# Synthetic Control Methods - Introduction

## Transcript video 1

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 with you my knowledge and learning of the synthetic control methods.

 First, what are synthetic control methods? It is a statistical method used to estimate causal effects or interventions. I'll use the example as shown on the left-hand side to explain this. This map shows the locations of mobile vaccination units that our research team helped to deploy and design during the COVID-19 pandemic. Our intervention or our objective was to help the disadvantaged neighbourhoods and the communities to promote the first dose of COVID-19 vaccine uptake.

 So, to summarise, this is not a randomised clinical trial setting, which for example researchers have very strict control, as in the lab environment. So, it doesn't have a clear-cut treatment or comparable control group.

 In considering this, we choose to use the synthetic control method by constructing a synthetic control group that is comparable to the intervention group or the treatment group so as to draw the causal effects.

 In this light, I compared natural experiments with randomised controlled trials. Randomised controlled trials are a golden standard for making causal inferences in public health and medical research studies. Basically, in randomised controlled trials, researchers have control of when and where to deploy the treatment or intervention and researchers could control the experiment process. But in natural experiments, these are mainly scenarios relevant to policy changes. As you can see, on the left-hand side, the animated map shows the policy changes during the early stages of the pandemic from national lockdowns to different tiers of restriction measures.

 Policy changes can be very dynamic, are most importantly, they are not controlled or deployed by researchers. Researchers do not have control of the process in the setting of natural experiments. It’s more feasible, it is an easier to use synthetic control methods rather than other traditional methods to make causal inferences.

 One of the most popular treatment effects is the average treatment effect. It’s a mathematical formulation can be summarised at the difference in the outcome variable between the intervention group and the synthetic control group, as you can see in the listed equation where one and zero represent the intervention and non-intervention group respectively and why is the outcome variable. If you are interested in the more detailed mathematical formulation, please refer to the link at the right-hand corner at the bottom of this slide.

 I would like to use a few examples for the explanation in simple linear regression based on cross-sectional data which means we only have observations at one timepoint. If we want to, for example, look at the relationship between blood pressure and sex, age and race, we can use this linear regression model. The outcome variable here would be blood pressure and the three features would be age, sex and race. The unit level would be each and every individual. So, this equation will help us to understand the association between the outcome and independent variables. But in making causal inferences, usually we need to use longitudinal data. That means we need observations at multiple timepoints.

 If you see this, the second figure, that means we have multiple, for example, control units like these places or these groups, these individuals without receiving the treatment, the observations of the outcome at various timepoints. If we transpose that right now, the column would become the control units and the role would become different timepoints of the observation. And finally, if we apply weights, usually we need to optimise the weighting to minimise the difference of this synthetic control group between the synthetic control group and the treatment group or treated group or the intervention group, the group that has been applied, the intervention. So, deriving the optimise weights here or collaborating these weights here is the key to construct the synthetic control group.

[0:07:57]

 To summarise the mathematical formulation process, the key to construct the synthetic control group to find the optimised weighting, to minimise the difference between the intervention group and the synthetic control group for the pre-intervention time periods in terms of the independent or the controlling variables.

 I will use a case study, the one I mentioned in the beginning of this course using the COVID-19 mobile vaccination units across the Cheshire and Merseyside region in 2021 to further explain this.

 This slide shows how we designed our study and the map in the middle should be quite familiar to you. I showed that one earlier in our course. So, the idea we had was to match intervention cohort with control cohort with similar trends in vaccine uptake in the pre-intervention time period and we will model the change in weekly uptake after the introduction of the mobile vaccination units.

 You can see in the figure at the right-hand side, the red dashed line at zero represents the timepoint when the intervention was implemented. So, before this intervention was applied, we wanted to develop or we wanted to find the synthetic control group that is perfectly matched that of the intervention group. So, after the intervention any difference in the outcome variables that can be used as the measurement for the treatment effect.

 The key then becomes how to define the intervention group. As you can see in the map on the right-hand side, the red dots represent mobile and static vaccination units that were deployed during our study period and then we categorised any lower super output area with population within radius within one kilometre of the deployed mobile vaccination units as catchment areas of the deployed units. These areas are coloured in yellow. The purple areas are the non-intervention areas from which we are going to construct a synthetic control group.

