Developments in official population statistics

Strand organiser: Phil Humby (Office for National Statistics)

13:30 – 15:00 Monday 11 September: Developments in official population statistics 1

The release of Census 2021 data for England and Wales: Meeting the varied requirements of our users
Sarah Garlick and Emily Green - Office for National Statistics

Data from Census 2021 provides the most detailed and accurate snapshot of society in England and Wales. The cost and effort of taking a census is only worthwhile if we can provide the results to users in a way that helps to support and inform the decisions they need to make. We recognise that our users have wide-ranging needs for census data. Therefore, it was essential that we designed an ambitious plan for the dissemination of Census 2021 data, which has led to us providing data and analysis through a variety of products and flexible formats. In this presentation, we will describe our innovative approach to the release of Census 2021 data and the products we are providing to help users interpret and contextualise those data. We will outline our Census 2021 outputs release schedule, in which we moved from topic summaries focussed on univariate data, to providing more complex cross-cutting data through the "Create a Custom Dataset" functionality on the ONS website. During the release schedule, we have also been providing detailed analysis and insights through bulletins, articles, and interactive products. In this way, we will show how we have ensured that all users of census data are able to gain rich and detailed insights into their communities.

Using administrative data to assist estimation for Scotland's Census 2022
David Rowley and Tom Macintyre

In Spring 2022 National Records of Scotland (NRS) collected information for Scotland's Census, the official count of people and households in Scotland. Census collection is complemented by the Census Coverage Survey (CCS), data from third parties (administrative data), and other statistical methodological work, to ensure high quality census outputs. Following appropriate governance processes, a number of administrative datasets were collected, processed and linked in the National Safe Haven. The datasets included information on health, electoral register, vital events registrations and education. The methodology used several linking variables so data could be linked, even without exact agreement between records. Records from across the data sources were resolved into individuals using these links. Strength of evidence calculations then indicated which individuals we were most sure were captured at the correct address on census day in the administrative data. This linking exercise created a population spine, which considered alongside the Census, CCS and other administrative data can be used to assist census estimation – the statistical process of making sure census outputs reflect the whole Scottish population at a national and local level. Census first outputs will be published on 14 September 2023, so the presentation will the discuss methodology. The census is of national strategic importance, being fundamental to resource allocation between UK countries, and across Scotland. Government, councils, the NHS, and public, private and third-sector users rely on census outputs.

Email: david.rowley@nrscotland.gov.uk

Intruder testing for Census 2021 England and Wales – checking the balance between risk and utility
Samantha Trace and Dominic Nelson - ONS

A fundamental legal requirement and ethical duty of Census 2021 is that individuals will not be identifiable in the data. A new challenge is that to make data easily available, a Create Your Own Dataset system will be used, allowing users to choose geography, variables, and classifications. To ensure that this Create Your Own Dataset system would still present low disclosure risk, three main methods are employed, targeted record swapping, cell key perturbation, and automated suppression rules. These measures are intended to make identifying records in the data difficult and doubtful, without damaging utility. In these times where there are such wide-ranging and detailed sources of publicly available information that might be used to make identifications alongside outputs from the Build-your-own system, we assessed the real-life risk level with an intruder test. The intruders used the

first version of the Create Your Own Dataset, loaded with post-swapping Census 2021 data for usual residents, which automatically perturbed and applied suppression rules as they requested the data. They also had internet access to help make claims. Claims were checked against the census database and levels of accuracy and confidence were recorded. Intruders were also surveyed on ease of use of the system and making claims. Our paper will discuss the challenges, explore the results and comment on the learnings from this exercise.

Email: samantha.trace@ons.gov.uk

Evaluation of the census data of 2002 and 2013 for Benin
Savia Coimbra Porto Santos - Cedeplar/UFMG, Sedami Deo - Gratias, Emmanuel Akossinou - Cedeplar/UFMG, Raquel Lessa Alves - Cedeplar/UFMG, Tânia Evelina Samuel Buchir - Cedeplar/UFMG

Benin went from a population of 7 million in 2000 to 13.35 million in 2022. This is in a scenario of lower mortality rates and, even though fertility rates are declining, these are still high. The challenges of evaluating the most accurate rates come from errors in the data. To evaluate the data from 2002 and 2013 Census in Benin, IPUMS data was collected for Benin. From the descriptive analysis, the age pyramid, sex ratio and intercensal survival ratio, there is an indication of underemuration for the population from 0 to 4 year. To find and adjustment for underenumeration for the age group 0 to 4 year, two techniques were used. The first was using back projection of births using census data from 2013. The second, was using back projections for the age group 10 to 14 years in 2013. Both results show a bigger population on the age group 0 to 4 year than observed in both Census years indicating the need to use an adjustment factor.

