

Putting administrative data at the core of the Population Statistics System – the experience in England and Wales

Rebecca Tinsley, Office for National Statistics, Royaume-Uni

Abstract

The Office for National Statistics (ONS) has been using integrated data to research into the Government ambition that “censuses after 2021 will be based on alternative sources of data.” The Administrative Data Census project has made progress in: producing a range of Research Outputs that are typically produced by the ten-yearly census; comparing these outputs with official statistics; and seeking feedback from users.

Research so far has covered a range of outputs including the size of the population, household statistics, and a range of population characteristics. These outputs have used a range of integrated methods and data sources including administrative data, commercial (aggregate mobile phone data) and survey data.

Recently, we have shifted the focus of our research to understand how admin data can be used to produce not only the stock count of the population, but also components of population change, including estimates of international migration.

This is a key milestone in an ambitious programme of work to transform to a new administrative data-led system for population and migration statistics for England and Wales by Spring 2020.

This work builds and expands upon our research on an Administrative Data Census to move into a routine understanding of population and migration using all available sources.

This presentation will cover:

- Progress made so far
- Challenges in estimating coherent stocks and flows of population from admin data, including outcomes from collaboration at an international workshop on the same topic
- Insights into user feedback

Key Words:

1. Context: why are we transforming the population and migration statistics system?

Population and migration statistics underpin a wide variety of other statistics (such as unemployment rates) and inform a vast range of decisions. For example, decisions about local services (such as the number of school places or the provision of health services for an ageing population) and decisions about where to site new businesses.

Users have told us they need us to provide coherent statistics on the size (or stock) of the population, and how it changes over time (flows, both nationally and locally). We also need to tell an understandable story about what is contributing to this change and show how different groups in the population impact on society and the economy, including on our workforce, communities and public services such as the NHS and schools. This needs to be recognised as the story that is being experienced by our users.

Our users have also told us that they want us to deliver these statistics frequently, and in a timely manner to enable them to make evidence-based decisions. Our statistics also need to be relevant in a rapidly changing society, and we need to be able to report on their quality.

This report brings together our transformation work on population and migration, building on our knowledge and investigating how the wide range of administrative data sources now available to ONS¹ can be used, alongside surveys, to improve the way we measure population stocks and flows in the future.

2. Key challenges for transforming the migration statistics system for the UK

¹ The [Digital Economy Act 2017](#) provides new opportunities for ONS to access existing data held by other government departments, for the purpose of producing research and statistics.

The ONS produces population statistics for England and Wales, and international migration statistics for the United Kingdom as a whole. Population statistics for Scotland and Northern Ireland are produced by the respective statistical agencies.

The size (or stock) of the population at a point in time is a key component of our population statistics system. However, of equal importance is measuring and understanding how the population is changing over time (flows, both nationally and locally), and what is contributing to this change. The key factors of the changing population are natural change (through births and deaths) and through migration (both international and internal migration). For internal migration, we already use certain administrative data sources such as health and education records to help us estimate movement of people within England and Wales. However, international migration – movement of people into and out of the UK - is one of the more challenging aspects to measure.

Current methodology for estimating international migration is based primarily on how long respondents to the International Passenger Survey (IPS) say that they intend to be in or out of the UK. Whilst the IPS clearly plays an important role in our understanding of international migration, ONS has long acknowledged that it has been stretched beyond its original purpose. For example, the size of the survey sample means that estimates at a local level or for certain groups of migrants can be subject to relatively wide margins of uncertainty. We currently use patterns in administrative data to distribute the IPS flows to local levels when producing population estimates. Our users have told us they need better evidence to support decision-making, and a coherent understanding of what different administrative and survey sources, including the Annual Population Survey (APS) and Labour Force Survey (LFS), tell us about international migration.

The way ONS transform international migration statistics needs to be consistent with the way we transform population statistics more broadly. If we do not align our approaches, we will not be able to deliver statistics that tell our users a cohesive story on how the population is changing and what factors are driving this.

We know that there is no single, comprehensive data source that tells us everything our users want to know. However, integrating multiple data sources together will give us a much richer understanding of how our population is changing, how many migrants are in the UK and what public services they interact with. The challenge here is twofold; linking these data without a unique identifier and; integrating data in a way that maximises the strengths of all the different sources. This is essential for creating a system that can deliver timely, robust and coherent statistics both now and in future. Administrative data is collected in different ways, for different reasons and therefore has strengths and limitations in terms of coverage and timeliness. We need to fully understand these to develop a robust approach to using this data to measure international migration and the wider population.

