in the control and treatment groups. We need these two things for an unbiased estimate of the treatment effect.

If we satisfy the necessary assumptions, then the difference-in-differences estimates the average treatment effect. This can actually be very useful when we are interested in evaluating the effect of something. This is particularly useful in program evaluation. I know both of these things have gained a lot more traction and momentum in health services research, in the VA in particular also.

I have listed references and resources at the end, that others can refer to. I have cited the papers I have referred to. I also listed a textbook that I find really useful. There is a nice overview of experiments and causal experiments in the Introduction to Econometrics textbook by Stock and Watson. It offers a very digestible introduction to this.

Now we can open up the floor for questions, and I will take questions until we hit the top of the hour.

Risha Gidwani: Christine, we have a lot of questions for you. The audience is quite interested in this topic. The first question is about the difference-in-differences method, and whether it can be used in retrospective research, in addition to natural experiments.

Dr. Christine Pal-Chee: Yes, we can use difference-in-differences. Everything I have described here, and all of the examples I have cited, are retrospective studies. They use data that has already been collected, or are analyzing situations that have already happened. The key thing there is that we need to satisfy the appropriate assumptions in order for us to make causal statements using difference-in-differences. Again there, what is most important is that we have exogenous or “as if” random variation in treatment, and that we have common underlying trends in the control and treatment groups. In the absence of those two things, we are actually unable to estimate causal effect.

Risha Gidwani: If you have confounding variables, something that you can include in your regression models to adjust for those covariants, correct?

Dr. Christine Pal-Chee: I am sorry? Was there more Risha? I did not mean to interrupt you.

Risha Gidwani: There is actually another question. It is about covariants. Are you able to include covariants in your difference-in-differences regression models, to adjust for potentially confounding factors?

Dr. Christine Pal-Chee: Yes, you definitely can. I omitted them from the lecture today for simplicity. We just focused on the main explanatory variables of interest. Yes, we can include additional control variables or covariants. We can include them if we believe that there are observable differences. The key here is observable. Control variables and covariants do not account for any unobserved differences. We can include them if we believe that, conditional on those covariants or control variables; we have “as if” random assignment for exogenous treatment. Including those variables can also, in some cases, improve our precision. It can result in smaller standard errors, if it actually helps to explain some of the variation in the outcome variables.

Risha Gidwani: The next question does help somewhat with that, whether one can still do a difference-in-differences model if you do not have the same set of cases or controls from your pre-intervention period to your post-intervention period.

Dr. Christine Pal-Chee: Risha, do you know if that means we are observing different people in the control and treatment periods?

Risha Gidwani: I believe that is what the question says.

Dr. Christine Pal-Chee: I am not sure if I understand the question.

Risha Gidwani: I think you did mention before in your lecture that you can look at different cohorts of people.

Dr. Christine Pal-Chee: Yes, you can use repeated cross-sectional data where the cohorts differ in each time period. For example, in the Medicaid payment paper by Gruber, Adams, and Newhouse, theirs was essentially a repeated cross-sectional sample because they did not observe the same people over time, or they did not necessarily observe the same people over time. What we do need is to observe people in Tennessee pre and post.

Risha Gidwani: If you do have two different cohorts, would this be the situation in which you do include covariants to adjust for differences in the pre-intervention cohort versus the post-intervention cohort?

Dr. Christine Pal-Chee: The covariants would actually control for difference. You could include covariants there. The covariants I think about as controlling for differences between the control and treatment groups. These covariants vary with person, so with time they would control for differences across control groups and across pre and post-treatment periods. The important thing there is to think through whether those changes in covariants affect our outcome variables, affect our trends, or affect treatment. The key there is that these things can differ. They can vary, but conditional on those things, we must have exogenous treatment variables.

Rishi Gidwani: Can you go back to the last slide, in which you showed results of the difference-in-differences model? We have a question about that.

Dr. Christine Pal-Chee: Which slide? Was that from the paper?

Risha Gidwani: Yes. I believe it is this one. The question is that the results seem to show the multiple estimates about what is causing the differences. You have estimates for physician’s offices, clinics, outpatient, and emergency room. The question is whether there would be multiple interactions with time in such cases, so that you would interact physician’s office times Time, clinics times Time, possible outpatient department times Time, and emergency room times Time.

Dr. Christine Pal-Chee: In this table physician’s office and clinic, the variables in the left hand column are actually the dependent variables. That is the Y variable on the left hand side. You can think of each set of rows as being a separate regression. In the first set of rows for physician’s office, the dependent variable is the share of enrollees for whom the physician’s office is their primary site of care. In the regression model there, the coefficients would be here on the first row.

The second row of clinics would be a separate regression. That would be a different dependent variable. We have the exact same explanatory variable of interest. We have the same treatment and the same pre and post time period.

Risha Gidwani: In the difference-in-differences model, the only thing you are interacting with time is the intervention.

Dr. Christine Pal-Chee: Time with the intervention, yes.

Risha Gidwani: Okay, great.

