Office for National Statistics

Article

# Dynamic population model for England and Wales: July 2022

The development of new methods to produce near real-time estimates of the population size, using a dynamic population model.

Contact:
Louisa Blackwell
pop.info@ons.gov.uk
+44 1329 444539

Release date:
14 July 2022

Next release:
To be announced

## Table of contents

# 1 . Summary of the dynamic population model

- We are introducing a new dynamic population model (DPM) to estimate population in a timely way, to better respond to user needs.

- The DPM will use a statistical modelling approach to draw strength from a range of data sources such as administrative and survey data; this builds on our earlier research to develop admin-based population estimates.

- The DPM will produce fully coherent estimates of population counts and changes because of births, deaths, and migration.

- To inform the population estimates, we are developing a real-time data dashboard to detect changes in demographic trends as they happen.

- Timely estimates will be provisional, subject to later confirmation as we move from projected to finalised data.

# 2 . Overview of the dynamic population model

## Transforming population statistics

The census has evolved throughout the decades, providing a snapshot and insight every 10 years into who we are and how we live. While the census and mid-year population estimates based on the census provide the best picture of society at a moment in time, how we produce our population and social statistics is changing.

We are using a variety of data sources, to provide more frequent, relevant, and timely statistics. This will allow us to understand population change in local areas this year and beyond.

This is the first in a series of publications over the coming months which will outline our plans to transform population and migration statistics.

## The dynamic population model (DPM)

The DPM is an approach to producing population estimates that:

- uses statistical modelling techniques to bring together a range of data and demographic insights to estimate the population

- builds on our previous research to develop admin-based population estimates

- can be flexible and adaptive to changing data inputs and include data inputs with known limitations (for example, containing missing data, or with known error)

- can formally include demographic trends to inform the model

This approach allows us to use existing data sources, such as cross-border moves between Scotland, Northern Ireland, England and Wales, used in the mid-year estimates (MYE). The approach also allows us to draw strength from new and emerging data sources, such as estimates based on administrative data.

## Ambitions

The DPM is flexible by design, allowing us to incorporate new data, incomplete data and expert demographic intelligence as these become available. This allows us to observe and reflect changes in demographic behaviour, during a time when society is rapidly changing as we emerge from the coronavirus (COVID-19) pandemic.

The DPM will use projection methods and early insight information to produce timely estimates that should be treated as provisional. These estimates will then be revised, replacing projected data with up-to-date data once they become available. This is similar to the approach we use to produce gross domestic product (GDP), where early estimates are provided and then later revised as more data become available.

The DPM will be central to the transformed population and social statistics system that we are building. It will produce the best estimates of the population of England and Wales at a point in time, and other Office for National Statistics (ONS) estimates will align with the outputs from the DPM. This will ensure that all estimates are consistent, including the denominators used for calculating rates such as mortality.

This output will be followed by early research showing progress toward monthly population estimates in Autumn 2022, as we further refine our models. The methods we are developing are regularly reviewed and assured by leading national and international experts in demography and population modelling, for example in the article, Integrated statistical design for the transformed population and social statistics system (PDF, 936KB).

We also welcome feedback from our users and will continue to engage throughout this process. Our ambition is that estimates from the DPM will achieve accreditation as National Statistics.

# 3 . Introducing the dynamic population model

## What the dynamic population model (DPM) will look like

The DPM produces a coherent set of population estimates. For example, the count of a population in an area at a particular time point is calculated as:

- the count of that population at a previous time point

- plus the number of people born and who arrived because of migration

- minus the number of people who died and who migrated out of that area in the intervening period

The DPM is provided with multiple input datasets about population counts and changes because of births, deaths, and migration. It also includes information about the reliability of that data, for example, patterns of under- and over-count. We can also include information about known demographic behaviours. Examples include childbearing patterns by the mother's age, migration patterns by age and sex, and how these can vary in different local authorities. The model then outputs information from which we can derive estimates and measures of statistical uncertainty as a range of plausible values.

## How the DPM is different to existing population estimation

As with the current mid-year estimation process, the cohort component method is at the centre of our population modelling.

The existing method suffers from increasing error the further we move away from the date of the census. However, the DPM addresses this problem by combining independent data on population with changes from births, deaths, migration, and demographic trends. We incorporate a wide range of administrative data and build on the progress in admin-based population research.

