

# Dynamic population model, improvements to data sources and methodology for local authorities, England and Wales: 2011 to 2022

Developments of methods and data used in the dynamic population model

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

Release date:  
28 February 2023

Next release:  
To be announced

## Table of contents

1. [Main changes](#)
2. [Overview of the dynamic population model](#)
3. [Summary of the dynamic population model estimation process](#)
4. [Census 2021 as an input to the dynamic population model \(DPM\)](#)
5. [Improvements to data sources](#)
6. [Improvements to methods](#)
7. [DPM estimates with improved methodology for local authority case study areas](#)
8. [Glossary](#)
9. [Future developments](#)
10. [Collaboration](#)
11. [Related links](#)
12. [Cite this methodology](#)

# 1 . Main changes

We continue to develop our research into the new dynamic population model (DPM). The DPM will estimate population and population change in a timely way, to better respond to user needs.

The outputs from the DPM include admin-based population estimates (ABPEs) for mid-year 2011 to 2022 for all 331 local authorities in England and Wales.

Since our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#), we have taken on feedback from users and included several improvements to the DPM estimation process. Improvements can be summarised as:

- updated and improved data sources
- included separate measures of statistical uncertainty for all of our primary population stock data sources
- estimated coverage ratios for administrative data that change over time
- improved smoothing of flow rates over age and time that better capture sudden peaks in migration, such as students aged 19 years

Results using these improved methods are published in our [Admin-based population estimates: provisional estimates for local authorities in England and Wales, 2011 to 2022 article](#).

These are not official statistics. They are estimates from a new methodology which is different from that currently used to produce official population and migration statistics. The information and research in this article should be read alongside the estimates to avoid misinterpretation. These outputs must not be reproduced without this warning.

## 2 . Overview of the dynamic population model

The dynamic population model (DPM) provides a coherent statistical framework for more timely population statistics. The current system using census has evolved over time, providing a snapshot every 10 years into who we are and how we live. The census and our census-based mid-year estimates (MYE) provide the best picture of society at a moment in time. However, the coronavirus (COVID-19) pandemic underlined the need for more timely population estimates and we are committed to maximising the use of administrative data. We are researching new ways to produce population and social statistics.

We introduced the DPM in July 2022 as our future proposal for producing timely, coherent population statistics. In our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#), we provided provisional population estimates for 14 local authorities. In this article, we describe changes made to the methodology and data sources since November.

Our [Admin-based population estimates: provisional estimates for local authorities in England and Wales, 2011 to 2022](#) presents admin-based population estimates (ABPE) derived from the DPM for all 331 local authorities in England and Wales. The estimates were calculated using the updated methods and data that this methodology describes.

Comparisons between census-based and admin-based estimates for 2021 are discussed in our [Transforming population statistics, comparing 2021 population estimates in England and Wales article](#). It provides some guidance on how best to interpret and use each of the estimates.

Our [Population statistics sources guide](#) helps users find the right population statistics for them.

### 3 . Summary of the dynamic population model estimation process

This section summarises the method used to produce our provisional admin-based population estimates (ABPEs), covering mid-year 2011 to 2022.

The DPM uses a demographic accounting framework. At its core is a set of internally consistent estimates of population, births, deaths, and migration by age, sex, geography, and time.

The first step in producing the population estimate is to approximate the components of the demographic account: in-migration, out-migration, and population stocks. Births and deaths are treated as known, in the current method.

These components are then used to estimate expected trends in fertility, mortality, and migration rates. These estimated rates and associated measures of uncertainty approximate what a skilled analyst would know about such demographic trends.

We also estimate statistical models for population stocks. They incorporate the variability we see in the data because of systematic inaccuracies, including coverage and reporting error. These statistical models approximate what a skilled analyst would know about the quality of the data sources.

Next, we estimate the individual demographic accounts for each local authority. A technique called particle filtering, or sequential Monte Carlo, is used to generate many estimates (for example, we used 10,000) to represent the probability distributions of these counts. This filtering approach is applied independently to each year of birth cohort by sex within each local authority. To produce point estimates, we use the mean of these probability distributions. To produce aggregate counts, we sum the means.

For a more detailed description of the estimation process, see our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#).

## 4 . Census 2021 as an input to the dynamic population model (DPM)

We did not incorporate the Census 2021 data into the DPM in our previous publication. This allowed us to compare our results with the actual census data and assess the accuracy of the model. In this paper, we have created three versions of admin-based population estimates (ABPE).

Our ABPE best estimates from the DPM include the results from Census 2021, rolled forward to mid-year, as accurate information on the population in 2021. These are referred to as Census 2021-based mid-year estimates (MYE). The Census 2021-based MYE, rolled forward from Census Day (21 March 2021) to mid-year (30 June 2021), are used as a population stock in 2021 and for estimating coverage adjustment in other population stock datasets.

