University of BRISTOL
Jean Golding Institute

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

# USING SECONDARY DATA IN RESEARCH

February 7, 2018

# Presentation summaries and speaker biographies

## Using Business Data for Research

### Professor Helen Simpson, University of Bristol

The talk will outline how anonymised ONS business data has been used in economics research which influences government policy. These data, typically collected for administrative purposes or for the production of National Statistics, have been made available to academic researchers via the UK Data Service Secure Lab. I will summarise the types of data available, using examples of research on firm location decisions, productivity, innovation and foreign direct investment.

I will also outline a current research project on the impact of High Speed Rail 1, which will link ONS business data with spatial data on rush-hour commuting times into London, Census data and Land Registry data on housing transactions. The project will examine the impact of this large-scale transport infrastructure investment on house prices, population growth, changes in demographics and business location decisions.

The talk will also outline the ESRC Secondary Data Analysis Initiative, which provides grants for research which uses the datasets discussed at this workshop.

### Biography

Helen is a Professor of Economics at the University of Bristol. She is a member of the ESRC Grant Assessment Panel for the Secondary Data Analysis Initiative. She is a Programme Director at the Oxford University Centre for Business Taxation and a Research Affiliate of CEPR. She has acted as an Academic Panel member of the What Works Centre for Local Economic Growth, as an Academic Associate of the HM Treasury Productivity Team and as an Academic Expert for DG Research, European Commission. She was previously Director of Productivity and Innovation Research at the Institute for Fiscal Studies. Her research covers firm location decisions, productivity, innovation and foreign direct investment.

## The use of a statistical analysis assistant for secondary data analysis

### Professor William Browne, University of Bristol

In our work at the Centre for Multilevel Modelling we have developed several statistical software packages over the years designed to give researchers access to cutting-edge statistical tools that fit models that realistically account for the complex structures that often exist in real data, in particular, in terms of hierarchies or multiple level of clustering. Our approach has been to produce user friendly software with lots of accompanying training materials so that we can bridge the gap between the background knowledge of the applied researcher and the complexity of the statistics required. In our latest software package Stat-JR, we have taken this approach a step further and developed a Statistical Analysis Assistant (SAA) interface, that takes a researcher's dataset and asks a few questions about their data before attempting to produce a complete analysis and report that aims to both assist the researcher through the modelling process, but also explain to the user what it is doing so that they are taught more about statistical analysis. In this talk, we will describe creating the Stat-JR software, show some examples of the software in action and mention a sister project looking at related tools in Stat-JR for teachers of quantitative methods.

## Biography

Professor William (Bill) Browne has been a Professor in Bristol since 2007 when he joined the School of Veterinary Science in Langford as Professor of Biostatistics moving from the Mathematics department at the University of Nottingham. His research is in statistical methodology, statistical software and applications of statistics to several disciplines, in particular education, veterinary science and animal behaviour. He is very much an interdisciplinary researcher having worked both in mathematics departments and applied schools. After 7 happy years out at Langford, he moved to his current post as Professor of Statistics in the School of Education in Bristol in 2014 and he co-directs the Centre for Multilevel Modelling, of which he has been a member since he was a postdoc when it was based at the Institute of Education, London in 1998! He also spent 2 years from 2015-2017 setting up the University of Bristol institute for data science, the Jean Golding Institute, as its founding director.

# Children growing up in the care of relatives in the UK: Establishing the prevalence and characteristics through analyses of Census microdata

## Dinithi Wijedasa, University of Bristol

The family environment that children grow up in can greatly influence their outcomes in terms of development and wellbeing. Although most children in the UK still live with at least one parent, some are unable to live with their parents due to reasons such as neglect or abuse by parents, parental sickness or death, parents being in prison, family circumstances such as poverty, parental drug and alcohol abuse or mental health issues. Although the majority of these children live with relatives or friends, there is no legal obligation placed on the families to inform authorities of these care arrangements. Therefore, most children and families are not known to the authorities and thus do not receive specific support from the State. This presentation provides the most current estimates of the number and characteristics of the children growing up with relatives in the UK, which were established through analyses of secure microdata from the 2011 Census.

## Biography

Dinithi is a Research Fellow at the Hadley Centre for Adoption and Foster Care at the University of Bristol with a background in developmental psychology, secondary data analyses and statistics.