This statistical modelling approach allows us to produce more timely and frequent statistics, for example, monthly as well as annual estimates of the population.

Demographic expertise and understanding of trends are formally included in the model, in a transparent way.

# Data used in the DPM

The DPM uses a range of data sources to measure population counts and the components of population change.

One of the inputs to the DPM are the admin-based population estimates (ABPEs) which we will now refer to as Statistical Population Datasets (SPDs). This reflects that they are not a finished estimate and will now be an important contributor to the DPM where the estimation is carried out. This means we can use the benefits of the SPDs with measures of uncertainty alongside other data inputs to estimate the population.

Data sources used include:

- births – birth registrations

- deaths – death registrations

- international migration - estimates of international migration

- internal migration - estimates of internal migration from the mid-year population estimates

- cross-border moves - estimates of cross-border moves from the mid-year population estimates

- population measures – Census 2011, estimates from Statistical Population Datasets version 3, NHS General Practice Patient Register, NHS Personal Demographic Service data of individuals who access NHS services

Our ambition to produce timely estimates requires timely "signal" data to give us early notice of emerging trends, such as a post-coronavirus (COVID-19) pandemic drift of people returning to live in inner city areas. To provide this timely information we have created a real-time data dashboard.

Visualising and comparing recent and historic patterns in the dashboard allow us to spot unexpected changes in the data. We are exploring new and innovative "signal" data to include in the dashboard, including mobile phone usage, wastewater, and energy consumption such as electricity or gas.

# What the DPM will produce

The following examples illustrate the flexibility of our DPM. We have combined data from similar local authorities to create a synthetic local authority. We compare two sets of modelled outputs, calculated under alternative assumptions. This synthetic local authority has a population distribution that is typical for areas that contain a university, with a concentration of university-age students.

In both examples we use the same input data. The synthetic estimates of population counts by single year of age and sex are:

### Mid-year estimates

Synthetic mid-year estimates (MYE) for June 2011, which was a census year. These are considered the most accurate estimate of population counts used as input data for the model. Synthetic MYE after 2011 are not included in the model but are shown in the figures below for comparison against the DPM modelled estimates.

### Statistical Population Dataset

The synthetic SPD is available at annual intervals from June 2016 to June 2020. This dataset is assumed to be less accurate than MYE in a census year, but more accurate than the Patient Register (PR) and Personal Demographic Services-based estimates of the Patient Register (PDSPR).

### Patient Register

Synthetic PR data are available at annual intervals from June 2011 to June 2020. This dataset is assumed to be less accurate than the SPD but is useful to include as it covers years where SPD data are not available.

### Personal Demographic Service-based estimate of the Patient Register

The synthetic PDSPR is available at annual intervals from June 2016 to June 2021. This dataset is assumed to be the least accurate estimate of population counts but does have the benefit of providing an estimate of population counts to June 2021.