2.1 Concepts and definitions

Whilst our current concepts and definitions cover a range of important information on population and migration, they are not comprehensive. People's lives are complex, as illustrated by our [previous research](#) on what Home Office administrative data can tell us about travel patterns into and out of the UK. In the future, we need a flexible approach that enables us to produce estimates of the population relating to the standard usually resident definition, but also other bases of the population, for example the daytime population to understand the impact that the increasingly more mobile population has on different services, such as health and education provision. Similarly, migrants enter and leave the UK for a variety of reasons, stay for different lengths of times and interact with society and the economy in different ways. We therefore believe that additional or alternative definitions may be needed to better reflect this complexity.

An important example of this complexity is circular migration, which is not covered by the statistics ONS currently publish. We have initiated research to better understand these patterns of movement using Home Office administrative data on non-EEA nationals who held, or went on to hold, a non-visit visa. This research helps us to assess whether we are including people in the correct definitions of migrants or visitors.

Our [update on our Population and Migration Statistics Transformation Journey research engagement report](#) gives more details on this work.

3. Our progress to date

We have already delivered a series of research outputs that have started to explore what administrative data sources can tell us about population and migration, and how we can use these to design an improved system for producing statistics in future.

3.1 International migration (UK)

We have previously published targeted pieces of research that strengthen what the IPS tells us about international migration. Our latest research on administrative data sources, [published in July 2018](#), focused on how we can triangulate these with existing survey data to develop a clearer picture of trends in migration. The report addressed how the IPS compares to Home Office visa data for non-EU migrants and set out the findings from an independent review into the quality of the IPS.

This work gave us a better, but not complete, understanding of what administrative data can tell us. Comparisons of Home Office visa data against the IPS showed some differences in patterns but that there are a multitude of reasons for this, including the fact that the data sources are intending to measure different things – intentions (IPS) compared with actual travel (administrative data). As these early findings related to non-EU migrants only, we were unable to draw firm conclusions about net migration at this stage.

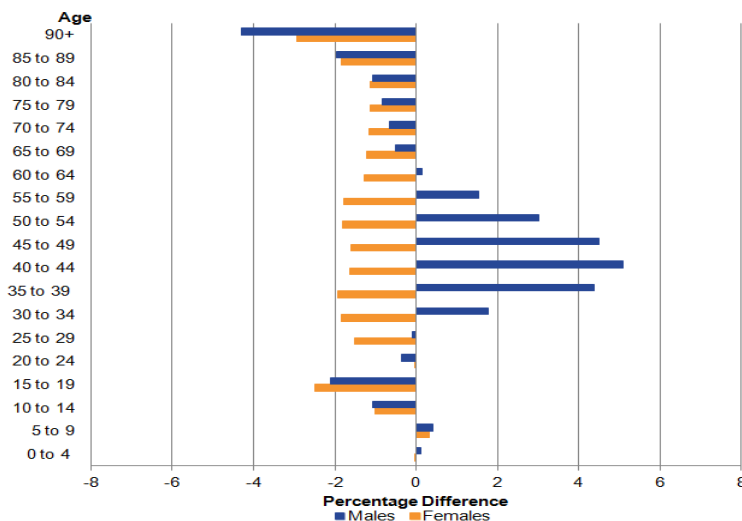
This highlights that no single source of information gives a clear view of migration. However, by bringing data sources together we can provide better evidence for our users - including adjustments to our estimates if the evidence supports this. As a result of this work, we set out our plans to take forward further research into wider data sources, including for EU migrants where our data coverage is more limited.

3.2 Population (England and Wales)

To deliver the ambition to conduct censuses using other sources of data, ONS set up the [Administrative Data Census](#) project in 2015. This aims to produce statistics on a wide range of topics usually produced by our decennial census, or topics which have a high need from census users such as; the size of the population, information on households and families, characteristics of the population, and housing.

We have already published a series of [research outputs](#) from this project, for which a focus has been producing [estimates on the size of the population](#) using alternative data sources such as tax and benefits records, health registrations and education data. We found that using linked administrative sources and a series of simple rules produced promising results on the size and structure of the population at local level, we call this Statistical Population Dataset v2 (SPDv2). Figure 1 shows the coverage patterns by age which suggest that the combination of data sources and the rules that we applied could be improved for particular ages.

Figure 1 showing under and over coverage of SPDv2 compared with the census population estimates, by 5-year age group and sex, for 2011.



As a result, we are now taking a different approach, using our knowledge of the data sources and how they relate to different population groups to produce a series of data-driven rules.

