Distr.: General
07 September 2022

English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Twenty-fourth Meeting
Geneva, Switzerland, 21-23 September 2022
Item 5 of the provisional agenda
Transitions in census methodology; plans, experiences and innovations

Transformed integrated population and migration statistics: an update

Note by Office for National Statistics*

Abstract

ONS is transforming its population and migration statistics to use new and existing data sources to meet our users' needs for better evidence to support national and local decision-making. This session will provide an overview of our research, including: developing Administrative-Based Population Estimates (ABPE) and Administrative-Based Migration Estimates (ABMEs); exploring groups, such as students, who may not interact with administrative sources in the same way as the rest of the population; exploring the viability of dynamic population models to produce timely, coherent population and migration estimates; and developing our ability to produce small-area multivariate statistics for a selection of topic. Census 2021 provides an exciting opportunity to evaluate our methods and outputs. We are linking Census estimates to our combined administrative data index and ABPEs to evaluate their quality. We will outline the aims of this linkage exercise and any published early findings. We are developing a Dynamic Population Model (DPM) to use all data available to us to produce frequent, timely and coherent population estimates. The session will include an update on what the DPM is, what we have learnt so far and our next steps.

*Prepared by Louisa Blackwell

I. Introduction

1. ONS is transforming its population and migration statistics to use new and existing data sources to meet our users' needs for better evidence to support national and local decision-making. While the census and mid-year population estimates based on the census provide the best picture of society at a moment in time, how we produce our population and social statistics is changing. We are using a variety of data sources, to provide more frequent, relevant, and timely statistics. This will allow us to understand population change in local areas this year and beyond.

2. At the heart of the new system is a dynamic population model (DPM). This model draws strength from a range of data sources including administrative and survey data, incorporating not just those data but also measures of their statistical quality. The model also includes data and system models, transparently using known features of demographic behaviour and data. The outputs will be fully coherent estimates of population counts and changes due to births, deaths and migration. For sub-national estimation this will include internal migration. Real-time estimation inevitably involves some forecasting. We have therefore developed a Real-time data dashboard, which aims to detect changes in demographic behaviour as they occur. Our timely estimates will be provisional, subject to later confirmation as we move from provisional to finalised data.

3. In this article we describe the DPM in more detail. We also provide an update on admin-based population and migration estimates, which are key feeds into the DPM. We also outline how we are researching and accounting for special groups such as students, who have characteristic demographic behaviours and notably different interactions with administrative systems. We describe ongoing research into small-area multivariate statistics in the years following Census 2021 and how the DPM will support and inform their production.

II. The Dynamic Population Model

4. The DPM uses statistical modelling techniques to combine a range of data and demographic insights to estimate the population. It is flexible by design, adaptive to changes in data inputs and can incorporate data with known quality limitations, such as incomplete or lagged data. It formally includes demographic trends and existing data sources, such as cross-border moves between Scotland, Northern Ireland, England and Wales.

5. The Covid-19 pandemic has highlighted the need for timely population estimates, for example to understand in real time the population at risk and for vaccine distribution purposes. Inevitably, improving timeliness involves a trade-off with other dimensions of quality, namely accuracy and reliability. There are lags in administrative data availability and early extracts of administrative data have 'edge effects' whereby the latest data can be altered as the data supply matures. We therefore consider the near real-time estimates from the DPM to be provisional, subject to later confirmation. This is a similar approach to our production of GDP (Gross Domestic Product), where early estimates are provided and then later revised as more data become available.

6. Like the current mid-year population estimation process, the DPM uses the cohort component approach whereby population at time t+1 is counted as:

population at time t
plus births and immigrations between t and t+1
minus deaths and emigrations between t and t+1

7. There are multiple input datasets and information about the reliability of the data, for example patterns of under- and over-count. Examples of demographic behaviours that we include are childbearing patterns by mothers' age and migration patterns by age and sex, and