### Data on population change

Synthetic counts of births and deaths are used in the model and are available for 2012 to 2021, along with fertility, mortality, and migration rates for 2012 to 2022. The demographic rates for 2012 to 2021 were calculated from synthetic counts of the events (births, deaths, and migration) and synthetic MYE for 2011 to 2021. These raw rates are then smoothed, as the DPM only requires the underlying rates. The rates for 2022 are forecast outside of the model. These rates are used in the model for estimating provisional counts of population and population change up to June 2022.
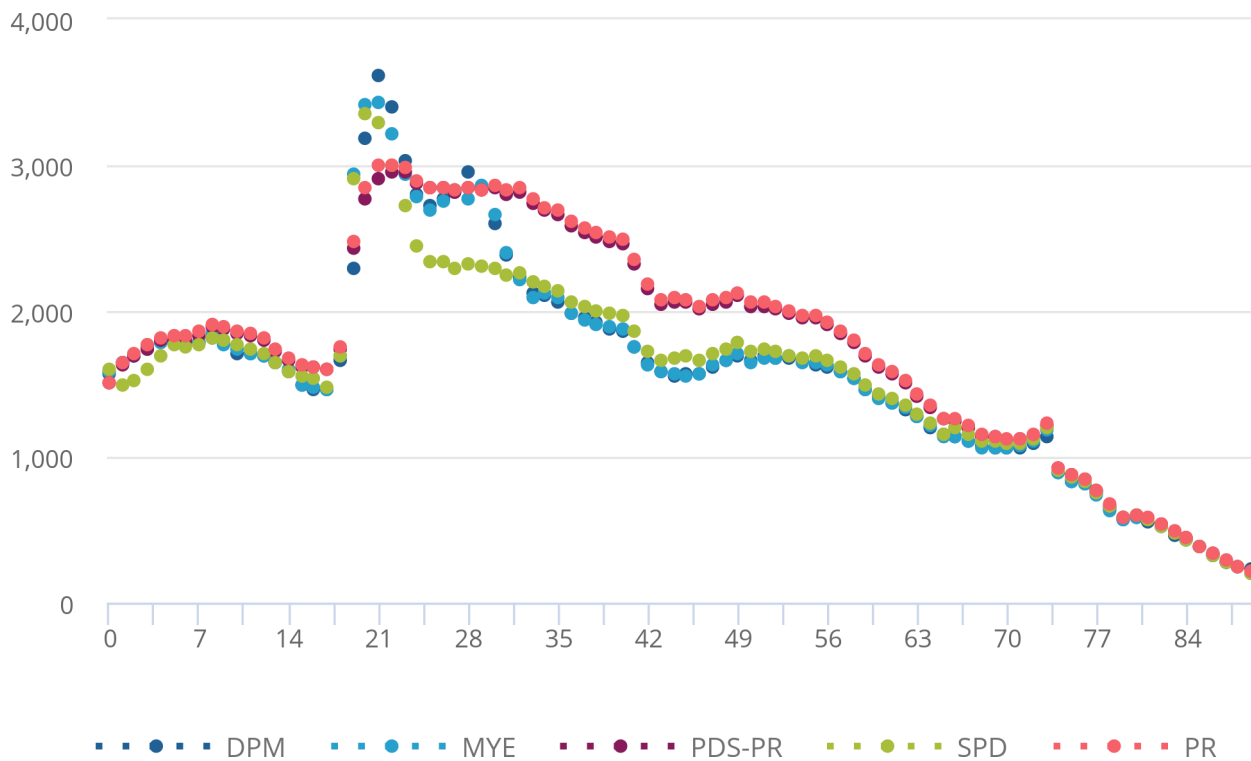
### The effect of alternative assumptions about input data quality on DPM estimates

The following two examples demonstrate how population and migration estimates from the DPM change when assumptions about the precision of the SPDs are altered. This shows comparisons that can be made in the model building process. As we further develop the models and publish estimates, we can use measures of quality of the input data to inform these assumptions, which will be clearly described and justified.