She has over 11 years of experience in conducting studies on outcomes of children in substitute placements such as foster, adoptive and kinship care and of management and analyses of longitudinal, sensitive and large datasets including Census secure microdata, longitudinal survey data (LSYPE) and the longitudinal national administrative data on children in care (SSDA903 data). Dinithi was one of the first academic researchers in the UK to be given access to national-level, longitudinal administrative datasets on children in care in England and Wales. She is also an ONS approved researcher and a current recipient of an ESRC Future Research Leaders Award (ES/K008587/1) (http://gtr.rcuk.ac.uk/projects?ref=ES%2FK008587%2F1).

# Introduction to the UK Data Service

## Dr Sarah King-Hele, UK Data Service

The UK Data Service provides access to a wide range of quantitative and qualitative datasets for social sciences research, including large scale survey datasets such as the British Social Attitudes Survey, data from longitudinal studies and the UK censuses, international databases such as OECD statistics and interview transcripts and other qualitative datasets. This talk covers the types of data we hold, how researchers can access the data and the support we offer to users including training courses and an email helpdesk.

## Biography

Sarah's background is in survey research. Based at the University of Manchester, she has worked for the UK Data Service supporting and training users of large-scale survey data since 2010.

## One in a Million': A primary care consultations archive

### Dr Rebecca Barnes, Bristol Medical School

Around one million primary care consultations happen in England every day. Despite this, much of what happens in these visits remains a "black box". Our aim was to test the feasibility of creating an archive of video-recorded consultations and linked data with consent for reuse. This was a cross-sectional study in 12 general practices. Two general practitioners (GPs) from each practice were invited to video-record up to 20 consecutive adult patients over one to two days. Patient questionnaires were self-administered immediately pre- and post-consultation and GPs filled out a checklist after each recording and a questionnaire. A follow-up questionnaire was sent to patients at 10 days, and data about subsequent related consultations were collected from medical records at three months. 421 (87%) of 485 patients approached were eligible and 334 (79%) consented to participate. Between July 2014-April 2015, 327 consultations with 23 GPs were successfully recorded. 300 patients (89%) consented to use by other researchers, subject to further NHS ethical approval. Most patients were therefore willing to allow their consultations to be video-recorded, and with very few exceptions, to allow recordings and linked data to be stored in a data repository for future use for research and training.

#### Biography

Rebecca K. Barnes is a Senior Research Fellow in Applied Conversation Analysis (CA) at the University of Bristol. She is also Data Steward for the One in a Million primary care consultations archive. Her main area of expertise is the study of interaction between patients and healthcare providers with the aim of improving communication. She specialises in the application of CA methods. She has studied routine consultations in primary care for a variety of projects including how sickness certification is negotiated, how drug and self-care treatment recommendations are initiated and responded to, and the delivery and uptake of safety-netting advice to caregivers in paediatric consultations. A number of these projects have combined CA-grounded formal coding with quantitative methods to test associations between communication practices and other variables. She has also pioneered using CA methods to inform provider training and the assessment of fidelity in trials of talk-based interventions.

## Qualitative secondary data: Spotlight on data from the secret lives of 4-year olds

### Dr Debbie Watson, University of Bristol & Hannah Brana-Martin, University of Bristol

The Secret Lives dataset is a large and unique collection of young children's free talk, acquired from the RDF & Channel 4 television series, 'Secret Life of 4 (and 5) Year olds'. The data is of huge potential research interest for those within areas such as childhood studies, psychology and education and there is no other data set available like this. In this session we will describe the content, formats, volumes and associated metadata, explaining the process we have established to archive this dataset to make it accessible for research. We will cover themes, challenges, breakthroughs and learning gained throughout the process. We will also discuss the sensitive issues around ethics and access for this type of data and showcase some examples of the content (audio only).

#### Biographies

**Debbie Watson** is a Reader in Childhood Studies at the University of Bristol and is the principal investigator of this project. She has extensive experience in research with children and families, research ethics, childhood wellbeing, identities and children's play and creativity.

**Hannah Brana-Martin** is a Research Associate and is currently in charge of archiving the Secret Lives data. The main aim of the study is to make the data accessible for research in childhood studies, education and beyond, in line with university ethics and data access protocols. She previously worked as an archivist for the BBC.

