Developments in official population statistics 1

Wednesday 3 September, 9am

Enhancing Official Population Statistics in Africa through Capacity Building and Partnerships
Edward Morgan, Steve Smallwood - Office for National Statistics

This presentation will discuss effective approaches to enhancing official population statistics in low-resource settings, with a particular focus on capacity building initiatives in African National Statistics Offices (NSOs). African NSOs often grapple with the challenge of generating precise population projections amid constraints such as incomplete census and survey data. The Office for National Statistics (ONS) has played a pivotal role in supporting these NSOs, notably in Tanzania, Malawi, Zimbabwe, and Namibia, through programmes like the Women's Integrated Sexual Health – Dividend (WISH-D) project, funded by the UK's Foreign, Commonwealth & Development Office (FCDO), and Overseas Development Aid.

The assistance provided by the ONS encompasses technical training in population projection methods and indirect estimation techniques. These techniques are indispensable for deriving accurate demographic estimates from limited data. Furthermore, the ONS aids in stakeholder engagement and project management, ensuring that demographic analyses are effectively incorporated into national policy and planning processes.

A central tenet of the ONS's strategy is the cultivation of partnerships among NSOs, fostering knowledge exchange and collaborative problem-solving. This approach not only builds sustainable capacity but also promotes the harmonisation of demographic methods across different countries, facilitating regional comparisons and joint research efforts.

Through these comprehensive efforts, the ONS significantly contributes to the improvement of official population statistics in Africa. Reliable population data are crucial for evidence-based policy-making and achieving sustainable development goals. The model of international cooperation and capacity development presented here offers valuable insights and a replicable framework for future initiatives aimed at strengthening statistical systems globally.

Email: edward.l.morgan@ons.gov.uk

Scotland's Census – how do I get the data?
Greg Blackadder - National Records of Scotland

Scotland's census is the official count of every person and household in the country. The answers people give to census questions help build a detailed picture of the population. The resulting datasets have enduring value due to the population level coverage of the information, and the range of variables recorded.

Over the past 2 years, National Records of Scotland have released a range of data products and services for the 2022 census. For researchers seeking to analyse and understand Scotland's population, these are an invaluable resource. For example, there are:

- Publicly available aggregated tables
- Online flexible table builder
- Microdata sample files
- Longitudinally linked data
- Custom data linkage projects

In this presentation we will describe our offer to researchers, to help you understand the resources available to you in analysis of Scotland's population.

Email: greg.blackadder@nrscotland.gov.uk

2022-based subnational projections for Wales
Martin Parry, Stephanie Harries - Welsh Government

The Welsh Government publishes subnational population and household projections for the 22 local authorities and 3 national parks in Wales. The last set of subnational projections were published in 2020.

The Welsh Government will be updating its subnational projections in 2025. These will be the first set of subnational projections for Wales that will reflect Census 2021 data and the latest demographic trends that we have seen since the COVID-19 pandemic.

In this session, Welsh Government statisticians will be providing an overview of the latest subnational projections for Wales. This will include any changes to the methods for calculating the projections, such as changes to membership rates used in the household projections and changes in the statistical programme used to calculate the projections. We will also discuss how we have engaged with users to inform the plans for our subnational projections as well as for our annual population and household statistics publications.

Email: stephanie.harries@gov.wales

Administrative Data-Based Population and Household Estimates for Scotland
Tom Macintyre, David Rowley, Uche Okoro - National Records of Scotland, Scottish Government

The Administrative Data-Based Population Estimates (ABPEs) provide an estimate of Scotland's usual resident population, derived exclusively from administrative data. Similarly, the Administrative Data-Based Household Estimates (ABHEs) offer an estimate of the number of households in Scotland using the same data sources.

To derive these estimates, multiple administrative datasets were linked together. Records were resolved into individuals to produce an Administrative Data Record Set (ADRS), which was trimmed using activity-based business rules to identify persons believed to be usual residents as of the reference date, resulting in Scotland’s Integrated Demographic Dataset (SIDD). The ABPEs represent the count of individuals within the SIDD. Household estimates are generated by grouping SIDD individuals by their addresses.