## Example 1: model assumes the SPD is not precise

**Figure 1: Example 1 population estimates for males by single year of age for synthetic local authority in June 2020**



Figure 1: Example 1 population estimates for males by single year of age for synthetic local authority in June 2020
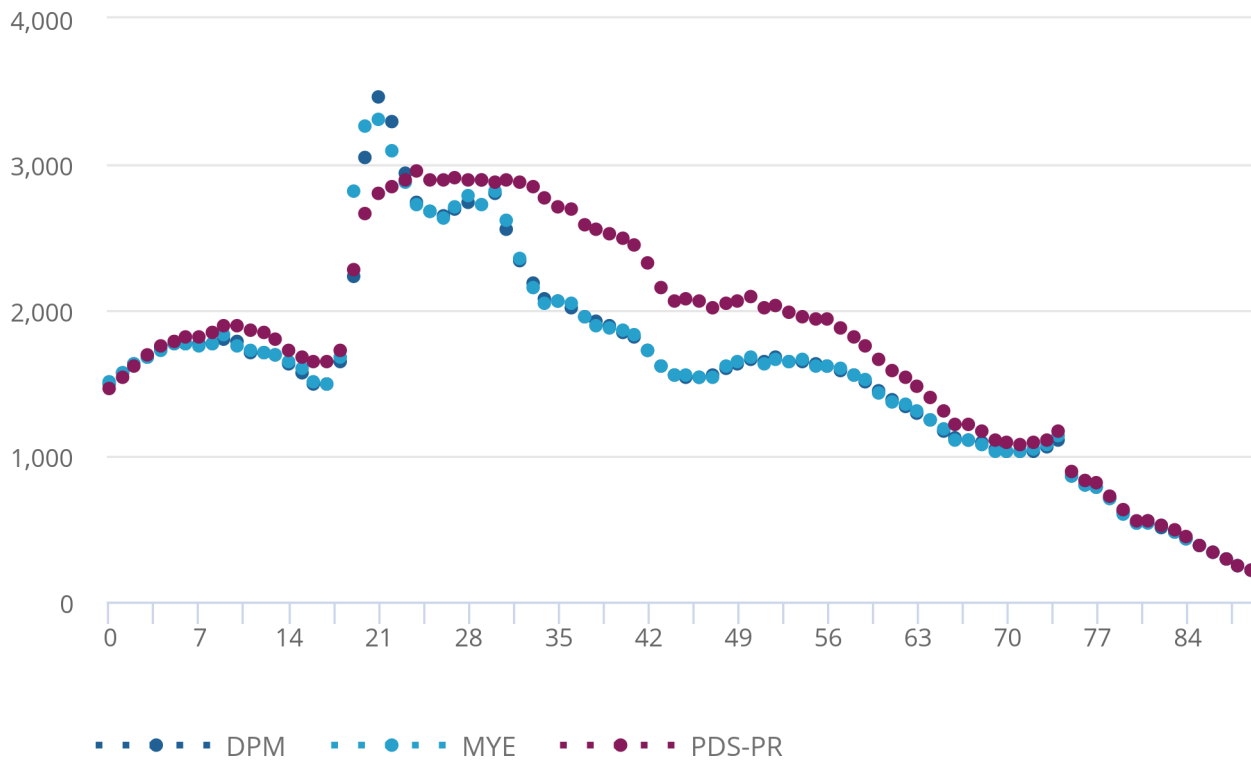
**Source: Office for National Statistics - Dynamic population model**

Figure 1 compares DPM estimates for males in June 2020 by single year of age. Here, we assume SPD input data are not precise (but more accurate than PR and PDSPR). DPM estimates are similar to the synthetic MYE for most ages. There are large differences between DPM estimates and the SPDs for ages 19 to 31 years. Modelled estimates are very different to PR and PDSPR around student ages (those aged 18 to 23 years). This is because the PDSPR and PR reflect the delays in registering with a doctor when students move to university. There is then overestimation in the PDSPR and PR after many students leave university. This is because they often delay registering with a new doctor when they move away from university.

Figure 2 shows the DPM estimates for males in June 2021 by single year of age. Although the only input data for population counts in June 2021 is PDSPR, DPM estimates still tend to be closer to the MYE.

Figure 2: Example 1 population estimates for males by single year of age for synthetic local authority in June 2021
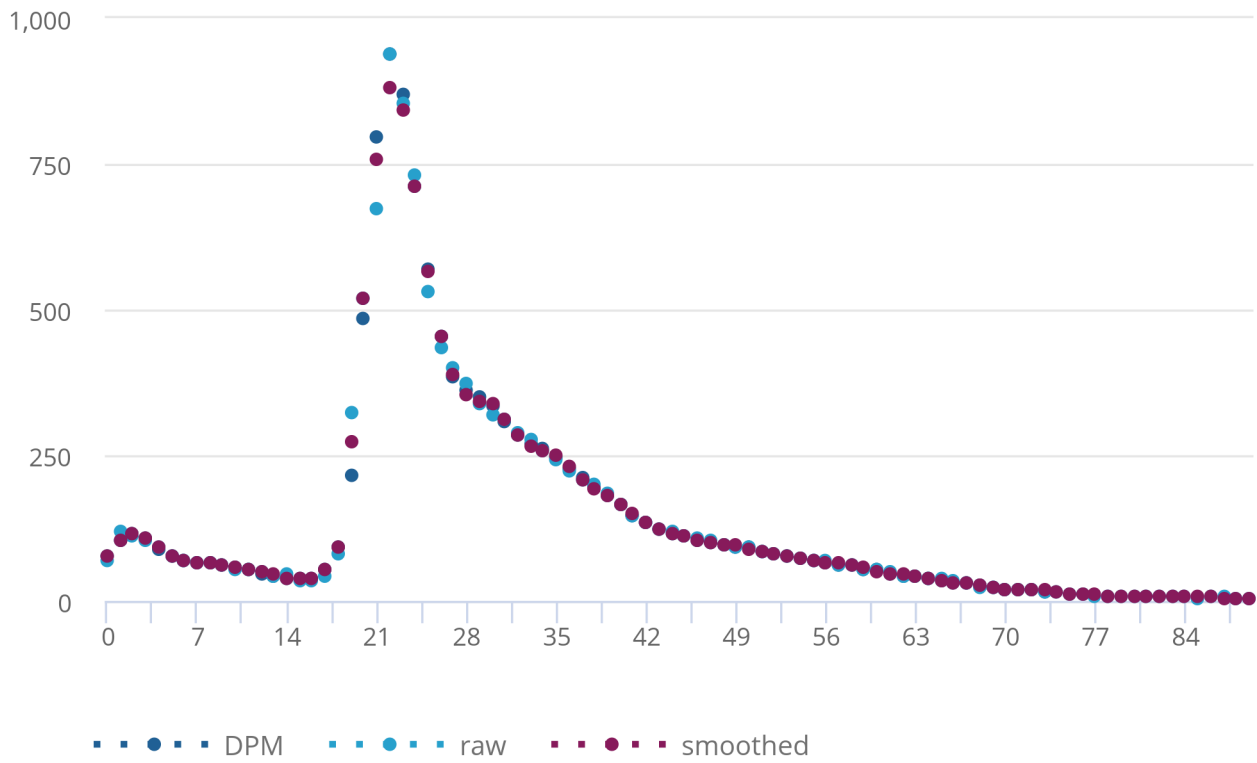