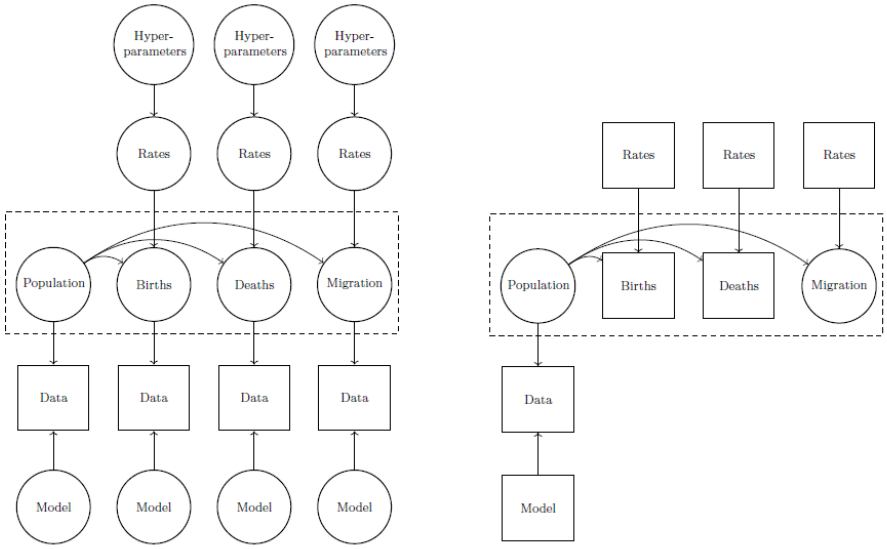
how these vary in different types of local authority. There are 331 local authorities in England and Wales, responsible for public services and facilities in their areas.

8. We are currently demonstrating proof of concept for the dynamic population model. In July 2022 we produced estimates from the DPM by single year of age and sex for a synthetic local authority for the years 2011 to 2022. In September 2022 we will provide proof of concept monthly estimates from the DPM to June 2022 by single year of age and sex for England and Wales, alongside annual estimates by single year of age and sex for a sample of local authorities. We aim to produce DPM estimates for all local authorities by month, age and sex for 2011- June 2022 in January 2023.

A. How the DPM works and how it is evolving

9. We have simplified the design of the dynamic population model address problems of computational speed. Figure 1 summarises the current development of the ONS dynamic population model. Beginning with the original framework that we considered (Bryant and Zhang, 2019), Figure I(a) shows the unobserved 'true' population stocks and flows within the dashed box. Circles within the dashed box represent components of the demographic account to be estimated: population stock, births, deaths and migration. The squares below are known quantities (data feeds), sometimes including multiple sources. For population stocks we use 2011 Census, National Health Service Patient Registers for 2011-2021 and Statistical Population Datasets for 2016-2020. Each count differs from the true, unobserved population. We construct a model to estimate the true population from them.

Figure I
Full and simplified demographic accounts frameworks



(a) Original framework (b) Exogenous rates framework

10. We could potentially have multiple datasets for births, deaths and migration, constructing models to estimate the unobserved true population flows.

11. In the area above the dashed box, we have underlying fertility rates, mortality rates and migration rates (though for immigration, counts are used). For example, in a university local authority we might expect high inward migration around university age, alongside high outward migration at graduation ages. Observed migration rates vary for many reasons but

the idea of an underlying rate of migration is useful, as are underlying fertility and mortality rates. The underlying rates are modelled using 'system models' that allow us to integrate prior information (capturing the true, underlying rates) to describe demographic regularities.

12. In Figure I(a) the original framework combines:

(a) knowledge about uncertainty of observed data (in data models at the bottom of the panel);

(b) demographic knowledge (in models for rates, described as 'system models';

(c) the demographic accounting identity, the concept at the heart of the DPM, which requires change in population stocks over time to be consistent with population flows of births, deaths and migration.

13. We require demographic accounts by single year of age, sex and Local Authority over time. This original framework is not computationally feasible at this level of granularity, largely because there are too many parameters to estimate.

14. Our solution is the simplified dynamic population model in I(b). This allows us to test new and considerably faster methods that make estimation feasible.

15. In I(b) we assume that counts of births and deaths are known and accurate, which is a safe assumption given the rigour of our vital events registration system. For now we are treating migration rates as true and known. Further, we estimate the demographic rates (system models) outside of the framework and treat these underlying rates as known.

16. We also simplify the data models below the dashed box. For example, we estimate outside of the framework the ratio of Patient Register population stocks to Census 2011 population stocks (the latter are assumed to be accurate) and model the underlying coverage issues in Patient Register data to use as inputs in our simplified framework. SPD data on population stocks is treated in a similar way.

17. The simplification in Figure I(b) produces reasonable estimates, although estimates of uncertainty are underestimated. As we develop the framework further, we move from the right panel towards the left panel of the diagram. The changes we make will be incremental, starting with underestimated uncertainty. Further details of our approach, which we consider to empirical Bayesian, can be found in the article, Integrated statistical design for the transformed population and social statistics system (PDF, 936KB).

B. How DPM compares to the current estimation process

18. Our existing method for estimating population suffers from increasing error the further we move away from the census date. The new method can use more data, accounts for uncertainty and adds demographic knowledge within a coherent and transparent framework. Most importantly, as we increasingly access more data, including very timely sources, this flexible framework can incorporate them all to produce more timely, consistent estimates with estimates of uncertainty.