Within the model, the coverage adjustment affects not only the 2021 population estimates, but also prior years. We assume a linear change in the coverage ratios by age and sex within local authorities between 2011 and 2021.

We also explore scenarios in which traditional census data are not available. We have created two further model versions.

Our second version, called ABPE future estimates, demonstrates the potential quality of the ABPEs assuming that a reliable coverage adjustment strategy is in place, and that there is no longer a census. This is indicative of what our future system could look like. For this publication, we use a proxy coverage adjustment strategy derived from census to demonstrate this, while our coverage adjustment methods are under development. We use the Statistical Population Dataset version 4.0 (SPD version 4.0) as population stock in 2021 and the smoothed ratio of SPD version 4.0 to Census 2021-based MYE for the coverage adjustment. The coverage adjustment assumes a linear change over time between 2011 and 2021.

The third version, called ABPE basic estimates, is based on similar methods to those used in our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#). It does not use the Census 2021-based MYE as input into the model. This version demonstrates how the DPM model performs without access to updated census information or up-to-date coverage adjustment data. This version is useful to compare with the current MYE rolling forward from 2011. The new SPD version 4.0 are used as population stocks in 2020 and 2021. We assume that the coverage ratio between the SPD and the census, estimated using 2011 data, remains consistent throughout the time period from 2011 to 2021.

## 5 . Improvements to data sources

The dynamic population model (DPM) uses a range of sources to measure population and the components of population change. We have improved and updated several data sources used in the model. Assumptions made in our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#) about individual data sources remain the same, unless expressly discussed in this article. Admin based population estimates (ABPE) and data sources used as inputs refer to the population at mid-year (30 June 2021).

## Statistical Population Dataset

Similarly to our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#), we use the Statistical Population Dataset (SPD) as our main source of population stock data. The SPD are admin-based population counts, derived from a set of inclusion rules to approximate the usually resident population. The new Statistical Population Datasets version 4.0 (SPD version 4.0) improves on the known undercoverage of Statistical Population Datasets (SPD) version 3.0 (SPD version 3.0) in our [Developing admin-based population estimates, England and Wales: 2016 to 2020 article](#). It does this by adding new sources and refining the inclusion rules. The construction of SPD version 4.0 is described in our [Developing Statistical Population Datasets, England and Wales: 2021 article](#).

SPD version 3.0 provides population stock data for 2016 to 2019 and SPD version 4.0 for 2020 to 2021. Here, we have used an early version of SPD version 4.0 for 2020 and 2021, which will differ from the results published in our [Developing Statistical Population Datasets, England and Wales: 2021 article](#). SPD version 4.0 data for 2016 to 2019 and updated versions of 2020 and 2021 were not available in time to be included in this publication but will be included in future versions.

The previous publication did not use actual stock data for 2021; instead, it applied population flow rates to the 2020 SPD, to predict the 2021 population. In this publication, we replace the predicted 2021 population with 2021 SPD data. In the absence of a 2022 SPD, we apply population flow rates to the 2021 SPD, to provide timely predicted population figures for 2022.

SPD data have been rounded to ensure that no small counts are identifiable. This has a particularly significant impact on age and sex combinations where counts are low, such as for people aged 95 years and above, or in local authorities with very small populations.

## Patient Register

The NHS GP Patient Register (PR) remains our stock data for 2012 to 2015, as the SPD is not available for this time period.

Now that we have expanded the DPM to cover all local authorities in England and Wales, we have corrected for the following two issues with the PR data that affect the counts in some local authorities in specific years.

In 2013, the PR count in Wolverhampton decreases substantially, which is believed to be an issue with the data.

In 2014, local authorities in Cornwall and Devon see large decreases in the PR counts relative to past and subsequent years. This is because of localised processes to remove individuals considered to no longer be living in the area from GP lists.

Sudden decreases in the PR count because of administrative processes, such as list cleaning, can lead to undercount in the ABPE. This will also be compounded by the coverage adjustment that accounts for the known overcount in the PR data.

We have corrected for these issues by removing the affected PR data and treating them as missing. In these instances, we rely entirely on the flows to estimate population in these local authorities. The rates for Wolverhampton are also adjusted to account for missing problematic years (see [Section 6: Improvements to methods](#)).

## Internal migration and cross-border moves

The internal migration and cross-border movements are based on the components of change used in the official mid-year population estimates from 2011 to 2021.

To estimate internal migration in 2022, we used the Personal Demographic Service (PDS). These PDS-based estimates have been refined to provide more detailed data, including further age breakdowns, up to age 105 years and above.