**Source: Office for National Statistics - Dynamic population model**

There are notable differences between MYE and DPM estimates for those aged 19 to 22 years. This is because the model uses smoothed migration rates while the synthetic MYEs use migration counts. Because we assume SPDs and PDSPR are not precise, the DPM estimates for June 2021 are driven by migration rates.

Figure 3 shows emigration estimates from the DPM for males, alongside raw and underlying smoothed emigration data. Modelled estimates are very similar to (smoothed) input migration data. This is because we assume that the population data for 2011 onwards are not precise.

Figure 3: Example 1 emigration estimates for males in synthetic local authority for June 2021
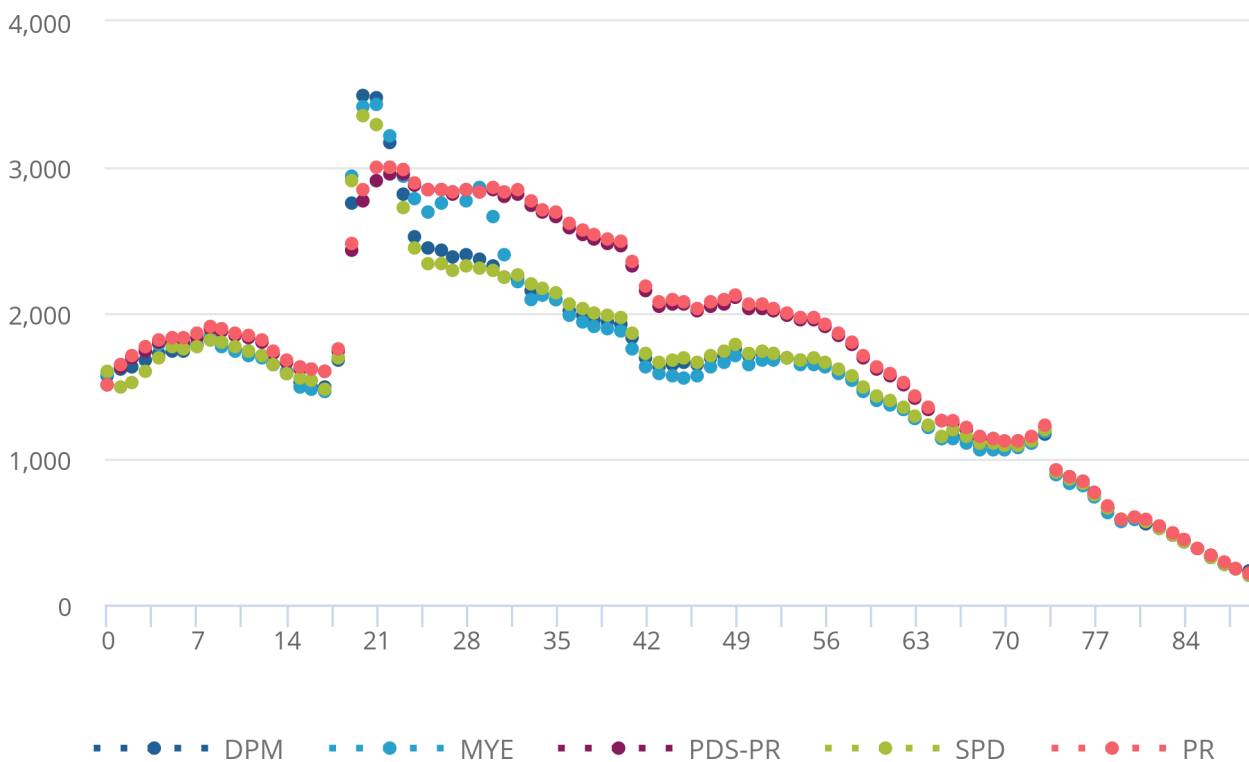
**Source: Office for National Statistics - Dynamic population model**

