

Article

Update on our population and migration statistics transformation journey: a research engagement report

Report updating our users on our population and migration statistics transformation journey, and seeking feedback on our future plans.

Contact:
Becca Briggs
pop.info@ons.gov.uk
+44 (0)1329 444661

Release date:
30 January 2019

Next release:
To be announced

Table of contents

1. [Executive summary](#)
2. [Disclaimer](#)
3. [Introduction and scope](#)
4. [Main challenges and progress to date](#)
5. [Our framework for transforming population and migration statistics](#)
6. [What are our user needs?](#)
7. [Concepts and definitions: what do we need to measure?](#)
8. [Data sources: what can we use to measure population and migration?](#)
9. [What methods can we use to analyse administrative data?](#)
10. [What might the pathway to transformation over the next five years look like and what are our next steps?](#)
11. [We want your feedback](#)
12. [Annex A: Measuring stocks and flows - what progress have we already made?](#)
13. [Annex B: Case Studies: spotlight on what administrative data can tell us about international migration](#)

1 . Executive summary

Office for National Statistics (ONS) is transforming the way we produce population and migration statistics, to better meet the needs of our users. Working in partnership across the Government Statistical Service (GSS), we are progressing a programme of work to put administrative data¹ at the core of our evidence on international migration (UK) and on population (England and Wales) by 2020². This ambition is based on our current plans for acquiring access to the further administrative data sources we need to deliver this. Our work programme is also an integral part of the work over the next four years to make a recommendation to the UK government in 2023 about the future of population and housing censuses in England and Wales.

This report provides an update on our GSS [Migration Statistics Transformation Programme](#) and builds on the previous research delivered through this and the ONS [Administrative Data Census](#) project. Previous publications have set out our progress in linking multiple data sources together to produce [estimates of the size of the population](#). We have also carried out targeted work to better understand international migration, including reports into [student migration](#) and work to [compare the International Passenger Survey \(IPS\) and Home Office visa data](#) for non-EU migrants.

ONS has long acknowledged that the IPS has been stretched beyond its original purpose and that we need to consider all available sources to fully understand international migration. Users have also told us that they want to have a coherent understanding of what different data sources tell us and how they compare, including other administrative data and survey sources across government.

At the same time, our previous research clearly demonstrated that no single source of information can tell us everything our users want to know, or fully reflect the complexity of our changing population. Instead it has shown the value that can be gained from using linked administrative data, while highlighting the challenges of using these data to measure traditional definitions such as short- and long-term migration, and usual residence.

This report brings together our transformation work on population and migration, building on our knowledge and investigating how the administrative data sources now available to ONS³ can be used, alongside surveys, to improve the way we measure population stocks and flows in the future.

We are still developing our future system and are in the process of acquiring the further administrative data sources needed to deliver this – for example, to address coverage gaps for EU migration. As such, this report does not directly compare administrative data with our [existing IPS-based migration statistics](#) or make any overall assessments of their statistical quality. Instead, it provides an update on our approach towards building an administrative data-based system that will, over time, give us new insights on the quality of the IPS and our official international migration estimates for the UK. It also provides evidence of how different administrative sources can help us to better understand aspects of migration and reveals some of the different travel patterns that migrants make, such as circular patterns of movement.

We have also developed a new framework that puts users at the heart of what we do and describes the important questions we need to answer to meet their needs: What concepts do we need to measure and what definitions support these? What data can we use to answer users' questions? What methods can we use to analyse administrative data? What outputs do we need to produce? This report is structured around this framework.

Main points from our latest research

As part of the Administrative Data Census project, we have previously produced [estimates of the size of the population](#) by linking four main data sources⁴. We applied a set of rules to include records into a Statistical Population Dataset (SPD) to represent the “usually resident” population. Since then, we have shown that our previous SPD rule of requiring records to appear on at least two of our four main data sources results in varying coverage patterns for certain age groups or geographies.

We have now made progress towards a new approach for producing population stocks and flows using administrative data, by bringing more sources together to fill gaps in coverage. We have linked immigration, education, health and income records, and have explored how we can use these sources to determine the usually resident population of England and Wales and immigration flows to the UK. This includes developing data driven rules, based on registrations and “signs of activity” we can identify from each data source.

We have also produced a series of case studies that put a spotlight on what different administrative data sources tell us about international migration.

This latest research has clearly demonstrated the benefits of combining multiple sources to provide new insights into migration. The main points are summarised in this section.

Our new analysis of circular patterns of movement using Home Office Exit Checks data clearly demonstrates the complexity of the travel patterns we can see in the data. Looking at individuals who arrived in the UK on a non-visit visa and their travel patterns for the following two-year period, we were able to identify a range of circular journeys into and out of the country, categorise these into groups and look at their characteristics. For example, those we defined as having a low or medium number of journeys tended to be here for around two to five months and travelled for the purposes of study or family. There is therefore potential to produce statistics on circular migration in future, so we will explore how we can do this based on feedback from our users about what aspects are most important to them.

The latest research has improved our understanding of what administrative data can tell us about migration from the EU, building on previous work, which focused mainly on non-EU citizens. For example, when we linked NHS Personal Demographic Service (PDS) data to the Migrant Worker Scan (MWS), we found median lags between arrival and NHS registration of 276 days for EU nationals and 60 days for non-EU nationals in our linked dataset. This indicates that those registering for a National Insurance number (NINo) do not tend to access health services immediately. For EU nationals it also emphasises that wider data sources will be particularly important for identifying migration into the UK, given the time lags in the health data.

It has also provided new insights into international student migration by linking Higher Education Statistics Agency (HESA) data with Home Office Exit Checks and HM Revenue and Customs (HMRC) Pay As You Earn (PAYE) data. We found that a greater proportion of EU students in our linked HESA and PAYE cohort were employed compared with non-EU students – which likely reflects immigration rules and the economic background of students – and also provides evidence that our move to bring more sources into our SPD will help improve our coverage for non-EU students. By linking HESA and Exit Checks data, we also found that almost half of non-EU students in our linked data spend between 300 to 400 days in England and Wales during their first 14 months of study within a 16-month period. Our existing definition of long-term migration counts those who spend 365 days in England and Wales as resident, so this work shows the need to explore what new or additional concepts and definitions would help our users better understand migration patterns.

Our latest research linking the MWS to both benefits and annual PAYE data showed clear potential for earnings data to be an important source for identifying and measuring migration patterns. Our analysis illustrated that four in five non-UK nationals in our linked dataset had signs of activity in income and benefits data following arrival in the country, with more identified in PAYE data. Further linkage work, using more comprehensive PAYE Real Time Information (RTI) data, will therefore be instrumental in helping us to identify patterns of migration to and from the UK.

We are using our knowledge of the different data sources to produce a series of data-driven rules with the aim of better representing the usually resident population. We are also developing a set of confidence-based rules aimed at improving the coverage of international migrants.

Next steps

We will continue to collaborate closely across the GSS to develop our approach for putting administrative data at the core of our statistics, and to address important evidence gaps identified by our users. The case study research in this report focuses on the findings from linking individual data sources together to explore specific topic areas. To take this further and maximise the benefits of administrative data, our next steps will be to link across a fuller range of data sources available to ONS, to continue to develop our data-driven rules and build an integrated system for measuring population and migration.

Important data sources that we plan to focus on in our next research phase are:

- further Home Office administrative data
- PAYE RTI (and Self Assessment)
- further healthcare data
- further linked education data (All Education Dataset for England (AEDE))
- Council Tax
- other data sources such as Electoral Register and Driver and Vehicle Licensing Agency (DVLA) registrations

This will improve the coverage of our data, particularly for groups such as EU migrants. Whilst we have improved our knowledge of what administrative sources such as health and income data can tell us about both EU and non-EU migration, our existing evidence base is much stronger for non-EU migration. Our next steps will focus on how we can use these further data sources to improve our coverage and address known challenges such as using administrative data to measure emigration from the UK. As we progress, we will also continue to consider the role of surveys in our future system, alongside developing our methods for producing improved statistics at a regional and local level.