To ensure consistency with the rest of the time series, we scaled the PDS-based migration in 2022 using the ratio of mid-year estimates (MYE)-based to PDS-based migration in 2018 and 2019 for each age, sex, and local authority combination. This accounts for some of the migration that was not captured by using PDS data alone. We used migration data for 2018 and 2019 because later data were affected by the coronavirus (COVID-19) pandemic. These provisional imputed estimates are not comparable with previously published estimates.

In the provisional estimates shown in our previous publication, we observed that the patterns seen in the PDS-based imputed migration were likely influenced by registrations for the COVID-19 vaccination program. We expect that the booster program for the vaccine will continue to affect PDS-based migration in 2022.

The estimates of internal migration will be updated in June 2023, when the MYE components for 2022 become available.

## International migration

Our previous provisional estimates used a combination of [Long-Term International Migration \(LTIM\)](#) for 2012 to 2020, [experimental modelled estimates](#) for 2021 and forecasts for 2022.

The modelled estimates did not contain data for people aged under 16 years, because children are not captured on the primary data source used for producing these estimates. We estimated child migration for 2021 and 2022 by scaling to the migration patterns for people aged 35 to 60 years. This age group was selected because the age distribution at a national level had similar proportions between the LTIM and experimental estimates, which suggested good coverage.

The scaling was based on a ratio of the 2012 to 2020 average total recorded moves for this age group in the LTIM data to the total moves for the same age group in the experimental modelled estimates data, for 2021 and 2022 independently. The ratios were then applied to the 2012 to 2020 average counts of people aged under 16 years, to impute child migration in the experimental modelled estimates data for 2021 and 2022.

For the current estimates, international data from 2012 to 2020 remain the same. This includes replacing emigration data for males aged 88 years and over in 2020 with data for 2019 because of known quality issues.

International migration by age, sex, and local authority are included as a component of processing the MYEs, with the latest release covering up to year ending June 2021.

Data were disaggregated, using Department for Work and Pensions (DWP) Registration and Population Interaction Database (RAPID). They included adjustments for international students and people under the age of 16 who have registered from overseas:

- Higher Education Statistics Agency (HESA) data were used to account for international students who are not working
- NHS Personal Demographic Service (PDS) data were used to account for anyone under the age of 16, as children are not captured in the RAPID data

These data are combined and converted into proportions (for example, the number of 18-year-old males in a given local authority may represent 1% of the data). These proportions are applied to overall immigration and emigration estimates to provide a breakdown by age, sex, and local authority. For the previous example, an overall immigration estimate of 100,000 will be apportioned to 1,000 18-year-old males in a particular local authority.

Data used for disaggregation are only available up until the year ending March 2021. For data up to the year ending June 2022, the same rates of migration for age, sex, and local authority are used from the 2021 MYE and applied to overall migration estimates. These were published in our [Long-term international migration, provisional: year ending June 2022 bulletin](#).

These estimates are not consistent with the estimates published in our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#) because of data availability. The local authority case study migration data was based on modelled estimates up to the year ending June 2022, which use fewer sources and are only indicative of migration trends. Estimates published alongside our [Admin-based population estimates: provisional estimates for local authorities in England and Wales, 2011 to 2022](#) are based on migration data published in our [Long-term international migration, provisional: year ending June 2022 bulletin](#). The bulletin uses administrative data to observe actual behaviours to measure international migration.

## Births and deaths

The DPM data for birth and death registrations are largely unchanged from our provisional estimates used in our previous DPM publication. However, there have been some small additions to the number of recorded births and deaths because of delays in registration. Birth registrations were temporarily paused in March 2020 because of the ongoing coronavirus (COVID-19) pandemic in 2020. From June 2020, registration services restarted where it was safe to do so. Consequently, our [User guide to birth statistics](#) show that 2020 births registrations came in much later than in normal years, with 42% arriving after 42 days (the usual legal limit). We continued to see delays in birth registrations in 2021, with 26% arriving after 42 days. Consequently we extended the usual cut-off date for including birth registration in the 2020 and 2021 data.

There are some differences in the births registration data published in our [Births in England and Wales: 2021 bulletin](#), the MYE components, and the births data used in the DPM. The largest difference was in 2020. Differences can largely be attributed to the different time periods used. Published birth registrations data are calendar year annual statistics that include births occurring in the reference year that were registered by 25 February the following year (the usual cut-off data for inclusion to allow late registrations). The DPM births data cover 1 July to 30 June.

There are small differences between previously published Office for National Statistics (ONS) mortality statistics, such as [Weekly death registrations data for England and Wales](#), the MYE components, and the data used in the DPM, for several reasons. Some deaths may be missing from one publication or another because of lags in registration and coroner-related delays, as explained in our [Impact of registration delays on mortality statistics in England and Wales: 2020 article](#). Some differences occur because the DPM uses date of death occurrence, rather than date of registration as used in most regular mortality statistics. Additionally, regularly published deaths data are compiled weekly, monthly, and for calendar years, whereas DPM deaths data cover the twelve months from 1 July to 30 June.