Email: saviacpsantos@cedeplar.ufmg.br

17:30 – 19:00 Tuesday 12 September: Developments in official population statistics 2

Methodological challenges in transforming population and migration statistics
Katie O'Farrell - Office for National Statistics

An increasing use of administrative data sits at the heart of the transformation of population and migration statistics at the Office for National Statistics (ONS) – and inherent to that, sits the methodological challenge of using such data for official statistics. The Methodology and Quality Directorate in ONS supports transformation by developing innovative methods in multiple fields including population stock estimation; the potential for machine learning methods for producing population counts; quantifying uncertainty in admin-based migration estimates; Bayesian accounting methods for reconciling stocks and flows; modelling mortality projections, and preliminary estimates of population mobility. With a focus on quality of these outputs in the context of official statistics, in particular understanding and quantifying accuracy, the use of administrative data is an exciting challenge. Much of our work requires the development of entirely new and complex methods; the application of a 'fail fast, fail often' approach maximises opportunities for scrutiny and external expertise and input. This presentation will provide a holistic overview of our work, drawing out common themes and challenges, with subsequent more detailed presentations on specific methods.

Email: Katie.O'Farrell@ons.gov.uk

Transformation of the UK Labour Force Survey through 2023
James Harris - Office for National Statistics

The Labour Force Survey and Annual Population Survey are a fundamental source of economic, labour market and demographic information about the population. They are used by central and local government, academics, thinktanks, and many others to derive insight into peoples' life circumstances and to help deliver effective policies to improve peoples' outcomes. Over the course of 2023 ONS is undertaking the final stages of transforming the LFS to improve its quality, design and functionality. We would like to inform the BSPS audience of the changes being made to the survey, the improvements this should mean to the statistical information they can use, and the impact the transformed survey may have on their analysis and outputs. We intend to provide details about the design and contents of the transformed survey, new approaches being used in data collection and methodology,

insights into the current performance of the survey and the planned timeline for when these changes will be implemented.

Email: james.p.harris@ons.gov.uk

A Longitudinal Population Dataset for England & Wales
Nicky Rogers and Dr Louisa Blackwell - Office for National Statistics

At the Office for National Statistics (ONS), we are exploring the feasibility of an anonymised person-level longitudinal data source for England and Wales, based on Census 2021 and then updated each year to reflect population change (births, deaths and migration); we call this the Longitudinal Population Dataset (LPD).
This presentation will set out our vision for the design of the LPD and how we plan to realise this through a proof of concept phase. The LPD forms part of the ONS' population and social statistics transformation programme which aims to provide the best insights on population, migration and society using a range of data sources. We will present the LPD design and data sources, the maintenance regime to roll the population forward, and how the LPD will support satellite cohort studies.
The LPD will support longitudinal and cross-sectional analysis; it will also support cohort studies (satellite cohorts) based on population samples drawn from the LPD. Our ambition is that the LPD supports satellite cohorts which will become statistical research resources for use by ONS accredited researchers in the ONS Secure Research Service environment (SRS) and the Integrated Data Service (IDS). Access to satellite cohorts will be governed in a similar way to the ONS Longitudinal Study (LS) and SRS.

Email: nicola.rogers@ons.gov.uk

The Future of Small Area Population Statistics, ten years later
Mark Fransham - University of Oxford

In October 2013 an event at the Royal Geographical Society called "The Future of Small Area Population Statistics" was held to discuss the Beyond 2011 programme's proposal to replace the decennial Census with an administrative data system. The presenters, who included this speaker and (more importantly) the future National Statistician Professor Sir Ian Diamond, were broadly supportive of proposals to exploit administrative data but sceptical that it could meet user requirements for statistics at local authority area and below. Amid another consultation on the future of the Census, I reflect on ten years advising the Office for National Statistics on the development of population statistics. I consider how well an administrative data system could cover the full range of small area statistics currently provided by the Census, focusing on three statistical outputs of interest to research and policy – housing tenure statistics, population by age and sex, and measures of poverty and deprivation. I also touch on the challenges associated with relying on data sources that are not directly controlled by the Office for National Statistics. In the light of this, I evaluate whether the last ten years has brought us closer to an administrative data system that can meet user needs for accurate, precise and timely population and housing statistics at subnational level.

Email: mark.fransham@spi.ox.ac.uk

09:00 – 11:00 Wednesday 13 September: Developments in official population statistics: Administrative data

Developing supporting information for users of administrative data - ADR Scotland's approach
Cecilia Macintyre, Scottish Government

ADR Scotland is a partnership between Scottish Government's Data for Research Unit and the Scottish Centre for Administrative Data Research (SCADR). Together they are transforming how public sector data in Scotland is curated, accessed and explored, so it can deliver its full potential. By linking administrative data together and conducting research , it can help address complex social and environmental challenges, filling evidence gaps for policy and decision makers to systematically improve the lives of the people of Scotland. ADR Scotland holds linkable ready datasets in a secure environment on a long-term basis. This reduces the time and effort required by

data controllers, data processors and researchers to provide and access data. ADR Scotland intends for researchers to be able to use comprehensive metadata to identify suitable datasets and variables that can answer critical questions. This also offers the potential to facilitate researcher feedback with data controllers, helping the latter to better understand their data and consider steps around data quality. Scottish Government has provided information to support researchers including metadata, a user guide and introductory videos. In addition, a SCADR researcher has recently produced a Data Explained document which provides a user perspective on using the data. This talk will describe the approach to date with our flagship dataset on looked after children. It will outline the next steps which will enable comprehensive metadata to be developed for other datasets recognising the range of data controllers and users supported by the programme and the drive to support wider use.