Dr. Christine Pal-Chee: Risha, I remember you asking a question about whether these are regression coefficients or predictive probabilities. I think I misunderstood that question. These are from regression, but they have calculated these out so that each of these coefficients or numbers corresponds to the probability for each of those groups in time.

Rishi Gidwani: There is one question about whether there is a significant test for difference-in-differences. I am looking at this output. I am not sure that I see it here. There should be a test of significance associated with your beta-3 coefficient. Is that correct?

Dr. Christine Pal-Chee: Yes. Let’s look at the first row. The difference-in-differences for physician’s office is the treatment effect on the share of patients who are going to the physician’s office for their primary site of care, 5.5 percentage points. The number underneath in the parentheses is the standard error. From those two, we can compute our T statistic and T value. In this here, the difference-in-differences estimate of 5.5 percentage points is statistically significant. That is a 5% significance level.

Risha Gidwani: I do these in \_\_\_\_\_ [00:55:07] and I know that it will give me a P value associated with beta-3, which is your coefficient of interest.

Dr. Christine Pal-Chee: Yes.

Risha Gidwani: Hopefully the statistical program will also show you whether your difference-in-differences term is significant.

Dr. Christine Pal-Chee: Yes. I should be clear that this last column, \_\_\_\_\_ [00:55:22], this corresponds to the regression estimate of beta-3. Each of these is the regression estimate of beta-3. In the curved parentheses are the standard errors for those regression estimates.

Risha Gidwani: We have one question about difference-in-differences design, comparing that to before and after controlled studies using an interaction trend between the study period, the time, and the treatment. Difference-in-differences is essentially pre and post with the control group, correct?

Dr. Christine Pal-Chee: Risha, I am not sure that I understand that question.

Risha Gidwani: The person is asking whether difference-in-differences estimators are similar or different from an analysis where you had a pre/post study with a control group, in which you included an interaction term between time and treatment. From my understanding, I believe these are the same thing. It is just different words being used to describe the same technique.

Dr. Christine Pal-Chee: That is my interpretation too. With difference-in-differences, we have the time/treatment interaction term. That is exactly it. We have the control group, the treatment group, and pre and post-treatment period. We are interested in the coefficient on this interaction term between time and treatment. Here, I called it post. It is an indicator variable that indicates whether the time period is in the post-treatment period.

Risha Gidwani: Great. We have a number of questions here. I do not think we are going to have time to get through all of them. I can propose a couple more to you Christine, but would you be available via email in case folks want to directly contact you with the rest?

Dr. Christine Pal-Chee: Sure. People can go ahead and send in their questions. I think Heidi will have a log of them, that I can get from her afterwards. We will try to follow up the best we can.

Risha Gidwani: We have a couple more questions in the remaining time we have available. Concerning the model that you used for difference-in-differences regression, it could be OLs or a logistic regression. It could be any type of regression model with an outcome that is categorical, continuous, or binary.

Dr. Christine Pal-Chee: Yes, that is true. It is just that the framework remains the same, or the framework that you can overlay over all of these different models.

Risha Gidwani: Okay, great. There is one request to show the references. I do not think it is coming up on your slide. Folks can take a look at those while I pose the last question to you.

Dr. Christine Pal-Chee: Yes. The slides are also available for download. I think there is a link in the email that was sent out, with login information. Here are the references and resources.

Risha: Okay, great. I will end with a question that is a bit complicated, whether one can combine difference-in-differences with propensity scores.

Dr. Christine Pal-Chee: I am thinking about what the purpose would be for doing that. We would want to include a propensity score if we believed that there were systematic differences in whether people got treatment. We would be concerned about endogeneity of treatment. Here, if we are concerned about that, then we actually do not have a natural experiment. We do not have exogenous variation in treatment. The fact that we are thinking about including a propensity score in this, or for this purpose, concerns me a little bit. If we truly do have a natural experiment, then we would not need a propensity score.

That said, I believe that the one way to control for differences in the propensity score is to include the propensity score, or we can control for the covariants or observable characteristics that we believe might influence the propensity to treatment. Those are the same characteristics we use to predict the propensity score. I think that answers the question.

Risha Gidwani: That is very clear. Thank you Christine. As I said, we do have a number of other questions. Hopefully those can be routed to you directly. I want to thank you. This was a fantastic lecture. We very much appreciate all of your insights here about difference-in-differences modeling.

Dr. Christine Pal-Chee: Thanks Risha.

Risha Gidwani: Heidi, I will hand it back over to you.

Dr. Christine Pal-Chee: I wonder if there are some technical issues on Heidi’s end. I believe we are at the top of the hour. We will wrap up here. I believe that when you close out of the webinar, there will be a poll that comes up. We greatly appreciate people’s feedback. This is a new lecture, so we would love feedback for how we can improve this, what worked, and what did not work. Please do take time to fill out that very short survey. We will read it and take it to heart.

Risha Gidwani: Wonderful. Thank you Christine, and thank you to all of our audience members for participating. Heidi tells me that she is indeed having technical difficulties. We will just ask that you click out of the seminar and fill in the poll that Christine requested. Thank you very much.