## 6 . Improvements to methods

Since our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#), there have been several improvements to the dynamic population model (DPM) methods. This section outlines the improvements made to the coverage adjustment, production of rates, and estimates of uncertainty.

## Coverage ratios

In the previous publication, the model used population stock data from the Patient Register (PR) and Statistical Population Dataset version 3 (SPD version 3.0). However, we recognised that these stock data were biased. To address this, coverage ratios were calculated based on the ratio between these stock data and mid-year estimates (MYE) for the 2011 Census. The coverage ratios were then smoothed across ages for each local authority and sex using generalised additive models (GAM). This helped to reduce error from fluctuations between years and aimed to capture the true underlying relationship, which should not change significantly.

To produce the most accurate population estimates possible, we now incorporate Census 2021 data rolled forward to mid-year (Census 2021-based MYE), into the coverage adjustment strategy. We do this for both the admin-based population estimates (ABPE) best estimates and ABPE future estimates. This allows us to assess the coverage of these stock data and provide the best estimates of the population more accurately.

The ABPE best estimates and ABPE future estimates include more accurate coverage ratios by comparing the SPD version 4.0 with Census 2021-based MYE. Similarly to the coverage ratios in our previous publication, we smooth across age for each local authority and sex. We then assume that the coverage ratio for each age, sex, and local authority estimate changes linearly between 2011 and 2021.

To ensure consistency between the different versions of the SPD data through time, we made some adjustments. Since we cannot compare SPD version 4.0 to the 2011 Census-based MYE, we corrected the SPD version 3.0 coverage ratios by comparing them with the SPD version 4.0 in 2020. The correction relies on 2020 data, because this is the only year where SPD version 3.0 and SPD version 4.0 overlap. This helped to make the changes in the data smoother over time.

The SPD version 3.0 available for 2011 is based on an earlier and less accurate method than later 2016 to 2019 data. This may cause some error in the earlier coverage ratios but will be improved by the comparison with Census 2021-based MYE in later years.

To calculate coverage ratios for the PR between 2012 and 2015, we used the same smoothing and linear interpolation methods as used for the SPD.

This linear change is unlikely to be realistic, given significant changes to administrative data sources during the coronavirus (COVID-19) pandemic period and following the UK leaving the European Union.

This [SPD Estimation Options paper \(PDF, 1.3MB\)](#) outlines our plans to develop and implement more accurate methods for assessing the coverage of administrative data on an ongoing basis.

## Rates

We have made improvements to the method for producing migration and death rates.

In our previous publication, the denominator for flow rates was SPD version 3.0, adjusted for coverage error using the 2011 Census. For 2012 to 2015, we do not have SPD version 3.0 data. For these years, we imputed SPD version 3.0 by calculating the ratio of PR to SPD version 3.0 from 2016 to 2020. We fitted a linear model through these ratios over time for each local authority by age and sex combination. We then extrapolated the model results to cover the period from 2012 to 2015 and applied the ratios to the PR in these years to get a proxy for SPD version 3.0.

For the ABPE best estimates and ABPE future estimates, we have applied the modified coverage ratios with SPD version 4.0 to SPD version 3.0 adjustment to the calculation of the rate denominators. These were used for 2012 to 2021.

To produce a 2022 stock in the absence of actual stock data, we took the coverage adjusted SPD version 4.0 for 2021 and rolled it forward to 2022, using flow counts.

We have also improved how we smooth our rates. Smoothing is applied to reduce the amount of random variation and attempts to represent the true underlying rates. Previously, the smoothing of flow rates across age was applied separately for each local authority, sex (if applicable), and time using an adaptive generalised additive model (GAM). We noted in our previous publication that this method led to over-smoothing that was particularly noticeable for sharp migration peaks around student ages.

The new method uses an extension of the previous adaptive GAMs, known as GAM least absolute shrinkage and selection operator (GAMLASSO), to more effectively deal with the sharp peaks in our migration rates. This has significantly improved our ability to capture student peak migration. Better capturing these sharp peaks also helps to improve estimates at other ages. The approach relies on the use of dummy variables at ages between 0 to 30 years, which are known to contain sharp peaks, to minimise the likelihood of these being over-smoothed.

This method was also applied to the calculation of death rates at age 0 years to account for the higher mortality rate at this age, relative to other young ages.

We also apply smoothing over time rather than fitting to each year independently. The GAMLASSO model determines the impact of dummy variables based on consistency over time and minimises the likelihood of these being identical to the raw rate.

Whilst smoothing over time, on average, produces better results, we have identified two situations where changes in the age profiles over time are not captured.