The SIDD is largely comparable to census data. The ABPE total is 0.22% lower than the Census 2022 estimate. Variations are more pronounced at the Local Authority (LA) level, although for all but one LA, the ABPE falls within the census' 95% confidence interval target of 3%. On average, the ABHEs are 6.8% lower than the rebased household estimates and approximately 7.1% below the Census 2022 estimate. The ABHEs indicate fewer smaller households (one or two persons) and more larger households (three or more persons) compared to official estimates.

When benchmarked against the re-based Mid-Year Estimates (MYEs) and the census estimates, the ABPEs perform well in comparison with the original MYEs. Differences by LA suggest greater uncertainty than reported in the census. Due to incomplete address information in some administrative records, the ABHEs are lower than official estimates.

Email: tom.macintyre@nrscotland.gov.uk

Developments in official population statistics 2

Thursday 4 September, 9am

Feasibility research into using administrative data and predictive modelling to develop disability statistics for England
Charlotte Standeven, Jesse Ransley, Matthew Minifie - Office for National Statistics

Our best, official estimates of population by disability status for England and Wales are from census, but there is a strong user need for more timely, robust protected characteristics statistics between censuses. This presentation outlines ongoing, exploratory research into the feasibility of producing person-level estimates for disability status in England by applying a predictive model to a suite of administrative data. This work focuses on the medical model of disability due to the scope of the datasets available.

For the model, a range of administrative data sources were used to predict disability status, including information on benefits, education, employment and health. Machine learning algorithms were used to train a model on a subsample of the data. Predictor variables include age, sex, geography and a variety of benefit, education, employment and health variables taken from the administrative data sources.

The performance of the model was then tested and evaluated on a different subsample of the data. A probability score of being disabled was generated for each record on the testing subsample and then converted to a binary disabled/non-disabled variable by reference to a calculated threshold probability.

As this is still work in progress, we will also discuss the model's evaluation metrics, and its efficacy, along with predicted disability prevalence rates by age, sex and geography compared with observed prevalence rates to assess the coherence between the predictive model and Census 2021, and potential next steps.

Email: charlotte.standeven@ons.gov.uk

Location Differences for Individuals on Scotland's Integrated Demographic Dataset and Census
David Rowley, Tom Macintyre - National Records of Scotland

The census is among the richest available population-scale, individual-level datasets. Scotland's Integrated Demographic Dataset (SIDD) is a population-scale individual-level dataset using only administrative data.

To build the SIDD, several administrative datasets were linked together, including those from health, education and the electoral register. The records were resolved into individuals to produce the Administrative Data Record Set. This was trimmed down using activity-based business rules to a set of individuals (the SIDD) believed to be usual residents on the reference date. The SIDD was linked to the census dataset to test its quality.

The SIDD is broadly comparable to the census. 95.6% of census records appear on the SIDD, while 84.7% of SIDD records appear on the census. 89.4% of individuals appear at the same address on the census and the SIDD. 98.2% appear in the same local authority (LA), although this is lower for individuals in their 20s. LA differences are not unbiased, e.g. there are more people in their late 20s in a suburban LA on the SIDD and another LA on the census, than vice versa. For example, East Dunbartonshire (on the north side of Glasgow) has 73, 196 and 403 individuals appeared there on the census, but other LAs on the SIDD, for age groups 18–22, 23–27 and 28–34 respectively. The equivalent numbers of individuals appearing there on the SIDD but elsewhere on the census are 212, 708 and 900.

The SIDD accurately captures the locations of most individuals, although there are systematic differences.

Email: david.rowley@nrscotland.gov.uk

Building an Admin-Based Census: How we are creating microdata on the residents of England and Wales for population research

Izzy Pickles - Office for National Statistics

Accurate and timely data on the population is a priority for local and national policy and planning departments, as well as an array of other users of population statistics. Currently the decennial census is the key source of these data, and accredited researchers can access samples of census records in a Trusted Research Environment at the record-level for their projects. As we move further away from a census however, the less reliable this data becomes, and is why the ONS have been developing methods to utilise administrative data to better meet user needs.

As part of ONS' programme of using more administrative data in the production of population and migration statistics, we have released experimental, census-like microdata on the usually resident population of England and Wales for 2023 to the Integrated Data Service (IDS), for accredited researchers to use in their work. This presentation will explain how we create this dataset from linked administrative data sources as well as how it relates to the Admin Based Population Estimates. It will detail our work to date - including limitations and opportunities - on enriching the individual records with characteristics typically found in the census such as ethnicity, as well as our plans to add further characteristics which are not as feasible to measure via a census.