Email: cecilia.macintyre@gov.scot

Producing an Administrative Data Based Occupied Address Dataset

Ayesha Barnes, Owen Maynard - Office for National Statistics

Data on the number, size and make-up of private households in England and Wales is regularly highlighted as a key need for national and local governments and charities. The decennial Census in England and Wales is the only time we get this data at a national level. In inter-censal periods, the Office for National Statistics produce household projections. Any inter-censal estimates for England and Wales produced on household make-up and family type are based on surveys. Other UK administrations do produce annual household estimates, but the methods employed for these are not consistent, with some relying heavily on Census data, which over time is outdated. The Admin Based Occupied Address Dataset (ABOAD) aims to fill this gap and produce annual data on the number, size and make-up of occupied addresses in England and Wales. We are looking to produce a person, and address, level frame which can then be used to replicate Census-style outputs, as well as providing a platform to further develop address level statistics, such as admin-based household income and over-crowding statistics. We will present a national level overview of our outputs, as well as some LA level case studies, to allow us to identify some of the strengths and weaknesses of the product so far, including the availability of address level data and uncertainty around address matching. Future developments of the ABOAD will look to identify within-address relationships to be able to better understand family and cohabiting units in England and Wales.

Email: owen.maynard@ons.gov.uk

Comparing the 2021 Census to administrative data to better understand the population estimation challenge

Emilie Woodhall, Elizabeth Pereira - Office for National Statistics

We will present on our analyses of the linked dataset between England and Wales 2021 Census/Census Coverage Survey and the Demographic Index (DI). This will provide insights on the quality and coverage of administrative data, enabling improvements in its use in informing population estimates. This will also provide evidence on additional administrative sources needed to estimate the population, and inform future linkages. The DI is a composite dataset made up of several linked administrative sources, therefore, the linked output is complex. It comprises individual sources, linked to form clusters, which have been linked to Census. To carry out the analyses, a flagging strategy was designed to form cuts of the linked dataset that are specific to research needs. High-level research questions were designed, and initial analyses completed, including: comparing the coverage of the DI and Statistical Population Datasets (SPD), which use inclusion rules to approximate the usually resident population, to Census; comparing how Communal Establishments are captured in the DI and Census and comparing the geography of those found in both the DI and the Census. Further analyses requirements to provide evidence to support the population estimation challenge have been established. The results will inform the National Statistician's 2023 recommendation, but also the future delivery of transformed population and migration statistics. This includes providing insights to improve the DI's linkage methodology, the construction of the SPD, and identifying the extent of over and under coverage in the SPD, allowing development of an estimation strategy to estimate the population using administrative data more accurately.
Email: elizabeth.pereira@ons.gov.uk

Estimating population by time of day

Robyn Hunt, Sam Butler, Ethan Moss, Hana Pauna and Michael Hawkes - ONS

Traditional population statistics estimate the usually resident population at their usual place of residence. Demand for statistics that employ alternative population bases is growing, including for those that estimate population present; an estimate of the number of all people located in a particular geography at a particular date and time. Use cases for population present statistics can include infrastructure planning, disaster/emergency response, and public health monitoring. The Office for National Statistics is undertaking research to develop methods to estimate small area population present ('daytime population statistics') as part of a wider population and migration statistics transformation programme. We have used a dasymetric redistributive model developed by Martin, Cockings, & Leung [Annals of the Association of American Geographers, 105:4, 754-772] to produce experimental small area daytime population statistics in a case study of 14 Local Authorities in England & Wales using a range of administrative, survey, and open data. The method is flexible, expandable, and can produce granular spatiotemporal estimates of population present given the appropriate input data. However, large coverage gaps for some types of population mobility in our input data (particularly around leisure and rail travel) limit the accuracy of these experimental statistics, and estimating their quality without a reliable benchmark is challenging. We are now investigating the use of high-frequency low-latency data, for example sensor footfall data, to validate these statistics, and ultimately to improve their accuracy and timeliness.

Email: michael.hawkes@ons.gov.uk

DWP's dynamic microsimulation models
Sam Mold, Ben Durcan, Stuart Grant, Gail Ludlam – Department for Work and Pensions

Department for Work and Pensions is responsible for welfare, pensions and child maintenance policy. It's the UK's biggest public service department, and delivers the State Pension and a range of working age, pension age, disability and ill health benefits to around 20 million customers, with over £230 billion spending per year, the majority of this on State Pension. We need great modelling to help manage this huge spending commitment. To do this, DWP develops and maintains a suite of microsimulation models for both working age and pension age. In this talk, we will introduce the models, focussing on the main pension age models:

- PENFORM – a medium term (next 10 years) pensioner benefits forecasting model
- PENSIM – a long term (next 80 years) pension accrual policy simulation model

We will also discuss our plans for the next generation of long-term modelling.

Email: sam.mold@dwp.gov.uk