An important next step is to put together all of what we have learned so far to produce administrative data-based estimates about the stocks and flows of the population. We will continue to publish the findings from our research on an iterative basis to demonstrate our progress towards our ambition to put administrative data at the core of our evidence on migration and population statistics by 2020. We plan to publish our next update on this work in spring 2019.

Alongside this, we will also carry out further work to compare what existing survey sources tell us about population and migration, including the International Passenger Survey (IPS), the Labour Force Survey (LFS) and the Annual Population Survey (APS). We will publish an update on our findings in February 2019 alongside the latest [Migration Statistics Quarterly Report](#), with conclusions following later this year.

We want your feedback

Your feedback is important. We want to hear what our users think about our plans to put administrative data at the core of population and migration statistics, to ensure these continue to meet their needs. We have set out the main areas we want feedback on throughout the report and there are more details on how to provide feedback in [Section 11](#).

Notes for: Executive summary

1. Administrative data refers to data collected by other organisations (such as government departments) to support the delivery of services or for other operational purposes.
2. International migration is a reserved policy area whereas population is devolved, so ONS official statistics cover migration at a UK level and population for England and Wales. See [Section 4](#) which explains the different systems for measuring the population in Northern Ireland and Scotland.
3. 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.
4. These data sources are: NHS Patient Register, Department for Work and Pensions and HM Revenue and Customs' Customer Information System, England and Wales School Censuses, and Higher Education Statistics Agency student data.

2 . Disclaimer

These Research Outputs are not official statistics on the population nor are they used in the underlying methods or assumptions in the production of official statistics. Rather they are based on exploratory research using linked administrative data. They cover specific cohorts of people and are not fully representative of the population as a whole. These outputs must not be reproduced without this disclaimer.

3 . Introduction and scope

This report provides an update on progress towards delivering our ambition to put administrative data at the core of our evidence on international migration (UK) and on population (England and Wales) in 2020. It sets out our latest research, in collaboration with the Government Statistical Service (GSS), to explore how we can use administrative data to provide a better evidence base for our users. We also ask our users for feedback on our research findings and next steps.

We begin by setting out an overview of the existing population and migration statistics systems across the UK, alongside the main challenges and progress we have already made in developing an administrative data-based system (see [Section 4](#)). This brings together what we have previously published through both the ONS [Administrative Data Census](#) project and cross-GSS [Migration Statistics Transformation Programme](#).

Having set out this context, we then introduce our new transformation framework that puts users at the heart of what we do (see [Section 5](#)) and sets out the main questions we need to answer to put administrative data at the core of our statistics in a way that meets user needs. The rest of the report summarises how our latest research has moved us closer towards answering these questions. We discuss the concepts and definitions we need to measure to understand population and migration ([Section 6](#)) and what our research has told us about the available data sources we can use to measure these ([Section 7](#)). We then explain our latest progress in developing an administrative-data based method for measuring population stocks and flows ([Section 8](#)). This includes important information on [how ONS manages and keeps data secure](#), to provide the best standard of statistical information for the public.

Alongside this report, we have also published a series of slide packs that set out our research in more detail. These are signposted within each section, so that users can easily find out more about our full methodology and research findings.

Whilst our research has continued to develop and refine our approach for putting administrative data at the core of our statistics, we know that there is more to do. We therefore conclude this report by setting out our pathway to transformation over the next five years, alongside our immediate research priorities ([Section 9](#)).

4 . Main challenges and progress to date

How population and migration statistics are produced within the UK

Office for National Statistics (ONS) produces population statistics for England and Wales, and international migration statistics for the UK as a whole. Population statistics for Scotland are produced by the National Records of Scotland (NRS), while the Northern Ireland Statistics and Research Agency (NISRA) produces the statistics for Northern Ireland. Detail is provided in this section on how population statistics are produced for the respective countries.

ONS produces international migration statistics for Great Britain and works with NISRA to produce migration statistics for Northern Ireland, which are combined to create statistics for the UK as a whole. Migration statistics for areas within Scotland and Northern Ireland are produced by NRS and NISRA respectively.

Any questions within this publication that refer to population statistics are for users of England and Wales statistics only. For questions relating to international migration, we would be interested to hear from users across the UK.

Main challenges for transforming the migration statistics system for the UK

The size (or stock) of the population at a point in time is an important 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 main 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 the Labour Force Survey (LFS), tell us about international migration.

This is why, in September 2017, Deputy National Statistician Iain Bell announced [our ambition to meet the changing user needs for international migration statistics](#). In May 2018, we published an [update on our Migration Statistics Transformation Programme](#), including our plans to transform the information that the Government Statistical Service (GSS) produces and the timetable for delivering this. This included consulting with migration statistics users on our latest research findings and the shape that a new administrative data-based statistics system may take.

Main challenges for transforming the population statistics system for England and Wales

The way ONS transforms 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. This is an important challenge for the ONS and the wider GSS, given that it has long been recognised that the way we produce population statistics also needs to change. In March 2014, the National Statistician recommended that the census in 2021 should be predominantly online, making increased use of administrative data and surveys to both enhance the statistics from the 2021 Census and improve statistics between censuses. The government's response to this recommendation was an ambition that "censuses after 2021 will be conducted using other sources of data."

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 for ONS and the wider GSS is integrating data in a way that maximises the strengths of all the different sources, and creating a system that can deliver timely, robust and coherent statistics both now and in future. Administrative data are collected in different ways, for different reasons and therefore have strengths and limitations in terms of coverage and timeliness. We need to fully understand these to develop a robust approach to using these data to measure international migration and the wider population.

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.

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 with 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 that stage.

However, our investigation into the unusual fall in student numbers recorded in the IPS for the year ending 2016, did find clear inconsistencies with the most comparable Home Office visa data and Higher Education Statistics Agency data. In this instance, our assessment of the available sources led us to publish an illustrative revised trend for non-EU students in both the July 2018 and [November 2018 Migration Statistics Quarterly Report \(MSQR\)](#).

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.

Our latest findings are discussed from [Section 5](#) onwards, where we look at how we can develop our definitions of migration using administrative data, what data sources are available to us and how we can use these to better measure international migration flows.

We also set out our plans for future work in [Section 10](#), which includes carrying out further work to review what existing survey sources tell us about population and migration, including the IPS, the LFS and the APS. This is an area where users have told us that they want to have a coherent understanding of how different sources compare, to help them interpret the latest trends.

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 that have a high need from census users, such as:

- the size of the population
- information on households and families
- characteristics of the population
- 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 the Department for Work and Pensions' (DWP's) Customer Information System (CIS), the NHS Patient Register (PR), Higher Education Statistics Agency (HESA) data and the England and Wales School Census.

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. However, analysis by age and sex showed some coverage patterns, which suggest that the combination of data sources and the rules that we applied could be improved for particular age groups. This research is described in more detail in Annex A.

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. This work is outlined in [Section 8](#).

Population and international migration statistics in Scotland and Northern Ireland

International migration statistics

In Scotland, NRS are working with ONS, and other GSS partners, as part of the programme of work to improve international migration statistics, which should help address user demand for more evidence on the impacts of international migration, particularly at local level, as well as provide the best estimate of international migration to feed into Scotland's population statistics.

In Northern Ireland, NISRA measure international migration using information from the local health card registration system along with other administrative data. This long-standing approach has been adopted as the IPS cannot measure international migration into and out of Northern Ireland accurately (for example, some international migrants enter and exit Northern Ireland through Dublin Airport in the Republic of Ireland). ONS incorporates the NISRA migration estimates with IPS-based migration estimates for Great Britain to complete the UK picture on international migration.