The first is when there are changes in behaviour or circumstances. For example, a new campus for Swansea University was built in Neath Port Talbot in 2017, including student accommodation. Since 2017, the raw in-migration counts for Neath Port Talbot show a large peak at age 19 years which was not present in earlier years of the decade. However, our smoothing method includes the influence of this, throughout. Consequently, our input rates significantly over-estimate immigration at this age in early years and under-estimate it in later years. Future iterations of the DPM framework would look to incorporate more of this qualitative information to inform our estimates.

The second is when there are changes in methodology. For example, the international migration data in 2021 and 2022 changes to a new methodology and the overall counts of child migration are much lower than in previous years. For more information on international migration data, see [Section 5: Improvements to data sources](#). The smoothing method does not capture this change.

The improvement in smoothing can be seen in Figures 1 and 2. Figure 1 compares raw rates with smoothed rates, using adaptive GAMs and DPM outputs from our previous publication. Figure 2 compares raw rates with smoothed rates, using the updated GAMLASSO method and DPM outputs from the current publication.

In both figures, DPM output migration flows are cohort-based, which differs from the previous publication where we presented age-based outputs of migration. Cohort-based outputs provides greater coherence across the components of change.

### **Figure 1: The adaptive GAM method over-smooths university-aged migration into Cambridge**

**Cambridge female immigration for 2021 for our DPM estimates for case study local authorities**

Download the data

[.xlsx](#)

## Figure 2: An improved GAM-Lasso smoothing method allows us to capture sharp peak migration around student ages into Cambridge

Cambridge female immigration for 2021 for ABPE best estimates

Download the data

[.xlsx](#)

## Uncertainty

In our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#), the statistical models for SPD version 3.0 and PR assumed that the uncertainty was equal to the standard deviations obtained from measures of statistical uncertainty for SPD version 3.0. Read more in our [Admin-based population estimates and statistical uncertainty article](#).

We have now included measures of statistical uncertainty obtained using a similar method for the PR and SPD version 4.0, benchmarking against the 2011 Census and Census 2021, respectively.

We have used measures of statistical uncertainty for the 2011 Census-based MYE and Census 2021-based MYE in our ABPE best estimate and ABPE future estimate.

One substantive change to the method was applied for the PR because of bias increasing over time in the PR. Comparison against the 2011 Census revealed that PR bias was not constant between 2011 and 2015. This was an issue for certain clusters of local authorities for specific ages. In these cases, we allowed greater variability between local authorities and greater uncertainty for each local authority.

Estimates of uncertainty (confidence intervals) are based only on observed data and represent the range of values that the true value of an estimate is likely to fall within. Credible intervals describe the range of values that the estimate is likely to fall within, given the data and contextual information that has been fed into the model.

## 7 . DPM estimates with improved methodology for local authority case study areas

In our [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#), we selected 14 local authorities as case studies and produced dynamic population model (DPM) estimates for mid-year 2011 to 2022. These 14 local authorities provide a useful comparator in demonstrating the impact of changes to the DPM methodology. This section focuses on comparing our new DPM estimates with those from our previous publication.

Table 1 summarises the percentage difference between Census 2021-based MYE and our previous estimates, admin-based population estimates (ABPE) best estimates, and ABPE future estimates.

As we would expect, the ABPE best estimates are very close to the Census 2021-based MYE and improve on our previous methods. The largest change was seen in Cambridge. Our previous estimates were 5.8% lower than Census 2021-based MYE. Our new ABPE best estimates are 0.6% lower. ABPE best estimates and Census 2021-based MYE differ here because of [higher uncertainty in the Census 2021-based MYE](#), fed into the DPM. We see the same in Westminster, with an estimate 0.3% lower than Census 2021-based MYE.

Our ABPE future estimates are closer to Census 2021-based MYE than those published in November, in most of our 14 local authority case studies. The largest differences are in Welsh local authorities. This is because SPD version 4.0, used as a 2021 population stock, includes some data sources for England only. Nationally, the SPD under-counts in Wales by 5.17%, compared with 0.89% in England. Read more in our [Developing Statistical Population Datasets, England and Wales: 2021 article](#). This is corrected in the ABPE best estimates by using the Census 2021-based MYE as population stock in 2021, which does not have this undercoverage. Overall, the smaller range of percentage differences shows how important a robust coverage adjustment strategy will be for future estimates.

