# An introduction to surveys and longitudinal data: cohort and household panel studies

## Transcript

So here are the topics that I'll cover in this webinar. We'll talk about the types of studies and data that are available and the types of research questions that they can be used to answer.

I'll also highlight some useful resources that will help you find out more about the data that are available and how you can access these data sets.

And we'll also briefly look at how you can maximise the comparability of mental health data in order to conduct cross study research.

The UK has a long history of world leading longitudinal research. Some of the world's longest running longitudinal studies are hosted here in the UK. The studies shown here are just some of the fantastic resources that we have access to.

And these studies have been driving research and influencing policy across many fields for generations. These longitudinal studies allow us to track changes over time, both within individuals and across different generations. And this can help us identify normal developmental trajectories and how development is shaped by risk factors that can be either fixed or time varying.

Given that longitudinal studies typically measure a wide range of variables from different domains, we can control for many extraneous variables, and in some circumstances, depending on the data that are available, make causal inferences about relationships between variables.

So, what types of studies are available to researchers who are interested in mental health? This talk and this webinar will deal with mainly longitudinal studies.

First, household panel surveys. A household panel study is a type of longitudinal research in which a group of households are selected and followed over a period of time, typically several years or more. The purpose of a household panel study is to observe changes in the behaviours, attitudes and characteristics of the households and their members. In a household panel study, data are collected through repeated surveys of the same group of households, usually on an annual or biannual basis. The surveys typically cover a range of topics related to the household, such as their economic status, demographic characteristics, health and wellbeing, and social and community engagement.

The advantages of a household panel study include the ability to track changes in households and their members over time, the ability to examine relationships between variables at different levels, such as individual family, geographic region, and the potential for more in-depth analysis of specific subgroups of interest.

Perhaps the best example of a household panel study in the UK is Understanding Society, which will be covered in-depth by Mark in the next talk.

So, for the rest of this webinar, I will be focusing on several British birth cohort studies. A birth cohort study is a type of longitudinal research that follows a group of individuals born during a particular time period, usually within a single year or a few years.

The individuals in the cohort are usually followed throughout their lives with data collected at various time points to investigate how different factors such as genetic, environmental, social and behavioural factors affect their health and wellbeing. Birth cohort studies can provide valuable information on the development of physical and mental health and on the impact of various exposures and health outcomes. They can also provide insights into the effectiveness of interventions such as public health programmes and policies.

Many of the UK birth cohorts are nationally representative. Therefore, they have the further advantage of being generalisable to the UK population. Some of the most well-known and utilised birth cohorts in the UK include the National Survey of Health and Development, which began in 1946 and is the UK's longest running birth cohort study. The study follows a randomly selected sample of 5,362 people born in England, Wales and Scotland during a single week in March 1946. This was approximately one-third of all babies born that week. These individuals have been followed up 25 times through a combination of paper questionnaires and visits by health visitors or nurses, either in the home or in a clinic. The most recent assessment of this cohort occurred during the COVID-19 pandemic when participants were aged approximately 74 years.

The 1958 National Child Development Study, or the NCDS, is following the lives of an initial 17,415 people born in England, Scotland and Wales in a single week of 1958. With 11 follow-ups, the study has collected information on their physical and educational development, economic circumstances, employment, family life, health, behaviour, wellbeing, social participation and attitudes.

Similarly, the 1970 British Cohort Study is following the lives of around 17,000 people born in England, Scotland and Wales in a single week of 1970.

Next Steps, previously known as the Longitudinal Study of Young People in England, follows the lives of around 16,000 people born in England between 1989 and 1990. Unlike the other studies, Next Steps did not commence at birth, rather it began in 2004 when the cohort members were aged 14. Next Steps has particularly rich data on education. For example, it can be linked to the national pupil database records.

And the Millennium Cohort Study, known as the Child of the New Century to cohort members and their families, is following the lives of around 19,000 young people born across England, Scotland, Wales and Northern Ireland between 2000 and 2002. Like all of the above studies, the most recent sweep of data occurred during the COVID pandemic studying how individuals coped during that unprecedented site.