Future population statistics

Separate programmes of work are taking place in Scotland and Northern Ireland, and ongoing collaboration will take place between ONS, NRS and NISRA to consider harmonisation and ensure population statistics across the UK remain comparable.

In Scotland, the Beyond 2011 programme was established by NRS in September 2011 to explore the future provision of population and socio-demographic statistics in Scotland. The programme closed in March 2014 following a recommendation to plan for a census in 2021. Looking to 2021 and beyond, NRS are keen to expand the use that they make of administrative data in the collection, production and quality assurance of population statistics.

Since 2014, NRS have been working with a number of public sector bodies to agree access to a range of administrative datasets, for example, information on students in higher education, to establish how these data can improve both current and future population estimates. NRS have been working closely with colleagues in ONS to develop methods for this work and to understand the strengths and weaknesses of different approaches.

NRS are currently considering how they might use administrative data to help quality assure and enhance the census data in 2021 and how they might be able to use it in data processing methods to improve data quality. NRS are continuing to research options for creating population and household estimates using administrative data and have been working with stakeholders in Scotland, including the Population and Migration Statistics Committee, and others on their plans. As part of Scotland's Census 2021 Programme, NRS will make recommendations for the approach to future censuses in Scotland, which will inform future developments to the population statistics system in Scotland.

In Northern Ireland, as with Scotland, a Beyond 2011 programme looked at the future provision of population statistics. A recommendation was made and accepted by Ministers to hold a census in 2021. NISRA is currently working on using administrative data to underpin the census in areas such as address register enumeration, post-collection data processing and possibly to enhance census outputs (via linked administrative data to the census database). The future provision of population statistics will be considered after the 2021 Census is complete. Any decision to reform the population statistics system in Northern Ireland will need to be put to Ministers for consideration.

5 . Our framework for transforming population and migration statistics

Our latest research has focused on bringing together the progress we have made in transforming population statistics for England and Wales, and migration statistics for the UK, and developing this further using the latest administrative data available to Office for National Statistics (ONS). We have structured this phase of work around a new framework, developed in collaboration with the Government Statistical Service (GSS), which sets out what we want to deliver and how we plan to deliver it. The framework puts users at the heart of what we do and describes the important questions we need to answer to meet their needs.

Figure 1: Our framework for transforming population and migration statistics

Source: Office for National Statistics

We have structured the rest of this research report around our draft framework, to make it clear how our latest research links to each aspect of this. We also welcome feedback from our users on whether our draft framework is covering the right areas and whether, and how, we should develop it further in future.

We want your feedback

- How do you currently use statistics on population and/or international migration published by ONS?
- What analysis or publications would you like to be able to produce from these in the future?
- Does our outlined framework miss any elements that are important for consideration?

6 . What are our user needs?

Our users sit at the heart of our framework. A rapidly changing policy context – including the government’s plans for a new [immigration system](#) once the UK exits the European Union – offers us a well-timed opportunity to reflect on the best way to deliver the population and migration statistics system to best meet user needs. The [Digital Economy Act 2017](#) also gives us further opportunities to collaborate across the Government Statistical Service (GSS) to better share and maximise the value of data that are already held, to deliver better evidence.

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.

We know from when we have previously engaged with our users that 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 be able 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.

Our recent discussions at a range of conferences and user forums has confirmed that moving towards an administrative data-based system is the right approach for delivering a stronger evidence-base and a more coherent set of statistics – but also emphasised that we need to continue engaging with our users as we develop this approach, to ensure we continue to meet their needs. In particular, users want to have an opportunity to feedback on any changes to our main definitions for population and migration statistics (such as those described in [Section 7](#)), our approach for integrating and analysing different data sources and how we can improve the information we publish to ensure it is fit for a range of purposes. This report provides an update on our latest research and offers users an opportunity to provide their feedback. [Section 10](#) explains how users can do this.

7 . Concepts and definitions: what do we need to measure?

As set out in our framework, statistical concepts and definitions – or what it is we are trying to measure – should be driven by what our users need to know.

What do we measure now?

Our population and migration statistics provide users with two sets of information:

- the size of the population at a point in time (the population “stock”)
- how the population changes over time (population “flows”) and what contributes to this change (both through births and deaths, and through international and internal migration)

How do we currently measure (or define) population stocks and flows?

For many users, we know it is important that our statistics are consistent with those produced by other countries. This ensures they can make “like-for-like” comparisons. Our current population and migration statistics are therefore closely aligned with main concepts and definitions set by the United Nations. This is set out in more detail in Table 1.

Table 1: Main concepts and definitions used in current UK population and migration statistics

Concept	How do we define it?	What data sources do we currently use to measure this?	What do we currently publish?
Usually resident population	ONS applies the standard UN definition for population estimates, which refers to people who reside in the area for a period of at least 12 months whatever their nationality.	<p>ONS population estimates are based on the Census and updated annually to reflect population change, including:</p> <ul style="list-style-type: none"> · birth and death registrations · health and education records, to account for migration within the UK · International Passenger Survey (IPS), to account for international migration <p>Note: population estimates for Northern Ireland and Scotland are produced separately by NISRA and NRS, respectively, and feed into the UK estimates published by ONS.</p>	<ul style="list-style-type: none"> · Mid-year population estimates (annual) · Small-area population estimates (annual) · Census population estimates (decennial)
Long term migrant	ONS applies the UN recommended definition of a long-term international migrant: "A person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of destination effectively becomes his or her new country of usual residence."	The International Passenger Survey (IPS), plus adjustments for people who change their intentions, asylum seekers and international migration to and from Northern Ireland.	Migration Statistics Quarterly Report: A summary of the latest official long-term international migration statistics for the UK

**Short
term
migrant**

ONS makes use of the UN recommended definition of a short-term international migrant: "A person who moves to a country other than that of his or her usual residence for a period of at least three months but less than a year except in cases where the movement to that country is for purposes of recreation, holiday, visits to friends and relatives, business, medical treatment or religious pilgrimage."

The International Passenger Survey (IPS).

Short-Term International Migration for England and Wales: Visits to and from the UK lasting less than 1 year analysed by the main reason for visiting, age, sex, citizenship and country of residence (annual)

However, to meet broader user needs ONS estimates are currently produced based on three definitions:

- moves made for between 3 and 12 months for employment or study (UN definition)
- moves made for between 3 and 12 months for any reason
- moves made for between 1 and 12 months for any reason

Overseas visitor	<p>ONS applies the UN definition of an overseas visitor: "A person who, being permanently resident in a country outside the UK, visits the UK for a period of less than 12 months. UK citizens resident overseas for 12 months or more coming home on leave are included in this category."</p> <p>There are therefore some overlaps between the definition of short term migrants and overseas visitors applied by the ONS. Estimates of visitors are considerably higher than those of migrants.</p>	<p>The International Passenger Survey (IPS).</p>	<p>Overseas travel and tourism: Visits to the UK by overseas residents, visits abroad by UK residents and spending by travellers (quarterly)</p>
Circular migrant	<p>The ONS does not currently measure circular migration, however the UNECE has set out the concept of circular migration as: "a repetition of migration by the same person between two or more countries".</p>	<p>Not currently measured.</p>	<p>Not currently published.</p>

Source: Office for National Statistics

Useful links:

- [UN definition for population and migration estimates](#)
- [ONS IPS survey background notes](#)
- [UNECE report on defining and measuring circular migration](#)

What do we currently know and what should we measure in future?

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. Similarly, migrants enter and leave the UK for a variety of reasons, stay for different lengths of time 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 is circular migration, which is not covered by the statistics that ONS currently publishes. The government has set out plans for the UK's future [immigration system](#), which includes a new transitional route for temporary short-term workers to come for a maximum of 12 months before a 12-month cooling off period, alongside the continued operation of already established temporary routes, such as the Youth Mobility Scheme and Tier 5 Visas. It is therefore important that ONS considers how our definitions reflect different types of movement into and out of the UK, both now and in the future.

