# Synthetic Control Methods - Introduction

## Transcript video 2

Full resource, see:   
<https://www.ncrm.ac.uk/resources/online/all/?id=20854>

Dr Xingna Zhang: Hello, everyone. My name is Xingna Zhang. I would like to share my learning and knowledge of conducting synthetic control methods in R or RStudio in this tutorial.

We will be using microsynth in R for this tutorial. As you can see from the map on the left-hand side, this intervention was implemented in Seattle. It's a drug market intervention. The aim was to reduce crime. There were 12 pre-intervention periods and four post-intervention periods in the data.

The intervention was applied to 39 blocks. These were our intervention group. We will construct our synthetic control group from the remaining 9,603 blocks. These blocks didn't receive in the intervention. Other data we have in the data set are block level sensors and demographics and incidences of crime reported by the Seattle Police Department.

We have five steps of analysis plan for this data. First we will try our best to understand the data. Secondly, we'll set up the key parameters for the modelling. Thirdly, we will run a basic microsynth model and then we will try to interpret the results. Finally, we will add a bit more advanced techniques to improve our model and results.

As a first step to understand the data, we pulled out all the column names of this data set contained within this R package. We use column names, seattledmi, and to read out all the column names we can see here this data set has 22 columns in total. The first column ID means the unique identifier for each block, time denotes the time periods in our data set. Remember we have 12 pre-intervention time periods and four post-intervention time periods. Intervention, the third column, has two values. Zero means no intervention implemented, one represents intervention implemented and the following few columns are different types of crimes. We will be looking at any crime, column number 12, as our outcome in this tutorial. We will also have some block level statistics in this data set, such as the number of total population, the number of black population, the number of Hispanic population, the number of men aged between 15 and 21, the number of households, the number of families who own their own house, the number of houses headed by females and the number of houses that are rented and the number of vacant houses. Remember all these are block level statistics.

Then use the corr on the left-hand side, we can produce the correlation matrix figure on the right-hand side. Because we’re going to look at any crime as our outcome variable, so we are going to focus on our outcome variable mainly, so we can look at both horizontally and vertically.

Then we notice two other types of crimes. Larceny and burglary are very closely related to any crime. So, in our later stages of analysis we may consider incorporating these two as additional outcome variables, but for the purpose of this tutorial, we will only focus on one outcome variable, any crime in this stage.

We will set up the model as having nine covariates or natural variables such as total population, the number of black or Hispanic population. These are all block level statistics and remember these covariates, they shouldn't be time varying. In other words, they should be stable. They should remain the same during our study period and the match out or our outcome variable is any crime as I mentioned earlier.

To run a basic model, first we need to set up the random seed so we can retry and reproduce our result. We called our first basic model sea1 and the ID variable first, which is spelled out microsynth, that’s the function, then we refer to the name of the data set, the ID variable. Remember, we're doing analysis at the block level. Then we specify the time variable which is the time column in our data set. The intervention variable in our data set is called intervention and we have the starting of the pre-intervention period is one in our data set and the end of pre-intervention period is 12. The end post-intervention period is 16. Our match out variable, in other words, time varying variable is called match out, as I mentioned earlier. It's called any crime. Our matching variable would be the nine covariance that we set up in our last slide. The result variable is the same as a match out variable. It’s called any crime. The test we set up is 95% confidence interval. The number of cores that we will use in our calculation, we’ll set it as one to avoid any errors. And then we will be able to summarise and print out the model result. If we can run this, it will set up the parameters correctly.

[0:09:27]

Our model indeed run successfully. Now let's look at the model summary. Remember we have nine covariates that is stable across all time periods and 12 timepoints before the intervention. We asked the model to match 9 plus 12 which means 21 rows of statistics for us. Then there's one additional component. It’s called intercept, so in total we are expecting to have 22 rows and three columns and then we have a look at the table. We did have what we are expecting. We just want to highlight the first two columns. The first column is called targets which represents the statistics that we observed in the intervention group. Then the second column in the table, we call it weighted control. This one is actually the statistics that we constructed for the synthetic control group.

As we can see, the first and the second column are exactly the same for each and every of these 22 rows. So, well done us. We have successfully constructed a very good synthetic control group.

Now let's interpret the results. When we print out this model result, the first part is the quick summary of the setting up of our model, such as the units, study periods, the constraints and also the number of matches, time variant outcome, time invariant covariates. And most importantly, we would like to know the estimated size of the treatment effect. So, we observed like a minus, almost -22% change and that effect size, if we look at the 95% confidence interval, we will look at the linear lower and linear upper beside the lower and upper 95% confidence interval. Both intervals are below zero. So, we are quite confident that this intervention was successful in reducing crime, at least in our outcome, in reducing any crime.

If we put out the result, both figures show us in quite intuitive ways that any crime did reduce, did decline compared to the synthetic control group.

If we want to check the real balance of our results for them, there are two techniques we can use. The first one is called jackknife, basically by taking out one observation or a group of observations each time and we calibrate the model estimates repeatedly. The second technique is called permutation, which is randomly reshuffling the data. Both techniques will help us to check the robustness of our model results against randomness and the caveats. We can do this very simply in a microsynth package just by simply specifying jack represent jackknife and then we can use a perm represents permutation. Here we will just use the default 250 permutation iterations to rerun our model. We can see the result. The linear estimate stayed the same with our basic model. What is interesting is the jackknife outcome. For the jackknife outcome, the mean estimates stay the same, but the confidence intervals get larger, get wider, and p value gets a bit bigger for permutation. We observed the p value also increased a little compared to the linear estimate technique, but the confidence intervals actually, get smaller for the permutation.

This is the end of our tutorial. If you have any questions, please reach out to NCRM or myself directly for any enquiries. I hope you enjoy learning about synthetic control methods. Thank you very much.

[End of Transcript]

National Centre for Research Methods (NCRM)  
Social Sciences  
Murray Building (Bldg 58)  
University of Southampton  
Southampton SO17 1BJ  
United Kingdom

**Web** www.ncrm.ac.uk   
**Email** info@ncrm.ac.uk  
**Tel** +44 23 8059 4539  
**Twitter** @NCRMUK