Table 1: Comparison of dynamic population model estimates for year ending mid-2021 and Census 2021-based MYE, for 14 local authority case studies, England and Wales  
Percentage difference from Census 2021-based MYE

Local authority	Census 2021-based MYE	DPM estimates for case study local authorities (published November 2022)	ABPE best estimates	ABPE future estimates
<b>Blackpool</b>	140,954	0.00%	0.00%	0.30%
<b>Boston</b>	70,798	2.10%	-0.10%	-0.20%
<b>Cambridge</b>	144,714	-5.80%	-0.60%	-1.30%
<b>Ceredigion</b>	70,658	1.40%	-0.10%	-1.90%
<b>Coventry</b>	343,320	3.30%	-0.10%	-1.20%
<b>Guildford</b>	143,929	-1.90%	0.00%	0.00%
<b>Gwynedd</b>	117,072	2.70%	0.00%	-2.20%
<b>Islington</b>	216,767	3.50%	0.00%	-0.30%
<b>Manchester</b>	549,853	0.30%	0.00%	-1.00%
<b>Newham</b>	350,626	1.10%	0.00%	0.40%
<b>North Norfolk</b>	103,257	-1.60%	0.00%	1.50%
<b>Swansea</b>	237,834	1.10%	0.00%	-2.10%
<b>Warwick</b>	148,729	-2.40%	-0.10%	-1.20%
<b>Westminster</b>	205,087	2.10%	-0.30%	1.40%

Source: Office for National Statistics

Figure 3 shows that people aged 18 to 20 years saw the largest changes in many of the 14 local authorities. These are the ages that are most impacted by the new smoothing methods, which now better account for sharp migration peaks in student ages. There are corresponding decreases in the 21-to-25-year age group caused by the shift in peak migration towards ages 19 and 20 years.

The reported high net international migration for 2021 and 2022, reflecting migration events discussed above, together with new methods for estimating international migration, are both likely to contribute to the sudden increases in people aged 18 to 20 years.

### **Figure 3: Greater percentage increases in the 18-to-20-year age group for 14 case study local authorities**

**Percentage difference between our DPM estimates for case study local authorities and ABPE best estimates by age group year ending 2021**

**Download the data**

[.xlsx](#)

## Cambridge

The [Dynamic population model for local authority case studies in England and Wales: 2011 to 2022 article](#) highlighted that the DPM estimates in Cambridge were being influenced by the SPD version 3.0 coverage ratios and low estimates of net inward migration. These factors led to our DPM estimates being lower than Census 2021.

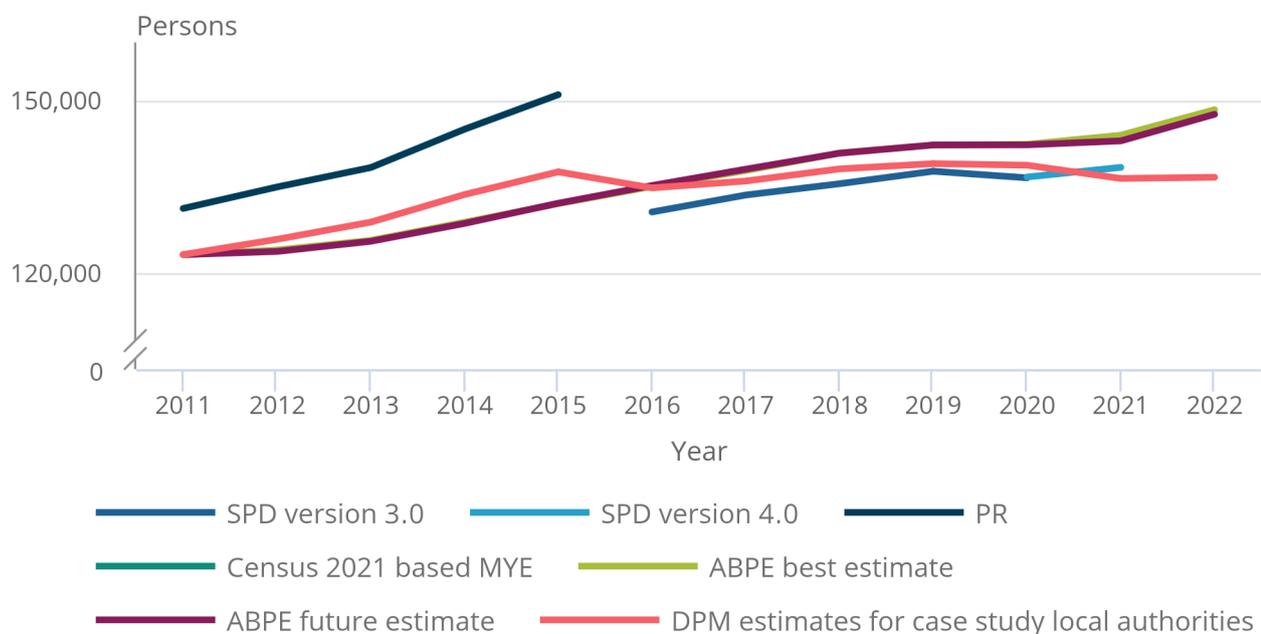
Figure 4 shows our DPM estimate from November, our two ABPE versions, and the stock estimates being used to produce our estimates.