## An introduction to the ALSPAC resource: a multi-generation study spanning a quarter of a century

### Dr Alison Teyhan, University of Bristol

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a world-leading birth cohort study based at the University of Bristol. ALSPAC has been tracking the lives of thousands of participants for over 25 years. This has generated a wealth of data - from questionnaires, clinic visits, biological samples, and through record-linkage to administrative and health data. Alison will give an overview of ALSPAC, the data available, and how they can be accessed. She will also highlight a few of the study's key findings to date.

### Biography

Alison is a Senior Research Associate in epidemiology who has worked for ALSPAC for over 6 years. Most of her research focuses on child and adolescent health, and particular interests include the determinants and consequences of childhood adversity, contextual influences on health, and the use of administrative data in epidemiology. As well as conducting her own research, she has experience of providing datasets and advice to external collaborators who wish to use ALSPAC data. Since 2014 she has been based within the ALSPAC data linkage team, where her role is to work on a series of exemplar projects to demonstrate the value of linking administrative data to cohort data. Prior to joining ALSPAC, she was a Research Assistant and PhD student at the MRC/CSO Social and Public Health Sciences Unit in Glasgow.

## Trajectories of depressive symptoms from childhood to young adulthood in ALSPAC: Methods and predictors

### Alex Kwong, University of Bristol

Depression is a common mental illness thought to affect more than 300 million people. Evidence suggests that those who experience depression early in life have a higher chance of relapsing. Research has therefore focused on childhood and adolescent depression as a potentially modifiable risk factor for later adulthood depression. The aetiology of depression is complex with many social, biological and genetic risk factors. In this talk, I highlight some of the key predictors for trajectories of depressive symptoms from childhood to young adulthood and the various methods that are available to aid researchers. Specifically, I will show that gender, neighbourhood deprivation and genetics are all associated with varying trajectories of depressive symptoms.

### Biography

Alex Kwong is a 2nd year PhD student from the University of Bristol, based across Geographical Sciences, the Integrative Epidemiology Unit and the Centre for Multilevel Modelling. His PhD explores social and genetic contributions to psychiatric disorders. He is supervised by Dr. David Manley, Prof. Nicholas Timpson, Dr. George Leckie, Dr. Evie Stergiakouli and Dr. Oliver Davis. His PhD is funded by the ESRC on an Advanced Quantitative Methods studentship.

## Using change in depressive symptoms over time to identify latent classes of ALSPAC children

### Dr. José López-López, University of Bristol

Mental health and educational achievement are important aspects of child development, as they play a key role in future employment, health, and social functioning worldwide. Thus, it is important to examine trends over time and to explore potential associations between both areas throughout adolescence.

Data: The ALSPAC resource includes the Short Mood and Feelings Questionnaire (SMFQ), a depression inventory that was completed by participants at nine occasions between 10.65 and 23.8 years, with one or more self-reports available from 9398 participants. Educational records, including the General Certificate of Secondary Education (GCSE), are also available through linked data.

Analyses: We used this information to fit growth mixture models, which enabled us to identify latent classes of individuals based on their change in SMFQ scores over time. Our analysis strategy followed three main stages: (1) we explored different trajectory shapes, (2) we compared models with different numbers of classes, and (3) we focused on one of these models and examined the association between the resulting latent classes and the GCSEs. We will present some preliminary findings and discuss their implications.

### Biography

Dr. José López-López is a methodologist interested in the optimization and application of statistical methods, especially in the areas of mental health and education. He currently leads an ESRC SDAI grant intended to explore the association between depression and educational and early employment outcomes in UK adolescents and young adults, using ALSPAC data.

## Introduction to CLOSER (Cohort & Longitudinal Studies Enhancement Resources)

### Professor Alison Park, Director of CLOSER, UCL Institute of Education

The UK is home to the world's largest and longest-running longitudinal studies. Following the lives of generations of Britons, these rich resources have been used for decades to answer some of the most pressing questions facing our society. Inequality, social mobility, parenting, children's outcomes, health, education, employment, ageing – the possible areas for investigation are endless. These resources are increasing in value year on year – new survey data is collected from participants, and older data is repaired using new methods. Some of the most exciting developments are in resources for cross-study comparisons. Studies in the CLOSER consortium are leading efforts to harmonise data across studies, link survey responses to administrative records, and – crucially – enhance the discoverability of meta data through a new search platform. This session will provide delegates with an overview of CLOSER and our work within the longitudinal research community. It will also highlight the value of longitudinal studies to academic and government activity.