The following section sets out our latest research into how we can use administrative data to measure circular patterns of movement into and out of the UK (of which "circular migration" is a part) and develop new statistics that better meet user needs.

We want your feedback

What additional or alternative definitions would support you in better understanding population and patterns of migration? This might include different population bases, such as daytime populations, public service populations, and so on.

What do we know about circular patterns of movement?

As set out previously, there is no existing ONS approach for producing statistics on circular patterns of movement by people into and out of the UK. However, depending on overall length of stay in the country, some people with circular patterns of movement will fall within our existing definitions of long-term and short-term migrants, whereas others will fall in our definitions of overseas visitors – even if we do not currently refer to them as "circular".

We have therefore taken forward research to better understand these patterns – using Home Office administrative data on non-European Economic Area (non-EEA) nationals who held, or went on to hold, a non-visit visa. This is important as it helps us to assess whether we are including people in the correct definitions of migrants or visitors, consider whether these existing definitions should be maintained or amended, and identify whether any further categories are needed to reflect the complexity of people's travel patterns.

Based on two years' worth of data, we analysed different patterns of movement by looking at the number of journeys people made into and out of the UK and grouped these by frequency. You can find a more detailed report of our research findings in our accompanying [slide pack](#).

What have we learned?

Our main findings from this research are that:

- individuals with a low or medium number of journeys in and out usually stayed in the UK for periods of two to five months at a time and travelled for the purposes of study or family
- people who have a high or very high number of journeys, usually stay in the UK for a period of up to one month a time; we do not have any definitive reason for travel for this group and are conducting further research to understand them, and to try to identify business travellers
- the differences between the two groups helps demonstrate that those with circular patterns of movement are not a homogenous group and that our current definitions do not reflect the complexity of travel patterns that we can identify using administrative data

What will we do next?

We plan to carry out further research to:

- develop our methodology for measuring repeated patterns of movement
- consider whether and how to define circular migration – who should we include and what important aspects should we measure – and how to ensure coherence with other definitions of migration and changes to the UK immigration system
- improve our data coverage for EEA and UK nationals
- consider if and how we should incorporate analysis on circular migrants into our regular outputs

To help us do this in an informed way, we want feedback from our users on what is most important to them.

We want your feedback

How should any grouping and definitions we develop in the future interact with our existing definitions of long-term migration, short-term migration, usually resident population and overseas visitors?

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

As set out in our framework, once we understand the concepts and definitions we are trying to measure, we need to identify the right data sources to measure these. What can we collect from the wide range of administrative sources already held across government? What do we need to collect from other sources, such as surveys, or even commercial data held by other organisations?

What have we learned?

Annex A describes our previously published research showing how linked administrative data could be used to produce outputs on a range of different census topics. We have also published source reports that describe in-depth analysis assessing the statistical quality of the following data sources:

- [NHS Personal Demographic Service \(PDS\) data](#)
- [Department for Work and Pensions' \(DWP\) and HM Revenue and Customs' \(HMRC's\) Benefits and Income Datasets \(BIDS\)](#)
- [NHS Patient Register](#)
- [Electoral Register](#)
- [The English School Census and the Welsh School Census](#)
- [Higher Education Statistics Agency: Student Record](#)
- [DWP and HMRC Customer Information System \(CIS\)](#)

Our published research has enabled us to report on how well the linked sources meet the existing definitions on a range of topics, both nationally and locally. In Annex A, we describe the progress made on producing stocks of the population by age and sex from Statistical Population Dataset (SPD)¹ version 2.0, and some of the known limitations from the combination of data sources used, and rules applied to produce an estimate of the usually resident population. Table 2 shows which data sources we have been using and which sources we will be using (including potential new data sources) in the future.

Table 2: Administrative data sources for transforming population and migration statistics

Use of Data Sources

	Producing estimates of the total population (stock)			Measuring international migration flows (UK)		
	Previous approach (SPD v2.0)		New approach (SPD v3.0)	Potential future use (2019 & Beyond)		
	Identification for Population Stock	Activity	Identification for Population Stock	Activity	Identification of New Migrants	Potential Future Use (2019 & Beyond)
DWP/HMRC Customer Information System						
DWP/HMRC Benefits and Income Datasets						
NHS Patient Register						
NHS Personal Demographic Service						
NHS Hospital Episode Statistics						
England and Wales School Census						
Higher Education Statistical Service student record data						
Births						- for associative linkage
Deaths						
DWP/HMRC Migrant Worker Scan						
Home Office Exit Check Data						
Extended Home Office Administrative Data						
Council tax						
HMRC Real Time Information						
International Passenger Survey						
Other Sources*						

Source: Office for National Statistics

Notes:

1. Other potential sources include surveys citizenship data Electoral Registrations, Driver and Vehicle Licensing Agency (DVLA) data, and private schools data.

As administrative data are not collected for statistical purposes, when we try to use the data to produce statistics that relate to specific definitions (such as those described in [Section 7](#)), 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; these sources are indicated in Table 2 as being used for “identification of population stock” and/or for “identification of new migrants”.

Conversely, other sources are useful as snapshot evidence of presence in the country at a particular moment in time, which together can provide a longitudinal picture of a population change; these sources are labelled in Table 2 as being used for “activity”.

The continuing role of the International Passenger Survey (IPS)

The IPS will continue to have a role in ensuring our outputs remain timely. Administrative sources are often retrospective, that is, they tell us about activity that has already happened. There may be notable time lags before we can use these sources to identify new arrivals to the UK – as people may not register for public services such as health care immediately and consequently, will not be present in the administrative data until they do. We may also need to wait to ensure that new arrivals or departures have been active or inactive in the data sources for long enough to be considered a long-term migrant (as stated in the [UN definition](#)).

The IPS will therefore continue to be essential as a leading indicator of international migration. The survey collects information on the future intentions of individuals moving to the UK and helps provide a timely picture of migration patterns. We can then potentially enhance it using the latest administrative sources, so that we reach our best assessment of migration – reflecting the strengths of what different sources can tell us.

[Section 8](#) describes the likely role that surveys will play to measure and adjust for coverage patterns seen in the data. We will continue to engage with our users as we progress with building our new administrative data-based system, to ensure that our future approach for using IPS and other survey data continues to meet their needs.

What will we do next?

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. This includes continued collaboration with the Home Office to understand how the government’s plans to build a new end-to-end [border and immigration system](#) may provide new opportunities to use administrative data sources to measure international migration.

The next section also sets out the role that surveys will play in the transformation of the population and migration statistics system.

We want your feedback

Are there other data sources that we have not outlined that would add value to our transformation work?

How ONS looks after and uses data for public benefit

ONS has recently published a series of [revised principles and policies on the use, management and security of data](#). These set out how we use, manage and secure data, while providing the best standard of statistical information for the public.

Our aim is to deliver the data and statistics needed to serve the public benefit, and at the same time assure citizens and businesses that they can trust us to safeguard their data. We are doing this by developing a [comprehensive framework](#) to define, manage and govern our data practices, enabling us to use data to inform better decision-making, and at the same time ensure that the data are protected and secure.

We use data from surveys, the census and administrative sources for statistical and research purposes only. In our work, we adopt statistical methods that are professional, ethical and transparent. We follow the principles and protocols for the production of official statistics set out in the [Code of Practice for Statistics](#) and on ethical considerations concerning data, we seek advice from the [National Statistician's Data Ethics Advisory Committee](#).

Notes for: Data sources: what can we use to measure population and migration?

1. A single, coherent dataset 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.

9 . What methods can we use to analyse administrative data?

As set out in our framework, once we have identified the right data sources to deliver a future administrative data-based system, 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 putting administrative data at the core of our official statistics for population and migration in future, but also strengthening of the evidence base on important areas such as the impact of international migration on society and the economy.