### Figure 4: ABPE estimates and stock data for 2011 to 2022 showing impact of not using interpolated coverage ratios

Total population estimates for Cambridge, 2011 to 2022

#### Figure 4: ABPE estimates and stock data for 2011 to 2022 showing impact of not using interpolated coverage ratios

Total population estimates for Cambridge, 2011 to 2022



Source: Office for National Statistics

The estimates from our previous publication were affected by changes to stock data over time. These occurred as we moved from PR to SPD version 3.0 in 2016 and SPD version 3.0 to no stock data in 2021.

For the ABPE best estimates and ABPE future estimates, we use the information from the 2011 Census-based MYE and Census 2021-based MYE to produce coverage ratios that change linearly over the decade. This helps to smooth changes in the stock data over time and leads to estimates that should be more accurate. These estimates are also less affected by changes to input stock data as admin data sources change over time.

The ABPE future estimates are similar to the ABPE best estimates. This demonstrates how the DPM might perform in the absence of census data in the future, if we have a robust coverage survey and adjustment strategy. Our work towards this coverage adjustment strategy will be iterative and may take time to match the quality of the proxy coverage adjustment used here.

Figure 5 further demonstrates the improvements from using Census 2021-based MYE as a proxy for a robust coverage adjustment. Our previously published estimate typically underestimated the female population, compared with Census 2021. When ABPE future estimates coverage adjustment is used, the overall age profiles are similar between the ABPE best estimates and Census 2021-based MYE.

The peak age, at age 20 years, reflects a similar pattern seen in the SPD version 4.0 age profile, which the ABPE future estimates use for stock data in 2021. In our previous publication, the peak age was at 21 years of age because of over-smoothing migration rates.

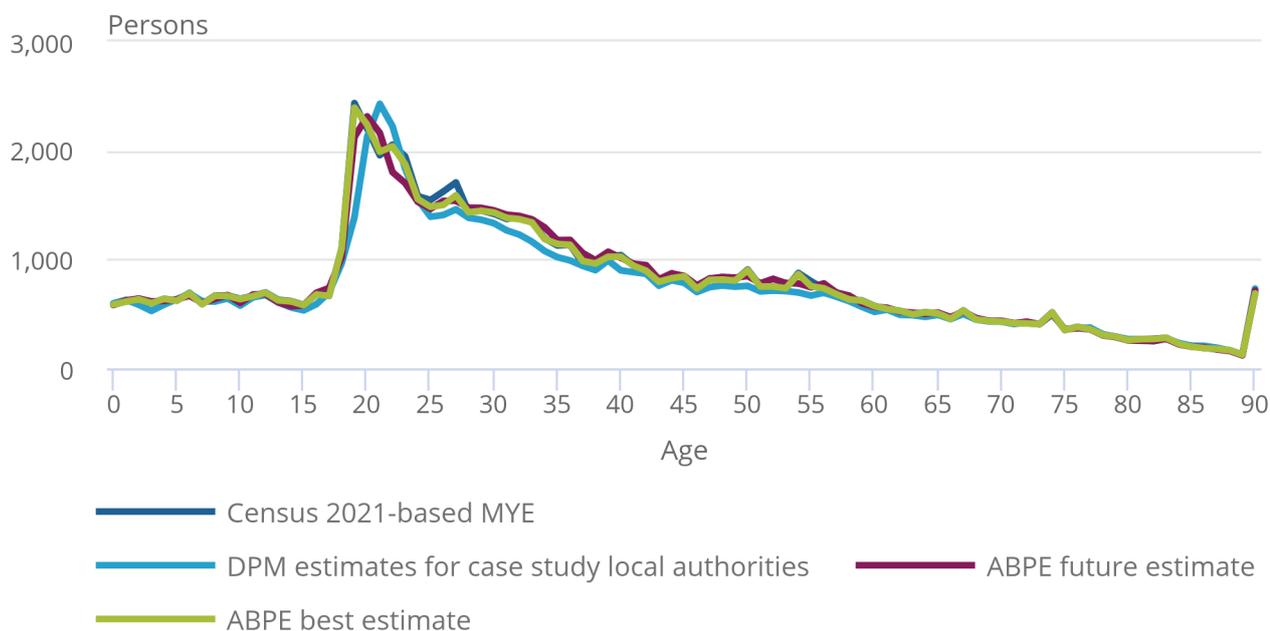
Figure 5 shows that our improved smoothing methods have moved the population peak to age 20 years in Cambridge for our ABPE future estimates. This is because we are better able to capture the sharp peak migration of students. Similar to Census 2021-based MYE, our ABPE best estimates places the population peak at age 19 years.