Email: izzy.pickles@ons.gov.uk

Building on 2022: Administrative Data for Coverage-Corrected Population Estimates

Ralph McDevitt, Benjamin Little - National Records of Scotland

Scotland's 2022 population estimates showed that linking administrative systems can cut census bias and tighten confidence intervals, but also revealed a trade-off: surveys miss people, while admin files include records that should have been deleted. Over-coverage dominates the accuracy problem when using administrative data. Even admin-based population estimates, while achieving high-level accurate population counts, tend to include over and under-coverage errors that balance out, making them unsuitable without modification for coverage estimation.

We ask: how should Scotland use administrative data in 2031—alone, with a survey, or beside a census—to deliver transparent coverage correction?

Using the Scottish Integrated Demographic Dataset we prototyped three designs: (1) admin-only (required for the Dynamic Population Model for mid-year estimates); (2) admin+ coverage survey; and (3) census estimation with/without a coverage survey. For each we built a person spine and tested three over-coverage controls: rule-based inclusion criteria, a trimming algorithm, and a predictive model. Guided by steering-group advice we prioritise deterministic rules, validating them with predictive models.

A flexible decision-tree workflow allowed evidence on over-coverage to guide the design census choice.

We conclude that administrative data should be treated as both opportunity and risk: capable of increasing precision and decreasing bias caused by dependence between lists in census population estimation. It is also essential for producing frequent high quality population updates, yet, alone, is not able to provide unbiased population estimates to replace the quality of census until wider data access and mature over-coverage methods are in place; factors that may change in the future.

Email: benjamin.little@nrscotland.gov.uk

Research and development of the admin-based households estimates, by number and size for England and Wales

Elizabeth Pereira, Sally Mylles, Niamh Walshe, Angharad Noonan, Beth Allen,, Hayleigh Ball, Ffion Jones and Paulina Galezewska – Office for National Statistics

Timely and accurate households estimates in England and Wales are a key user need, however between Censuses, surveys cannot produce the same granularity. Our ambition is to produce the first admin-based household estimates (ABHEs) by number and size for each Local Authority (LA), in England and Wales.

To devise our research plan we conducted an international literature review, a quality options strategy and a methodology options paper. We utilise linked person-level administrative data, and also draw strength from survey data.

The first step of our plan is to explore seven different methods to produce admin-based household population estimates, aligned with the admin-based population estimate. The second step is to create the admin-based number of households by LA, again considering seven potential methods. The third step is to produce the admin-based household estimates by size focusing on three potential methods.

Presently we are undertaking the research and will share results if available, discuss considered interpretation and success measurements. In addition to comparing our ABHEs to Census and official surveys we are researching other administrative sources to produce a proxy for the number of households.

This research faces multiple challenges including: household definitions, identifying statistical methods and quality metrics and administrative data coverage, quality and access. We will explain how we addressed these issues, highlight remaining challenges, and outline our plans to tackle them further.

Producing ABHEs presents challenges but also involves innovation and ambitious research aiming to produce a proof of concept to later enhance further to Official Statistics in Development.

Email: hayleigh.ball@ons.gov.uk

Innovation in travel and tourism statistics: Measuring the UK's temporary population
Claudia Jenkins - Office for National Statistics, Sabina Kastberg - Office for National Statistics, - , - ,

Following a review of our statistics in 2019, the Office for National Statistics have improved the way we collect travel and tourism (T&T) data to deliver more efficient, accurate and coherent T&T statistics. This involves moving away from the sole use of the International Passenger Survey (IPS) and instead using a new mixed methodology approach. The new approach from July 2024 includes the introduction of household survey data (e.g. The Great Britain Tourism Survey), with the ONS's longer-term ambition being to move to an administrative data model with machine learning methodologies.

As outlined at BSPS last year, ONS have been exploring the possibilities that short-term accommodation, mobility and financial transaction data can bring to better understand international mobility and the temporary population in the UK.

However, this year, our presentation will go one step further to discuss the complexities involved in introducing novel methodologies to measuring T&T and provide a deeper dive into some initial results from integrating survey data with administrative data sources: these administrative sources being Home Office Borders and Immigration Data, Visa Card data and predictive modelling.

Email: claudia.jenkins@ons.gov.uk