Using administrative data to develop a stocks-based approach

Our previous [Administrative Data Census](#) research (outlined in Annex A) focused on our progress in developing a future stocks-based approach, where we combined 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 that 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 main 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” and so on). Developing rules in this way will enable us to be flexible as new administrative sources become available, or as sources change over time. Table 3 shows our initial approach for specific age groups based on our understanding of the coverage and quality of each data source.

Table 3: Outline of our approach for developing data-driven rules for determining usual residency

Age	First hierarchy		Second hierarchy		Ongoing research
	Source	Rationale	Source	Rationale	
0- to 4-year olds	Birth registrations	Covers all births in England and Wales (EW).	Personal Demographic Service (PDS)	High chance of registering non-UK born children.	Aim is to establish 'top of hierarchy' sources for each age group that account for the majority of the population with minimal over-coverage.
		Births to usual residents in EW that occur outside EW are not included.		Will only qualify with evidence of active adult family member at same address.	
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:	Provides evidence of economic activity, including: •periodic benefit payments on or near reference date	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).
	•Pay As You Earn (PAYE) •Benefits datasets (BIDS/NBD)	•annual earnings exceeding expected income threshold			
Aged 65 and over	Likely to be pensions and health registrations				

Source: Office for National Statistics

As an example, our work on the age group 5 to 15 years 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 Department for Work and Pensions’ (DWP’s) Child Benefit data, and have included records from the NHS Patient Register 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).

Other main findings from developing a stocks-based approach using administrative data are summarised in this section.

We are making progress towards a new approach for producing population stocks and flows researching the use of “single data sources” combined with “signs of activity” to determine the usually resident population. This data-driven approach may help improve our coverage issues identified in our research so far (Annex A) and enable us to produce more relevant information on the population.

For deriving population stocks, we have looked at various population age groups and considered particular combinations of data (register based and activity based) to measure the resident population. The next steps will be bringing together learnings across the age groups to refine the rules, for example, where there is activity for parents, we can also assume any dependents are also “active”. We will pull these rules together with the linked data sources to produce a new data-driven stock estimate.

We have incorporated births and deaths registrations into our population stocks and flows. We are currently considering how we can use other information collected on birth registrations (such as mothers’ and fathers’ details) in developing our data-driven rules across other age groups.

Initial analysis is presented in our accompanying [slide pack](#). You can also find a series of case studies that explore how we can use administrative data to identify “activity” for the migrant population in Annex B and our slide packs.

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 10-yearly census. Each year, we age everyone on, use administrative data to add births, remove deaths and make adjustments for internal migration. We use the International Passenger Survey (IPS) 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 births 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 the Patient Register/ (PR) and Personal Demographic Service (PDS), the Higher Education Statistics Agency (HESA), and the Migrant Worker Scan
- we then look across longitudinally linked “activity” sources including HESA, Benefits and Income, PR and PDS, England and Wales School Census and Migrant Worker Scan 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 HESA and Exit Checks data to better understand international student migration in particular

Main findings from using administrative data to develop our future flows-based approach are summarised in this section.

This is the first time we have tried to measure long-term immigration flows using administrative data. This work demonstrates exploratory data-driven rules, which are based on our understanding of the data sources. This is ongoing work and we will be collaborating across the Government Statistical Service (GSS) to develop these rules (particularly around migration).

We have investigated what administrative data sources can tell us about international migration, by linking sources such as immigration, education, health and income records. This work illustrates how each data source has different strengths and can play an important part in understanding international migration, but that no single source can tell us everything our users need to know¹. It is therefore important to link multiple data sources together as each one allows us to pick up different “signs of activity”, such as when migrants interact with public services.

Our new analysis of circular patterns of movement using Home Office Exit Checks data clearly demonstrates the complexity of the travel patterns we can see in the data. Looking at individuals who arrived in the UK on a non-visit visa and their travel patterns for the following two-year period, we were able to identify a range of circular journeys into and out of the country, categorise these into groups and look at their characteristics. For example, those we defined as having a low or medium number of journeys tended to be here for around two to five months and travelled for the purposes of study or family. There is therefore potential to produce statistics on circular migration in future, so we will explore how we can do this based on feedback from our users about what aspects are most important to them.

We also have an improved understanding of what administrative data can tell us about migration from the EU, building on previous work, which focused mainly on non-EU citizens. For example, when we linked PDS data to the Migrant Worker Scan (MWS), we found median lags between arrival and NHS registration of 276 days for EU nationals and 60 days for non-EU nationals in our linked dataset. This indicates that those registering for a National Insurance number (NINo) do not tend to access health services immediately. For EU nationals, it also emphasises that wider data sources will be particularly important for identifying migration into the UK, given the time lags in the health data.

It has also provided new insights into international student migration by linking HESA data with Home Office Exit Checks and HM Revenue and Customs (HMRC) Pay As You Earn (PAYE) data. We found that a greater proportion of EU students in our linked HESA and PAYE cohort were employed compared with non-EU students – which likely reflects immigration rules and the economic background of students – and also provides evidence that our move to bring more sources into our SPD will help improve our coverage for non-EU students.

By linking HESA and Exit Checks data, we also found that almost half of non-EU students in our linked data spent between 300 to 400 days in England and Wales during their first 14 months of study within a 16-month period. Our existing definition of long-term migration counts those who spend 365 days in England and Wales as resident, so this work shows the need to explore what new or additional concepts and definitions would help our users better understand migration patterns.

Our latest research linking the MWS to both benefits and annual PAYE data showed clear potential for earnings data to be an important source for identifying and measuring migration patterns. Our analysis illustrated that four in five non-UK nationals in our linked dataset had signs of activity in income and benefits data following arrival in the country, with more identified in PAYE data. Further linkage work, using more comprehensive PAYE Real Time Information (RTI) data, will therefore be instrumental in helping us to identify patterns of migration to and from the UK.

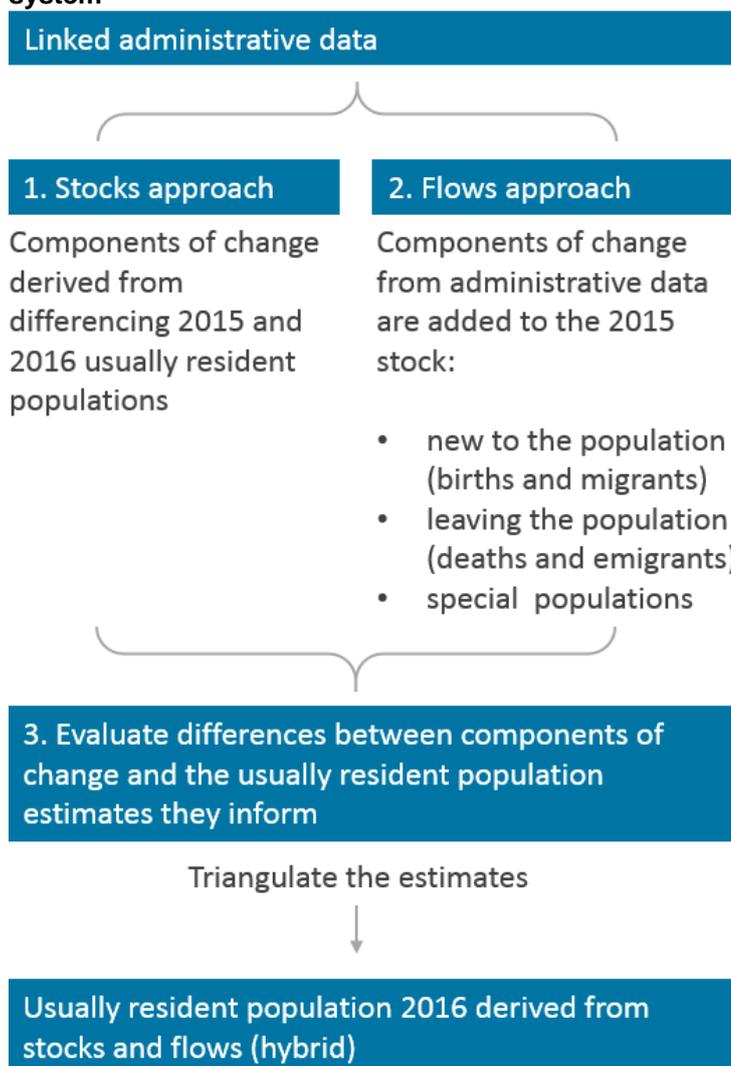
Further analysis is presented in the case studies described in Annex B, and in our accompanying slide packs which can be found in related links.