**Figure 5: Estimates from ABPE future estimates are very similar to ABPE best estimates for the population of females in Cambridge in 2021**

Cambridge female population estimates for 2021

Figure 5: Estimates from ABPE future estimates are very similar to ABPE best estimates for the population of females in Cambridge in 2021

Cambridge female population estimates for 2021



Source: Office for National Statistics

## 8 . Glossary

### Dynamic population model

The 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.

### Credible intervals

The range in which the true value of the estimates is likely to fall. We use 95% credible intervals by taking 2.5th and 97.5th percentiles from the distributions of counts produced by our model as the lower and upper bounds, respectively. For more information on this, see [Section 3: Summary of the DPM estimation process](#). Thus, the probability that the true value lies in the credible interval is 95%, given the data and the contextual information fed into the model.

## 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 Patient Register (PR), since it is updated more frequently and by a wider range of NHS services. The PDS data available to the Office for National Statistics (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.

## Generalised additive model (GAM)

A generalised additive model allows the modelling and smoothing of non-linear data. GAMs have been used within the DPM to model and smooth raw stock and flow data. This was done to reduce the amount of random variation and attempt to represent the true underlying pattern. This approach is particularly useful when working with noisy data (data that fluctuate a lot, year-on-year) or rare events.

## GAM Least Absolute Shrinkage and Selection Operator (GAMLASSO)

GAMLASSO is a statistical modelling technique that combines the flexibility of generalised additive models with the regularisation properties of lasso regression. It can be used to identify the most important predictors of a response variable and to model non-linear relationships between the predictors and the response. GAMLASSO has been used to model and smooth raw flow data, while maintaining sharp peaks in migration, as seen in student local authorities.

## 9 . Future developments

The admin-based population estimates (ABPEs) published alongside our companion results article are provisional estimates, which will be refined later in the year.

In summer 2023, we will publish an update to our ABPEs for all local authorities from 2011 to 2022 and compare against the new official mid-year estimates (MYE) for 2022.

We plan to incorporate the following improvements:

- replace Statistical Population Dataset (SPD) version 3.0 with version 4.0
- updated internal and international migration data
- first implementation of a coverage adjustment method using survey data

In further work, we will continue to research how we adjust our population stock data to account for coverage error. [This work is under development](#). This will also include investigating how we are accounting for special populations, including armed forces and students. We will continue to develop our methods for generating estimates of uncertainty. We will investigate our ability to produce credible intervals for aggregate estimates, giving estimates of uncertainty at local authority level, or by grouped age or sex combinations.

We will continue to investigate possible adjustments to improve the coherence across data sources by using age at time of event, rather than age at mid-year.

We will continue to develop our methods for splitting combined migration into internal and international migration components.

## Provide feedback

We welcome your feedback on the dynamic population model (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](mailto:pop.info@ons.gov.uk).

We have launched our [Local population statistics insight feedback framework](#), which enables users of population statistics to provide feedback at local authority level and suggest data sources for us to better understand the quality of our estimates.

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.

## 10 . Collaboration

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

We are also indebted to the insights, expertise, and feedback provided by local authorities:

- Blackpool
- Boston
- Cambridge
- Ceredigion
- Coventry
- Guildford
- Gwynedd
- Islington
- Manchester
- Newham
- North Norfolk
- Swansea
- Warwick
- Westminster

## 11 . Related links

[Admin-based population estimates: provisional estimates for local authorities in England and Wales, 2011 to 2022](#)

Article | Released 28 February 2023

Admin-based population estimates for all local authorities in England and Wales from the dynamic population model

[Developing Statistical Population Datasets, England and Wales: 2021](#)

Article | Released 28 February 2023

Aggregate comparisons between the Statistical Population Dataset v4.0 and Census 2021

[Reconciliation of mid-year population estimates with Census 2021, England and Wales](#)

Article | Released 28 February 2023

Analysis of differences between mid-year population estimates rolled forward from mid-2020 and official estimates rolled forward from Census 2021

[Transforming population statistics, comparing 2021 population estimates in England and Wales](#)

Article | Released 28 February 2023

Evaluating progress towards a transformed population statistics system, using comparisons between census-based and admin-based population estimates and Census 2021

[Understanding quality of Statistical Population Dataset in England and Wales using the 2021 Census - Demographic Index linkage](#)

Article | Released 28 February 2023

Analysis of Statistical Population Dataset version 4.0 2021 using a linkage between Census 2021 and the demographic index.

## 12 . Cite this methodology

Office for National Statistics (ONS), released 28 February 2023, ONS website, methodology, [Dynamic population model, improvements to data sources and methodology for local authorities, England and Wales: 2011 to 2022](#)