What type of information are routinely collected and available in these studies? So obviously most people here are interested in mental health data. This is something that is routinely collected throughout the life course in these studies for both the cohort members themselves, but also frequently for parents and other family members. Mental health problems are typically assessed using well-established screeners, usually self-report questionnaires in adults and parent and teacher proxy reports in children.

In terms of other variables that may be related to mental health, these cohorts are designed to be used in many disciplines. Therefore, they capture a range of relevant information across many domains, such as demographic characteristics, socioeconomic status, physical health, cognitive ability and education, and also social and community factors.

Most of this information is freely available, however, more sensitive data such as linked administrative records and genetic data can be obtained via secure access routes.

In terms of how this data can be used and what research questions we can answer, the answer is well, we can answer an endless amount of research questions really. If we narrow our focus on mental health, many studies look at time trends. For instance, how does mental health develop across the life course? Are there universal periods of high risk? And is this consistent across different cohorts?

Other studies may be interested in looking at how certain risk factors are associated with the development of mental ill health, either at different stages of the life course or across different cohorts.

Really, the limits of the research questions we can answer are determined by what you're interested in and what data are available. However, when we want to explore the data that are out there we are faced with two issues.

First, we need to find out what studies are out there and have the necessary data to answer our research question.

Second, once we've identified the studies that can do this, we need to find out what specific variables are available in these studies, when they were measured and how they were operationalised.

How do you go about sourcing this information? Each individual study has a website, often with detailed cohort profiles and technical manuals. However, routing through these can be a time-consuming and imprecise process.

Thankfully, a number of organisations and initiatives exist to help with data discoverability, in other words, making it easy for researchers to find out what studies are out there and what information is gathered in each study.

One example is Closer, the UK's home for longitudinal research. This consortium provide a searchable engine, Discovery, that allows you to browse the content of a large number of UK based longitudinal studies. It includes a simple search function where you can enter the construct you're interested in and it will provide you with a list of variables. You can then filter these by things like study or particular life stage. To find out more about the value of longitudinal studies, have a look at the Closer website.

For those who are specifically interested in mental health data, I recommend using the Catalogue of Mental Health Measures, a tool that's been developed by Professor Louise Arseneault and her team at King’s College London. This tool covers over 50 longitudinal studies and its easy to use interface allows researchers to quickly get to grips with the mental health data that are available in each cohort.

Once you have identified your relevant studies and variables, you will no doubt be keen to access this data. And while there's no one stop shop for UK longitudinal data, at least not yet, many of the studies discussed today can be downloaded via the UK Data Service. This provides researchers, government analysts and charities access to high quality resources free of charge. On the UK Data Service website, researchers can browse a range of metadata and download the data they need quickly and easily. Along with full sweeps from individual studies, a range of linked and harmonised data sets for cross cohort comparisons are also available.

It is worth noting, however, that not all data can be downloaded directly. Sensitive information such as linked administrative records, for instance, from health or education databases, these often require secure access. And further information about this can be found on the UKDS website.

For other studies, bespoke applications must be made to the host institutions. This usually involves submitting a research proposal with the main aim of avoiding overlap between separate projects and research groups. Also, depending on the cohort and question, a fee may be required to access the data.

Increasingly, researchers look to combine data from multiple birth cohorts to study how certain phenomena are changing across different generations. When conducting cross cohort studies of mental health, researchers are often confronted with measurement problems, mainly the fact that different questionnaires were used to assess mental health depending on when the assessments occurred.

For instance, child mental health in the 1946 cohort was measured with a precursor to the Rutter Behaviour Scales, whereas in more recent cohorts the Strengths and Difficulties Questionnaire is the most common instrument used.

Mental health screeners such as these often differ in a range of features from the content of specific questions to the number of response options.

In order to help researchers make the most of existing data, I'm currently leading a project which is funded by the Wellcome Trust in which we're developing an online tool that uses machine learning, specifically natural language processing, to match and quantify the similarity of questions from different questionnaires based on their semantic content.

This tool, Harmony, will allow researchers to identify subsets of questions from different instruments that can be used to make direct comparisons across studies.

For more information on Harmony, visit our project website or Open Science Framework page.

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