## Example 2: model assumes the SPD is precise

Figure 4 shows population estimates for males in June 2020 by single year of age when the SPD input data are assumed to be precise. Compared with Example 1, the modelled estimates are further from the synthetic MYE and much closer to SPD.

**Figure 4: Example 2 population estimates for males in synthetic local authority for June 2020**

Figure 4: Example 2 population estimates for males in synthetic local authority for June 2020
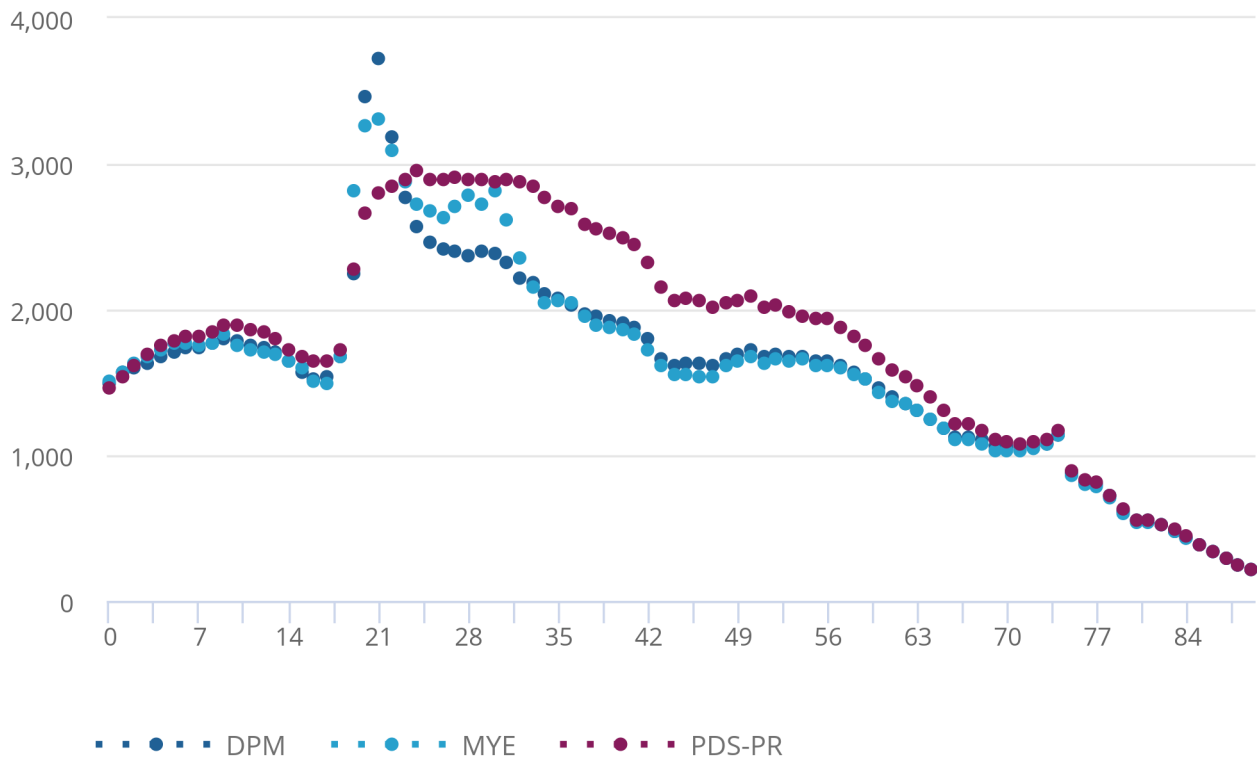


**Source: Office for National Statistics - Dynamic population model**

The DPM estimates for June 2021 shown in Figure 5, are again further from the synthetic MYE. This leads to larger differences between the emigration estimates and input emigration data, as seen in Figure 6. Because we assume SPDs are precise, the DPM adjusts migration to make them coherent with population counts.