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 2 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 2: Our proposed hybrid model for delivering a transformed population and migration statistics system



Source: Office for National Statistics

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)).

We have updated on our progress about developing a PCS in our [previous publications](#). 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. This model would provide other vital characteristics of the population, which, along with information from administrative sources, will help us shed light into the impact of different groups of population on society and the economy.

We will ensure that we develop the hybrid model, described previously, 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 important policy and research areas, such as the economic outcomes of international migrants depending on length of stay and age at arrival.

We want your feedback

- Does the hybrid approach seem like a sensible way to produce a coherent set of stocks and flows? Are there alternative approaches we should consider?
- Are there specific methods we should consider when triangulating the two approaches?
- Are there estimation methods that you are aware of, that would enable us to produce a coherent set of stocks and flows?

Thinking ahead to the outputs that we should produce from a new population and migration statistics system:

- Are there specific new outputs that would be helpful to you?
- What impact will any new outputs we may be able to produce have on the way you use population and migration statistics?
- Which aspects of quality (timeliness, frequency, coherence, relevance, accuracy, accessibility and interpretability) are most important to you in the production of population and migration statistics?
- At what geographic level do you need international migration statistics? How important is it to be able to use these at a regional or local authority level?

Notes for: What methods can we use to analyse administrative data?

1. Due to the coverage of the different data sources available to ONS, some case studies cover England and Wales whereas others cover Great Britain or the UK more widely. We will continue to develop our methods and data coverage – in collaboration with the GSS – to produce research at a local level in future.

10 . What might the pathway to transformation over the next five 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 estimates. We will test this survey over the next couple of years, aiming to operationalise a Population Coverage Survey from 2022 onwards.

We will also continue to develop our admin-based outputs for the characteristics of the population and housing. This culminates in a recommendation to government in 2023 about the future of population and housing censuses in England and Wales. Table 4 highlights important milestones along this transformation journey.

Table 4: Our transformation timeline

2019	Share approach for transforming the population statistics system
2020	Administrative data at the core of our evidence on international migration and population
2021	2021 Census
2022	Produce national stock estimates from the 2021 Census, administrative data and coverage survey
2023 and beyond	Recommendation from the National Statistician on the future of the Census

Source: Source: Office for National Statistics

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 important evidence gaps identified by the users of our statistics
- investigating new data sources, including: further Home Office administrative data, PAYE-RTI (and Self Assessment), NHS Hospital Episode Statistics, Further Education data (AEDE), Council Tax and others such as Electoral Register and DVLA registrations; this will improve the coverage of our data, particularly for groups such as EU migrants, as whilst we have improved our knowledge of what administrative sources can tell us about EU and non-EU migration, our existing evidence base is much stronger for non-EU migration
- continuing to consider the role of surveys 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 the GSS, including NRS and NISRA – as we know this is vital to our users
- developing our approach to producing population stocks and flows further, based on feedback from our users; an important 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 more these can tell us 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.
- Carrying out further work to review what existing survey sources tell us about population and migration, including the International Passenger Survey (IPS), the Labour Force Survey (LFS) and Annual Population Survey (APS). These surveys capture information in different ways, so we will publish an update on our work into the comparability between these sources in February 2019 alongside the latest [Migration Statistics Quarterly Report](#), with conclusions following later this year.

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](#). You (users) have the opportunity to help shape our research by feeding back on your requirements.

11 . We want your feedback

Your feedback is important.

We want to hear what our users think about our plans to put administrative data at the core of population and migration statistics.

Throughout this report, we have highlighted the key topics and questions that we want feedback on.

Please send any feedback on these questions to pop.info@ons.gov.uk

Please indicate in your response if you do not wish for the Centre for International Migration (ONS) to keep your details. Your personal information will be stored and processed securely as outlined in the '[Privacy information for our Stakeholders](#)' document.

How will we use your feedback?

We will use your feedback to inform the way that we transform our statistics. Our plan is to publish further research updates on an iterative basis in 2019, taking on board the feedback we receive. As part of these updates, we will continue to ask our users for further feedback on any key developments and decisions along the way.

What other opportunities are there to engage with our transformation work?

We regularly engage with users of our population and migration statistics to keep them informed about our work. Some of our recent engagement activities include:

- Presentations at the Migration Statistics User Forum (which last met on 19 October 2018)
- Presentations at the 2018 British Society for Population Studies (BSPS) and Royal Statistical Society (RSS) conferences and other international conferences
- Our regular ONS Centre for International Migration and Centre for Demography and Ageing newsletters

In spring 2019, we plan to run a series of events and roundtables with key stakeholders, to discuss our latest research findings and gather feedback.

If you would like to get involved with future events or have suggestions for ways we can engage with our users, please let us know by contacting pop.info@ons.gov.uk.

12 . Annex A: Measuring stocks and flows - what progress have we already made?

This Annex sets out a brief overview of the work we have already done to produce estimates on the size and structure of the population as part of the Administrative Data Census project.

Size of the population

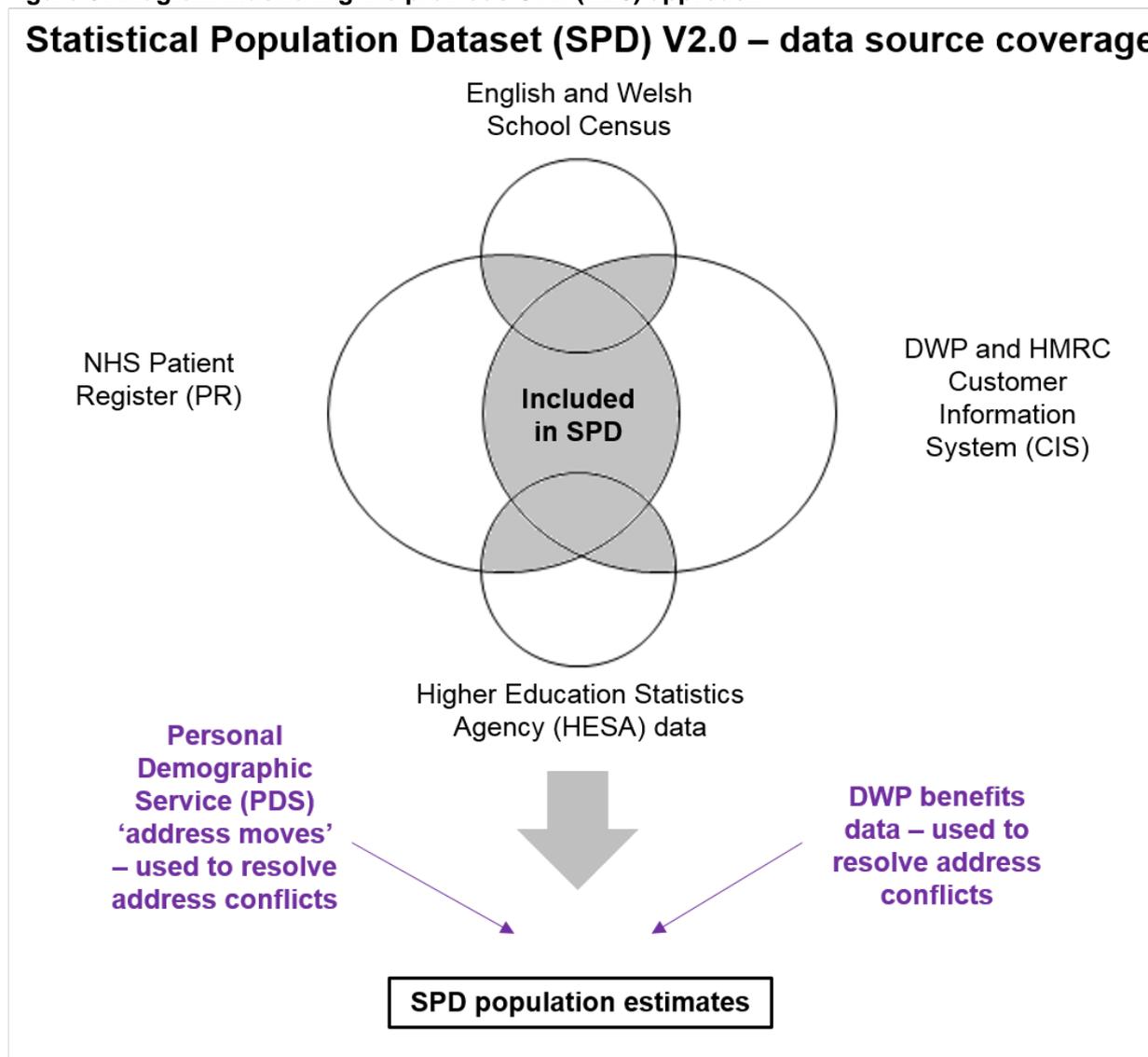
The [Administrative Data Census](#) project was set up to demonstrate progress towards the ambition that “censuses after 2021 will be conducted using other sources of data”. We have published Research Outputs on a wide range of topics usually produced by our decennial census, or topics which have a high need from census users. Previous research can be found here:

- [size of the population](#)
- [population characteristics](#)
- [households and families](#)

An important focus of this research has been on producing estimates on the size of the population using alternative data sources. Our latest [Statistical Population Dataset \(SPD\) version 2.0](#) (v2.0) was created by linking multiple administrative data sources, and applying a set of “rules” to determine records considered to be part of the “usually resident” population. The rules were relatively simple, as shown in Figure 3.

To be included in the SPD as “usually resident”, a record must be found on both the Department for Work and Pensions’ (DWP’s) and HM Revenue and Customs’ (HMRC’s) Customer Information System (CIS), and the NHS Patient Register (PR), unless they are a Higher Education student (found on the Higher Education Statistics Agency data), or a pupil aged 5 to 15 years (found on the English or Welsh School Census). In this case, these records must also be found on at least one of the CIS or the PR. We also use only the NHS PR for those aged zero to four years. We then used activity data from the NHS Person Demographic Service (PDS) and DWP’s Benefits data to resolve address conflicts.

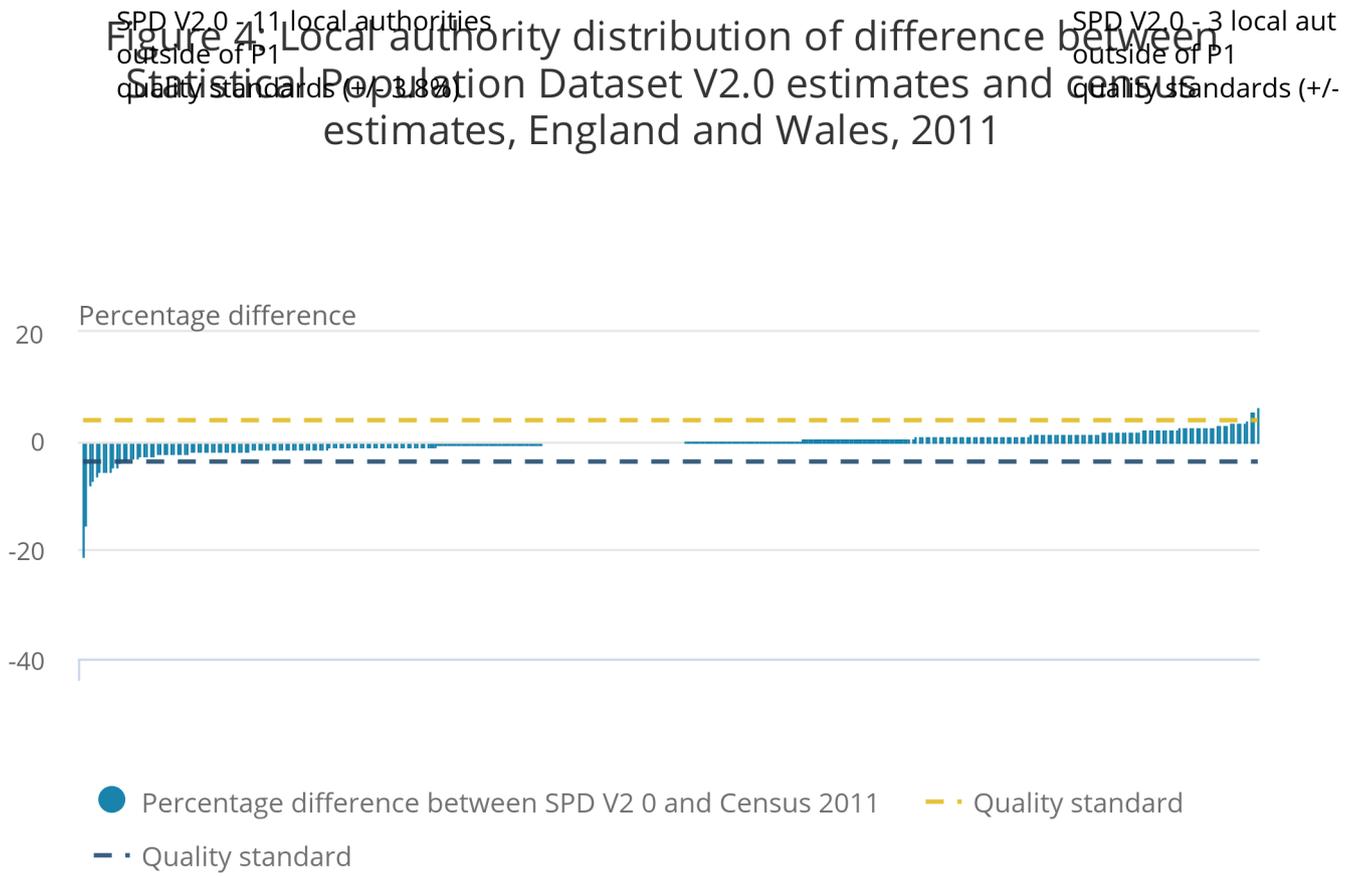
Figure 3: Diagram illustrating the previous SPD (v2.0) approach



Source: Office for National Statistics

We published SPDv2.0 population estimates by five-year age and sex groups down to Output Area for 2011 and 2016. The results showed promise. The method showed that for 96% of Local Authorities, the SPD approach produced total population estimates to a similar level of quality achieved by the 2011 census (they were within +/- 3.8% of the Census estimates).

Figure 4: Local authority distribution of difference between Statistical Population Dataset V2.0 estimates and census estimates, England and Wales, 2011



Source: Office for National Statistics

However, analysis by age and sex showed some coverage patterns which suggest that the combination of data sources and the rules that we applied could be improved for particular age groups.

Figure 5: Comparison of Statistical Population Dataset version 2.0 with the official mid-year estimates by five-year age group and sex, England and Wales, 2011

Figure 5: Comparison of Statistical Population Dataset version 2.0 with the official mid-year estimates by five-year age group and sex, England and Wales, 2011



Source: Office for National Statistics

Notes:

1. Bars to the right of the central line (zero) are those ages and sexes with SPD V2.0 estimates higher than the 2011 Census. Bars to the left are those with SPD V2.0 estimates lower than the 2011 Census.
2. SPD - Statistical Population Dataset.