 I already explained about the exposure which was deployed in the early stages of the COVID-19 and the mobile vaccination units conducted 54 visits in total and visited 37 sites in total. And also, we looked at seven weeks before and three weeks after the exposure. To construct the synthetic control group, we matched the intervention group and the non-intervention group in terms of demographics such as proportion of minority groups’ social economic condition measured by index of multiple deprivation, vaccine access measured by distance to the nearest vaccine sites, and additional neighbourhood characteristics such as population density.

 Our results show the intervention was effective in improving weekly vaccine uptake and that actually led to a 25% increase on average in the uptake within three weeks after the first deployment. As you can see from the figure on the right-hand side, the purple solid line represented the synthetic control group we constructed and the yellow dashed line represents the intervention group. The slightly lighter yellow and purple colour surrounding both lines are the 95% confidence intervals of both groups and red dashed line, the vertical dashed line at zero represented the starting time of the intervention.

 As I mentioned already, we look at seven weeks before and three weeks after the intervention. If we look at the pre-intervention time period, both groups perfectly matched each other which means a synthetic control group was conducted quite successfully and any difference after the intervention between the two groups in terms of the outcome would be the key measures that we can use to measure the average treatment effect.

 Now I would like to introduce a second case study to further explain how we constructed and used synthetic control methods in our research. And this research project was looking at the impacts of tiered restrictions introduced in England during October and December 2020 on COVID-19 cases.

[0:16:08]

 These two maps shows the Tier 2 and Tier 3 areas in October and December 2020 in England. Tier 2 and 3 were high alert and very high alert areas back then. Tier 3 were restrictive compared to Tier 2 in terms of restrictions. More restrictions on social interactions and the hospitality sector.

 These two figures show the result of a number of cases per 100,000 population four weeks before and four weeks after implementing Tier 3 restrictions in October and December 2020 respectively. The solid purple lines represent the synthetic control group and yellow dashed lines represent the intervention group.

 As you can see, it's quite obvious in both time periods additional restrictions in social interactions and the hospitality sector were effective in reducing case rates.

 We further conducted a sensitivity test to check whether our results were robust against spatial spillover effect. Our hypothesis was people in Tier 3 areas, they could have travelled to Tier 2 areas to take advantage of the less restrictive measures over there, such as going for a meal in the restaurant and that would have diluted the intervention effect. We were quite concerned about that, so that's why we conducted this sensitivity test by removing the bordering areas between Tier 2 and Tier 3. Our sensitivity tests showed a very similar result with our main model which means there wasn't a lot of spatial spill over. In fact, in our study, I will just show you one more time the study area of our main result is to confirm.

 Now I would like to quickly summarise the strengths and limitations of synthetic control methods. Synthetic control methods have merits in terms of ethical considerations because researchers usually are going to use natural experiment data. So, researchers, we don't need to consider ethical applications or ethical evaluations because the intervention experiments have been conducted by the government or the politicians.

 In terms of data requirement, you would need to have panel data to use synthetic control methods and you would have to have pre and post-intervention periods in your panel data.

 For example, if you have quite good panel data, but you are struggling to identify a clear-cut pre and post-intervention periods or you are struggling with identifying interventions in your data, then synthetic control methods may not be the right method for you.

 Finally, you will need to have a large and representative donor pool to help you to construct the synthetic control group. There are six methodological assumptions for synthetic control methods.

 First, no anticipation. This is quite easy to understand. For example, if the government plans to implement restrictions on alcoholic drinks, the public know about this before the government start the intervention. The public then can go out of their own way to buy a large amount of alcoholic drinks. That would have diluted your estimate and biased your results. So, be aware of this.

 Secondly, no interference, which is the spillover effect that I mentioned in the second case study.

 Thirdly, common support, the underlying mechanisms of the intervention and control group should have common and similar underlying driving mechanisms, which is to say the independent variables that you put in your model, they should be able to measure and control the intervention and the synthetic control group.

[0:23:05]

 Fourthly, homogeneous treatment effect across units and time. So, that means across different units and across pre-intervention and post-intervention time period. The treatment effect should remain stable.

 Fourthly, the treatment assignment shouldn't be affected by the outcome or any unmeasured confounders.

 Finally, no unmeasured time varying confounders.

 That will be all. Thank you very much. If you have any questions, please reach out to NCRM and myself for any inquiries. Thank you so 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