4. Data sources: what can we use to measure population and migration?

As administrative data is not collected for statistical purposes, when we try to use it to produce statistics which relate to specific definitions, we find that each data source has its own unique coverage patterns and statistical quality considerations. For some sources, presence on that source means we can be relatively sure an individual is usually resident in the country eg registration for an undergraduate degree. Conversely, other sources are useful as a snapshot evidence of presence in the country e.g. receiving a student loan. Together these impressions can provide a longitudinal picture of population change.

Our previous research, and working closely with data suppliers to better understand the data, has helped us learn where each source has its strengths and limitations. We have put this understanding into developing a new way to produce a Statistical Population Dataset (SPDv3). This approach relies on us using the knowledge we have built to produce data-driven rules for including records we believe meet the usually resident population definition. We are taking a similar approach to determining confidence-based rules for identifying migrants who meet the existing definitions of long-term and short-term migrants, and extending this to expand our research on circular migrants. We will also continue to work closely with data suppliers to understand administrative data sources and how we can build these into our future systems.

5. How should we combine these data sources to produce population and migration outputs?

Once we have identified the right data sources, we need to develop the right methods for using them to measure what users need. Our aim is to integrate administrative data sources in a way that creates a flexible system for producing the range of insights our users need. This is both in terms of our ongoing official estimates for population and migration, but also our strengthening of the evidence base on important areas such as the impact of international migration on society and the economy.

5.1 Using administrative data to develop a stocks-based approach

Our previous [Administrative Data Census](#) research focused on our progress in developing a future **stocks-based approach** where we combine linked administrative data and applied a set of rules to produce a Statistical Population Dataset (SPD). This work demonstrated the potential for producing estimates of the usually resident population using administrative data, but early research showed the need for further refinements to produce estimates about the flows of the population between two points in time. It has also highlighted the importance of using a greater range of data sources and the need for a survey which can measure and adjust for coverage issues seen for different areas and different groups in the population.

Using the knowledge that we have developed from analysing our SPDv2 and understanding key data sources, we have developed a set of data-driven rules that we can use as part of a future system for determining which administrative records are part of the usually resident population. This approach focuses on identifying the data source that provides the best coverage for a given age group ('first hierarchy'). We then supplement any gaps in coverage, or limitations of that source by using other sources, to create a 'hierarchy' of rules ('second hierarchy' etc.). Developing rules in this way will enable us to be flexible as new administrative sources become available, or as sources change over time. Figure 2 shows our initial approach for specific age groups based on our understanding of the coverage and quality of each data source.

Figure 2 underscoring our approach to build data-based rules for determining usual residence.

| First hierarchy | | | Second hierarchy | | |
|---------------------|--|--|--|--|---|
| Age | Source | Rationale | Source | Rationale | |
| 0- to 4-year-olds | Birth registrations | Covers all births in England and Wales (EW). Births to usual residents in EW that occur outside EW are not included. | Personal Demographic Service (PDS) | High chance of registering non-UK born children. Will only qualify with evidence of active adult family member at same address. | Ongoing research: <ul style="list-style-type: none"> To establish "top of hierarchy" sources for each age group. Aim is to account for majority of population at top of hierarchy with minimal over-coverage. |
| 5- to 15-year-olds | English and Welsh School Censuses | State school pupils represent large proportion of population. Annual submissions reduce risk of over-coverage. | Child Benefit Extract (from CIS) | High take-up to account for children not in state schools but on child benefit. Child Benefit unlikely to be paid for children resident abroad and those from high income families. | |
| | | | NHS Patient Register (PR) | The new arrivals on the Patient Register, to include those not included on Child benefit or School census. | |
| 16- to 64-year-olds | DWP/HMRC Customer Information System (CIS) linked to: • Pay As You Earn (PAYE) • Benefits datasets (BIDS/NBD) | Provides evidence of economic activity, including: <ul style="list-style-type: none"> periodic benefit payments on or near reference date annual earnings exceeding expected income threshold evidence of low income longitudinally | Higher Education Statistics Agency (HESA) | Will cover a large proportion of the adult population not economically active with DWP or HMRC. Identify and remove short-term foreign students (using longitudinal analysis or nationality). | Our findings from PDS-MWS work show evidence of lags in registration, particularly for EU. We do find presence of migrants who are on neither the CIS-linked sources nor HESA (so PDS does fill some gaps). |
| Aged 65 and over | Likely to be pensions and health registrations | | | | |

As an example, our work on the 5-15 age group illustrates this approach. We have used the England and Wales School Census as our ‘single source’ as this provides high coverage of children attending state schools, which represents a large proportion of this age group. We then supplement this with ‘activity’ from Child Benefit data, and have included records from NHS patient registrations to account for gaps in coverage (for example, children attending private schools and home-schooled children who are not covered by the School Census data).