Figure 5 shows age and sex groups as totals for England and Wales, compared with the 2011 Census. The SPD showed between 2% and 5% over-coverage (compared with the 2011 Census) for males aged 30 to 60 years, under-coverage for females for most age groups, and increasing levels of under-coverage for the oldest age groups. This highlights a limitation of this previous approach for producing an SPD, and shows why we are moving towards an approach whereby we bring more sources together to fill gaps in coverage. [Section 7 of Research Outputs: Estimating the size of the population in England and Wales: 2016 release](#) provides interactive maps to show how the SPDs performed against official population estimates for the 2011 Census at local authority level, by age group and sex.

Some of these coverage patterns may be due to the inclusion of records of those people who have died, left the country or who are short-term migrants in the SPD, or the non-inclusion of records of those people who are international migrants who might not register for multiple services. Therefore, it is important to understand how different groups of the population interact with administrative sources to produce estimates of the population and population change (including international migrants).

13 . Annex B: Case Studies: spotlight on what administrative data can tell us about international migration

We have produced a series of in-depth case studies, which put a spotlight on what linking together specific data sources – such as immigration, education, health and tax records – can tell us about international migration.

This annex provides a summary of the case studies, and links to the slide packs that provide more detail. These case studies provide good evidence that will help us develop our data-driven rules.

Case study: What can linking HESA and Exit Checks data tell us about non-EU international students' departure patterns and length of stay in England and Wales?

We have built on our earlier research into international student migration by investigating non-EU students' departure patterns and their length of stay in England and Wales. How long do undergraduate students tend to remain in the country at the end of their studies? And how long do they spend here during their studies?

We did this by linking two cohorts of Higher Education Statistics Agency (HESA) student records with Home Office Exit Checks data, which gives us information on non-EU travellers arriving into and departing from the UK. This allows us to test the assumption that HESA records provide good evidence that these students are resident in England and Wales and informs our future approach for bringing different data sources together to better measure population and migration.

Main findings

- We matched the majority of records across both datasets (around 70% to 80%) but will carry out more work to understand the student records we were not able to link; for example, our analysis found evidence that certain groups of females may be under-represented in our linked data.
- Most students in our linked data departed the UK on a long-term basis at the end of their studies, around 70% for both cohorts; this pattern is consistent with the findings from the previous analysis we reported in [August 2017](#) and [July 2018](#).
- Combining HESA and Exit Checks data together allows us to identify when non-EU students arrive and depart, and understand their actual travel patterns and we found that:
 - around 10% of graduating non-EU students (that emigrated long-term) leave within a week of their course end date and a further 24% leave between one week and one month of their course end date
 - almost half of students in our linked data spend between 300 to 400 days in England and Wales during their first 14 months of study during a 16-month period; this shows that combined HESA and Exit Checks data are important sources for identifying “signs of activity” that students are resident in England and Wales.
- We believe we can use this linked data to help us understand migration at a local level in future and have set out an illustrative approach for how linked HESA and Exit Checks data can identify departure patterns for graduating students at local authority level; we will carry out work to develop this further.

You can find more details on our methodology and findings in the [slide pack](#) that accompany this report.

Case study: What can linking MWS and PDS data together tell us about when and how new international migrants appear on administrative data sources?

The Migrant Worker Scan (MWS) contains information on all adult overseas nationals who have registered for and been allocated a National Insurance number (NINo) since 2002. A NINo is generally required by any overseas national looking to work, or claim benefits or tax credits, in the UK. By linking this to the Personal Demographic Service (PDS) – which holds information on registrations with health services – we can start to understand when overseas nationals arrive in the country and how long it typically takes them to register for public services.

Our findings tell us about how useful these data sources are for measuring population and migration in the UK at a given point in time, or across certain time periods.

Main findings

- EU nationals tend to register quicker for a National Insurance number (NINo) than non-EU nationals and out of all records in the MWS, 84% had registered for a NINo within a year; the median time lag between arrival and registration for a NINo was 72 days for EU nationals and 135 days for non-EU nationals.
- Non-EU nationals tend to register quicker with the NHS than EU nationals and when we linked PDS data to the MWS, we found a median time lag between arrival and NHS registration of 276 days for EU nationals and 60 days for non-EU nationals; this indicates that those registering for a NINo do not tend to access health services immediately.
- Analysis showed that 41% of potential migrants identified in the PDS could not be linked to MWS as they had not (yet) registered for a NINo; the PDS contains information on new registrations from people overseas, which we can compare with NINo data.

These findings add to the evidence that no single data source can give us the complete picture of who is arriving and resident in the UK. If we used only MWS or PDS data, we would miss certain groups who are either not looking for work or not accessing health services quickly. It is therefore important to link multiple data sources together, to give us a much better understanding of population and migration in the UK.

You can find more details on our methodology and findings in the [slide pack](#) that accompany this report.

Case study: What can linking HESA and HMRC data tell us about employment and economic activity of international students at Higher Education Institutions in England and Wales?

We have built on our earlier research into international student migration, by using administrative data on earnings to investigate economic activity by international students in England and Wales. Do international students work during their studies? How does this differ by nationality?

We took HESA records for a cohort of undergraduate students, then matched these to two years' worth of Pay As You Earn (PAYE) annual records. We used this to identify whether students appear to be earning during their studies.

Main findings

- A greater proportion of EU students in our cohort were employed compared with non-EU students; however, these differences may reflect both immigration rules (in terms of rights to work) and the economic background of students.
- We could see some differences in employment by nationality of students, for example, within EU students, EU8 nationals were most likely to be employed; again, these differences may reflect other factors, including immigration rules and students' economic backgrounds.
- We found around 21,000 (5%) HESA student records in our cohort that did not link to the ONS Statistical Population Dataset (SPD v2.0) and we believe this is likely to be due to the linkage rules, which require that individuals are present on multiple data sources to be included; this finding strengthens our rationale behind changing our approach to developing the SPD, so that this is based on a "hierarchy of belief" rather than our previous approach.

You can find more details on our methodology and findings in the [slide pack](#) that accompany this report.

Case study: What can linking MWS data to HMRC and DWP income and benefits data tell us about activity patterns of non-UK nationals in Great Britain?

As set out earlier in this report, the Migrant Worker Scan (MWS) contains information on all adult overseas nationals who have registered for and been allocated a National Insurance number (NINo). A NINo is generally required by any overseas national looking to work, or claim benefits or tax credits, in the UK.

By linking this to data held by HM Revenue and Customs (HMRC) and the Department for Work and Pensions (DWP) on income and benefits, we can use this to identify "signs of activity" that they are resident in the country. If we can identify groups of migrants that are working or claiming benefits, this provides evidence that they should therefore be included in our estimates of population and migration.

Main findings

- Income and benefits data can help us measure the overall number of international migrants present in Great Britain by providing evidence that they are living here (showing "signs of activity"); our cohort analysis illustrated that four in five non-UK nationals had signs of activity in the income and benefits data following arrival in the country.
- Further analysis found more international migrants were earning through Pay As You Earn (PAYE) than claiming benefits and this is consistent for both EU and non-EU groups, who had similar patterns of activity across income and benefits datasets; this analysis provides additional evidence that, as an indicator of migrants' presence in the country, the MWS needs to be triangulated with income and benefits data to show patterns of arrival, earning and claiming benefits.

These findings add to the evidence that no single data source can give us the complete picture of who is arriving and resident in the UK. If we used benefits or earnings data in isolation, we would miss certain groups who are not present on either data source. It is therefore important to link multiple data sources together, to give us a much better understanding of population and migration in the UK.

You can find more details on our methodology and findings in the [slide pack](#) that accompany this report.