III. Development of Admin-based population and migration estimates as data feeds for the DPM

A. Population

19. Through the last decade ONS has been researching how to optimise the use of administrative data for population estimation. The latest iteration of this research produced

Statistical Population Dataset (SPD) version 3 (though research on version 4 is currently underway). The SPD is a database that forms the basis for estimating the size of the resident population. It is produced by linking records across multiple administrative data sources and applying a set of inclusion and distribution rules.

20. Data used in the SPD are Pay As You Earn and Tax Credits, National Benefits Database and Housing Benefit, NHS Patient Register and Personal Demographic Service data, Higher Education Statistics Agency data, English and Welsh School Census and Birth Registrations. Records are linked and selected for inclusion into the SPD using signs of activity in the administrative data.

21. The SPDs were formerly known as admin-based population estimates. The name has changed in recognition that they are not a finished estimate, but are an important contributor to the DPM, where the estimation is carried out. This means we can use the benefits of the SPDs with measures of uncertainty alongside other data inputs to estimate the population.

22. Comparisons against ONS official estimates has demonstrated that 2021 counts from the SPD compare well with 2021 Census, prior to coverage adjustment. The DPM can use intelligence on coverage alongside SPD counts. We are currently using high-quality record matching to understand the relationship between SPD and the Census, which will be used to provide a coverage adjustment within the DPM. In non-census years we intend to use survey data to monitor and quantify SPD coverage patterns.

B. Migration

23. In parallel with population estimation, ONS has been researching the use of administrative data to measure international migration. This was accelerated by the necessity to produce estimates of international migration during the Covid-19 Pandemic. The International Passenger Survey (IPS), used since 1961, was stood down in March 2020, due to the pandemic. Stretched beyond its original purpose, the IPS had known limitations around measuring migration and our research into administrative data usage was well established as part of our population statistics transformation at this point.

24. In 2021 we published estimates of international migration for 2020 using statistical modelling to nowcast historic IPS trends using up-to-date administrative data. This was heavily assumptions-driven and not sustainable. In July 2020 we published new, experimental statistics making greater use of data from the Home Office and Department for Work and Pensions. Comparisons against 2021 Census results suggest that these estimates perform relatively better than the previous modelled equivalents.

25. While the DPM requires annual estimates of population stock from the SPD, we need monthly international migration data for nowcasting using population flows. Research into using administrative data and statistical modelling to produce timely and more granular migration estimates is ongoing.

IV. Special populations; the case of students

26. Student migration is a major contributor to both internal and international migration. Given their statistical significance for population estimation it is essential for us to understand how students are represented in administrative data and in our statistical outputs.

27. Our research has shown that not all students update their address in our NHS systems at the start of the academic year when they move. Graduates do not update their administrative data when they graduate. Taken together, these make it difficult to identify student households.

28. Detailed research on the accuracy and timeliness of administrative data was essential for quality assuring the 2021 census, especially since there was additional Covid-19-related disruption to both students' migration behaviour and data quality through this period. This research involved comparing the National Health Service Personal Demographic Service data with Higher Education Statistics Agency data.

29. As part of the quality assurance of the DPM we will be monitoring performance of the model in capturing student populations for every local authority in England and Wales. Options for further refinement of the DPM include separate sources and models for the student population.

30. A number of population groups differ in their demographic behaviour from the population living in private residence. Prisoners, school boarders and armed forces personnel on military bases are all examples where the population gets replaced rather than ageing in place through time. We will assess for each local authority the adequacy of DPM estimates for these population sub-groups.

V. Multivariate characteristics beyond the 2021 Census

31. We are developing methods for using administrative sources to capture population characteristics that are usually provided by Census, at a granular level. Granularity refers to providing estimates in the inter-censal period and at small area level. Topics being investigated include ethnicity, housing, highest qualifications, overcrowding, income, labour market status. Proof of concept for multivariate statistics has delivered income and housing statistics by ethnicity.

32. We recognise that some census topics are not available in administrative sources. Examples would be occupation, religion and unpaid care. We are researching the use of survey data applied to the DPM modelled outputs as a means for updating characteristics beyond the Census year.

VI. Conclusion

33. Results from our research are highly encouraging and over the coming months we are engaging with key stakeholders and the public to share and develop our findings and future plans.

34. The Office for National Statistics (ONS) has been supported in the DPM research by the University of Southampton. Specifically, we would like to thank John Bryant, Peter Smith, Paul Smith, Jakub Bijak and Jason Hilton for their guidance and support

References

Bryant, J. and Zhang, J.L. (2019) Bayesian Demographic Estimation and Forecasting, Boca Raton: CRC Press.