5.2 Using administrative data to develop a flows-based approach

In the existing system for producing official estimates of the England and Wales population, we use a cohort component method. Our starting point is the ten-yearly census. Each year, we age everyone on, use administrative data to add births, remove deaths and adjust for internal migration. We use the International Passenger Survey to estimate international migration flows (people immigrating into the UK, and those emigrating from the UK), and distribute that to local levels using administrative data. Once these flows have been added to the previous year’s stock total, we are able to produce a stock total for the current year. This can be thought of as a **flows-based approach**. Our ambition is to make far greater use of administrative data to produce these national estimates of international migration in future, so we have carried out new research to explore how we can develop a flows-based approach based on wider data sources.

We have outlined the challenges in measuring and understanding how the population is changing over time, particularly for immigration and emigration (flows, both nationally and locally) earlier in the report.

We are confident that birth and deaths are well recorded in administrative sources. For long-term immigrants and emigrants, this is more challenging as no single data source captures the patterns of movement for all types of migrants. Like the data-driven rules described in the stocks-based approach, we are developing a similar inclusion approach for long-term immigrants - we are calling this confidence based rules.

Our aim is to bring together the multiple sources to build a comprehensive and granular evidence base for migration to (and eventually from) the UK. This approach involves several steps

- Understanding which records are potential long-term immigrants (usual residents) using variables from health, education, national insurance and immigration data sources.
- We then look across longitudinally linked ‘activity’ sources to assess how different types of migrants interact with these systems over time and what that tells us about how long they stay in the UK.

- Build our understanding on the way different types of migrants interact with various data sources by linking these data sources together. We have already carried out further exploration of how we can use student enrolments with travel events data to better understand international student migration.

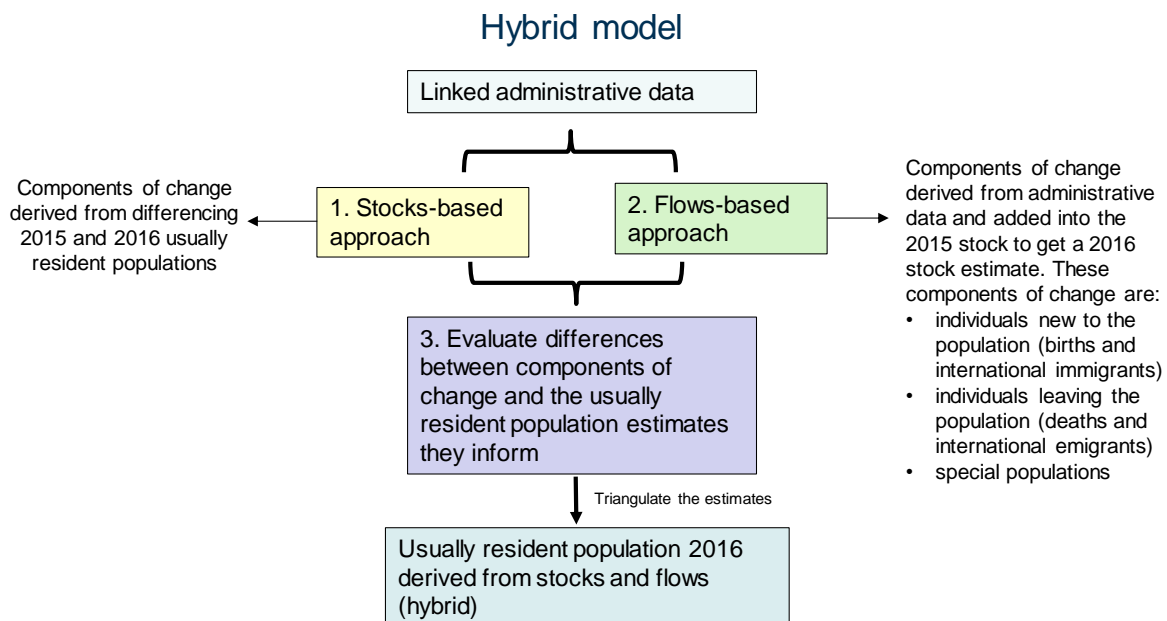
Both approaches are underpinned by our understanding of how international migrants and other population groups interact with these sources. Our [Transforming population and migration statistics report](#) includes our detailed findings which inform these data driven rules.

5.3 Bringing it together: our proposed hybrid model

Focusing on these approaches separately offers us the opportunity to produce the best-possible estimates for stocks and flows in future, using the best data and methods available to us. However, to produce a coherent set of statistics for our users requires us to develop an approach that brings the two methods together.