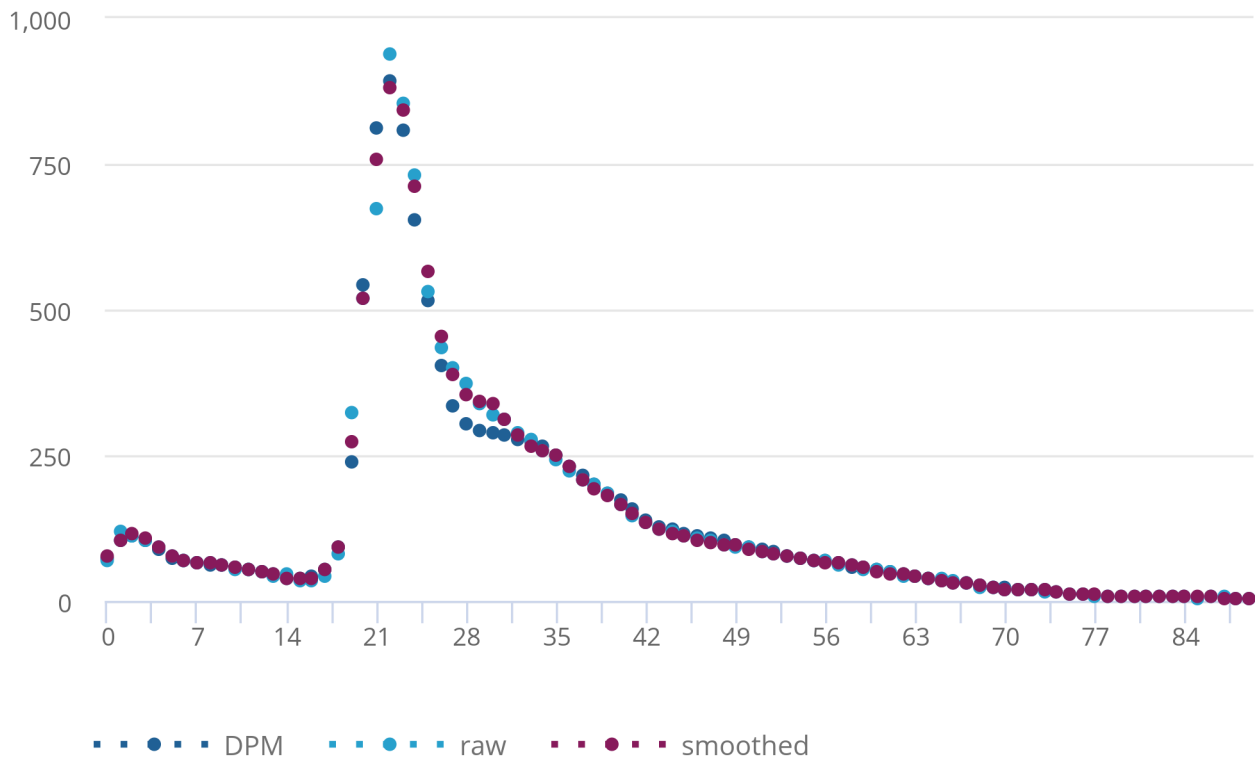Figure 5: Example 2 population estimates for males in synthetic local authority for June 2021



**Source: Office for National Statistics - Dynamic population model**

Figure 6: Example 2 emigration estimates for males in synthetic local authority for June 2021



**Source: Office for National Statistics - Dynamic population model**

## Estimating to June 2022

We have no input data for population counts in June 2022, because of data availability. We therefore forecast migration, fertility, and mortality rates for June 2021 to June 2022. Figure 7 compares population estimates for males and females from the two example models in an interactive population pyramid for June 2011 to June 2022. The differences between the two estimates increase over time. This is because SPD is only available from June 2016, and so the impact of assuming it is precise is most notable from 2016.