### Biography

Professor Alison Park is Director of CLOSER (Cohort and Longitudinal Studies Enhancement Resources) at the UCL Institute of Education. CLOSER brings together eight leading longitudinal studies, the British Library and the UK Data Service to share resources and expertise, provide training and stimulate longitudinal research. Its research activity has focused particularly on data harmonisation and data linkage. CLOSER is funded by the ESRC and MRC.

Before joining UCL Institute of Education, Alison led a research team at NatCen Social Research where she designed, resourced and implemented a range of commissioned and grant-funded research studies. Alison is a Fellow of the Academy of Social Sciences.

## There and Back Again: Using national insurance data to estimate commuting costs

### Isabel Stockton, University of Bristol

How much of a wage cut would you accept to reduce your commute by one kilometre? In joint work with Annette Bergemann and Stephan Brunow, we use national insurance data from Germany including daily information on job spells, wages and commuting distances to estimate this parameter. We explore how it varies by gender, parenthood and other factors and find that women's marginal willingness to pay as a share of their wage is 40% higher than men's and jumps up further after the birth of their first child. The large sample size and panel structure allows us to account for unobserved individual differences. The database is maintained by IAB Institute for Employment Research which provides anonymised Scientific Use Files and remote data access to researchers, who can also visit one of their locations to work with the administrative data or one of their linked surveys. The institute's Research Data Centre (http://fdz.iab.de/en.aspx) provides detailed information on the available datasets and access options.

## Biography

Isabel is an economist currently in the third year of her PhD at the University of Bristol. She first worked with German national insurance data during her Master's degree at the University of Mannheim. In her research, she uses a range of secondary data sources to study the way women trade off wages and job attributes such as commuting distance and schedule autonomy.

# How ONS supports statistical research using unpublished data


## Nick O'Donnell, Head of Research Data Access, Office for National Statistics


As set out in the UK Statistics Authority's Better Statistics, Better Decisions strategy for official statistics, ONS aims to provide high quality statistics, analysis and advice on the UK economy and society to support democratic debate and improve decision making.

However, there are times when it is not possible to openly publish data because of the risk of disclosing information that breaches person (individual or business) confidentiality. ONS permits access to these detailed data for the purpose of statistical research that delivers a public benefit. The Approved Researcher scheme is one such way in which ONS grants access to unpublished data as permitted by the Statistics and Registration Service Act 2007 (SRSA), to over 1,500 researchers. This presentation will explain more about how and why ONS grants access to these data and the arrangements in place to protect data confidentiality. It will also expand on future priorities for ONS research data access.

## Biography

Nick heads up the Research Data Access team in the Office for National Statistics (ONS). Each year, the team accredits over 400 research projects accessing unpublished ONS data to deliver a public benefit to the UK. They work closely with researchers and sponsoring organisations, to help demonstrate the impact of the research on policy and/or decision making. Nick joined ONS in 2009 as Head of Stakeholder Engagement for the 2011 Census of England and Wales. Prior to joining ONS, Nick worked in a range of policy roles across government, including better regulation, Promotion of UK oil and gas service businesses, health and safety enforcement and local government regulation.

Away from work, Nick enjoys mountain biking, walking, tennis and travelling around South America with his wife. He's yet to meet Paddington.

# Wage inequality and returns to workplace training for male and female employees: A comparison between Germany and England


## Rossella Icardi and Anna Hagglund, University of Bath


Whilst there is evidence that workplace training may have a positive effect on wages, whether training returns differ by gender has been under researched. This study aims to explore whether workplace training returns differ between men and women. In addition, it investigates whether they vary across the wage distribution.

## Biographies

**Rossella Icardi** is a Research Associate within the NEWFAMSTRAT project at the University of Bath. She has a background in Economics and Social Policy. Her research interests include social and educational inequalities, gender differences and quantitative methods. She has experience in conducting analyses of large administrative and datasets and complex surveys using a wide range of statistical techniques.

**Anna Erika Hägglundlund** is a Research Associate in the NEWFAMSTRAT project at the University of Bath. Her research interests include life course research, social stratification, and gender inequality from a cross-nationally comparative perspective. She has worked with several large-scale cross-sectional and longitudinal surveys, as well as register-based data, and is familiar with several longitudinal and multilevel modeling techniques.