Figure 3 below sets out our hybrid model for delivering a transformed population and migration statistics system. We call this a hybrid model to reflect that we are aiming for an approach that produces the best possible stocks and flows of the population.

Figure 3 : Hybrid model



Once we’ve produced a set of statistics from both approaches, we need to evaluate the stocks and flows that we have produced, and triangulate them to enable us to produce a coherent set of population and migration statistics. This is likely to involve the need for a survey to measure and adjust for coverage patterns seen in the data (we call this a ‘Population Coverage Survey’ (PCS)). As part of our Survey Transformation Work, we are looking at how we can integrate the PCS with the Labour Market Survey and other residual data requirements into an Integrated Survey Framework.

We will ensure that we develop the hybrid model, described above, to be as flexible as possible, to enable new sources and methods to be used as they become available, or as they change over time. This approach also opens up the potential for providing a longitudinal aspect into better understanding the dynamics of population change. This could give us the chance to offer more insights into key policy and research areas, such as the economic outcomes of international migrants depending on length of stay and age at arrival.

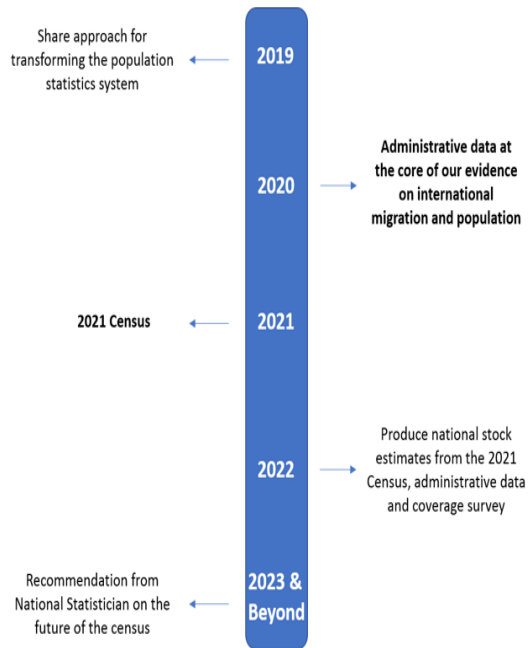
6. What might the pathway to transformation over the next 5 years look like and what are our next steps?

We have set ambitious targets to put administrative data at the core of our evidence on international migration and population by 2020, and will deliver a predominantly online census in 2021. It is also important that work done between now and the delivery of outputs from the 2021 Census paves the way for continuous improvement, and supports the ambition that “censuses after 2021 will be conducted using other sources of data”. Most importantly, the outputs that we publish should be coherent, and meet the needs of our users.

For this reason, we have taken the decision not to benchmark the Administrative Data Census outputs with the outputs from the 2021 Census. Instead, we will ensure that we use the best available sources to produce the best-possible outputs from the census, and to put administrative data at the core of our population and migration statistics system.

We will iteratively develop our transformed population and migration statistics system, taking on board feedback from users and making the best use of new data, and new methods as they become available. Alongside this, we will be developing our understanding of how surveys can measure and adjust for coverage patterns in our admin-based population

Figure 4 : Our transformation timeline



Our next immediate steps of research include:

- Continuing to collaborate closely across the Government Statistical Service (GSS) to build our expertise in the administrative data held across government, and to address key evidence gaps identified by the users of our statistics.
- Investigating new key data sources to improve the coverage, such as self-employed and EU and other non-visa requiring migrants.
- Continuing to consider the role of surveys (both new and existing) in our future system and developing methods that allow us to produce a coherent set of administrative data-based population stocks and flows at national, regional and local level in future– working in collaboration with our partners across the UK.
- Developing our approach to producing population stocks and flows further, based on feedback from our users. A key area of focus will be how we can measure emigration, given there are limited data sources that capture this well.
- Exploring what further breakdowns we can produce from linked administrative data, and what these can tell us more about the characteristics of migrants and the way they interact with public services – including at local level.
- Developing our approach for evaluating our transformation work, to help us demonstrate to users our progress towards delivering an administrative data-based population and migration statistics system.
- Continuing to review our existing suite of population and migration statistics publications– such as the Migration Statistics Quarterly Report – in light of our ongoing research. This is a key part of our transformation framework, to ensure we continue to meet user needs.

This work is iterative. We will be sharing the outcomes from our research into administrative data as we progress and plan to publish an update on our work in spring 2019. We also plan to publish further work in 2019 investigating the impact of migration on the health and education sector, as set out in the timetable we published in our previous [transformation update](#).