**Figure 7: Population pyramid comparing Example 1 and Example 2 for synthetic local authority by age and sex between 2011 and 2022**

**Download this chart**

.xlsx

These two examples demonstrate that we can produce coherent population estimates by single year of age and sex at local authority level. These are shown for June 2011 through to June 2021 and provide provisional estimates for June 2022. Varying the assumptions about quality of the SPDs changes the DPM modelled estimates of population. This, in turn, generates differing estimates of migration, ensuring a coherent set of population estimates.

# 4 . Glossary

## Dynamic population model

A dynamic population model (DPM) is a statistical modelling approach that uses a range of data to measure the population and population changes in a fully coherent way.

## Administrative data

These are data that people have already provided to government, for example, when accessing public services. Some of these data could be re-used by the Office for National Statistics (ONS) to produce statistics about the population.

The ONS has been using administrative data for many years. For example, we use annual births and deaths statistics, as well as NHS patient registrations, to roll forward the population estimates between censuses.

## Statistical Population Dataset (SPD)

A 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.

## Patient Register (PR)

The Patient Register from NHS Digital contains a list of all patients who are registered with a General Practitioner (GP) in England and Wales.

## Personal Demographic Service (PDS)

The Personal Demographic Service (PDS) from NHS Digital is a national electronic database of NHS patients, which contains only demographic information with no medical details. The PDS differs from the PR, since it is updated more frequently and by a wider range of NHS services. The PDS data available to the ONS consist of a subset of the records, including those which show a change of postcode recorded throughout the year or a new NHS registration.

## Local authority

The general term for a body administering local government services.

In England, local government is administered by either single tier or two-tier local authorities. The single tier authorities comprise unitary authorities, metropolitan districts and London boroughs, though some services such as transport planning are carried out by the Greater London Authority. The two-tier authorities elsewhere comprise counties and non-metropolitan districts.

In Wales, there are single tier unitary authorities.

## Communal establishment

An establishment providing managed residential accommodation. "Managed" in this context means full-time or part-time supervision of the accommodation.

# 5 . Future developments

## Methods and research

Our research has focused on implementing these methods, to reduce run-time and make the dynamic population model (DPM) a feasible way to estimate the population at local authority level. Now that we have achieved this, we are making several improvements, including:

- extending the DPM to produce monthly population estimates to enrich our understanding of demographic behaviours, including seasonal patterns of population and population change that vary for different local authorities; local authorities with large residential educational establishments are a typical example where these insights will support service and resource planning

- developing our models to account for uncertainty in the local authority level data

- further incorporating demographic knowledge and real-time emerging trends from alternative data sources

- for every local authority, assessing how well the DPM estimates special populations such as armed forces, school boarders and others living in communal establishments

- engaging with local authorities to ensure that local knowledge, insights, and data are embedded in the DPM; we are developing a framework for receiving user insight into local population levels and change, which will build on the processes used for local authorities' input to quality assurance of census

- continuing to develop our algorithms to improve the efficiency of the system and reduce run-time

- as real-time estimation requires more timely data feeds and data engineering, processing, and analysis at pace, we are working with other government departments and additional data suppliers to gain access to more frequent and timely data to suit the DPM requirements

- improving measures of statistical uncertainty for component datasets, including a detailed comparison between the 2021 Statistical Population Dataset (SPD) and Census 2021; from this we will develop measures of population under- and over-coverage for the SPDs, which will improve our modelled estimates

In autumn 2022, we will publish our research on provisional June 2022 population estimates for a selection of local authorities.

In winter 2022, we will publish our research on provisional June 2022 population estimates for all local authorities.

We plan to make the software that we are developing open source (publicly available) for transparency and to invite recommendations for further, continuous improvement.

## Feedback

We welcome your feedback on the DPM, our transformation journey, and our latest progress and plans. If you would like to contact us, please email us at pop.info@ons.gov.uk.

You can also sign up to email alerts from the Office for National Statistics Population team for updates on our progress, and to hear about upcoming events and opportunities to share your views.

# 6 . Related links

Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2020
Bulletin | Released 25 June 2021
National and subnational mid-year population estimates for the UK and its constituent countries by administrative area, age and sex.