

# Administrative Data Based Population Estimates Scotland 2016

## Statistical Research

Published on 17 November 2020

Disclaimer: These statistical research outputs are **not the OFFICIAL STATISTICS** for Population Estimates for Scotland. The Official Statistics can be found in the statistics and data section of National Records of Scotland's website.

This publication reports on the results of research into how population estimates might be produced using a range of administrative data.

Any presentation or use of these research outputs should make clear to users the nature and purpose of the statistics.

## Contents

<b>Key Findings</b> .....	<b>3</b>
<b>1. Introduction</b> .....	<b>5</b>
<b>2. Acknowledgements</b> .....	<b>5</b>
<b>3. Background</b> .....	<b>6</b>
Aims of the project .....	6
Governance .....	6
Why Statistical Research rather than Official Statistics .....	7
Overview of the project: flow of data .....	7
Methodology .....	9
Quality Assurance of Administrative Data (QAAD).....	11
<b>4. Analysis of Administrative Data Based Population Estimates, 2016 ...</b>	<b>12</b>
Scotland by single year of age and sex, 2016.....	12
Comparison of administrative based data population estimates within the UK .	14
Council areas, 2016 .....	15
Urban-Rural Classification by sex, 2016 .....	18
Scottish Index of Multiple Deprivation (SIMD) by sex, 2016.....	19
Limitations of analysis .....	20
<b>5. Future Developments</b> .....	<b>21</b>
<b>6. Background Note</b> .....	<b>22</b>
Administrative Datasets Timeline .....	22
Stakeholder Engagements.....	22
Revisions .....	23
Links to related statistics .....	23
Supporting Documentation.....	24
<b>7. Glossary</b> .....	<b>24</b>
<b>8. Notes on statistical publications</b> .....	<b>25</b>

## Key Findings

The aim of this statistical research publication was to see how population estimates using only a range of administrative data, compare with the National Statistics mid-year estimates, which are based on adjusting census data with migration and vital events data.

The Administrative Data Based Population Estimates 2016 (ABPE) compare well with the 2016 Mid-Year Population Estimates (MYE) for Scotland's overall total.

- The ABPE for Scotland's population is 0.7 per cent higher than the published Mid-2016 Population Estimates overall.
- The ABPE estimates the population of Scotland to be 5,440,486 compared with 5,404,700 for the published Mid-2016 Population Estimates.

Looking at age and sex breakdowns, when compared with the Mid-2016 Population Estimates the ABPE was:

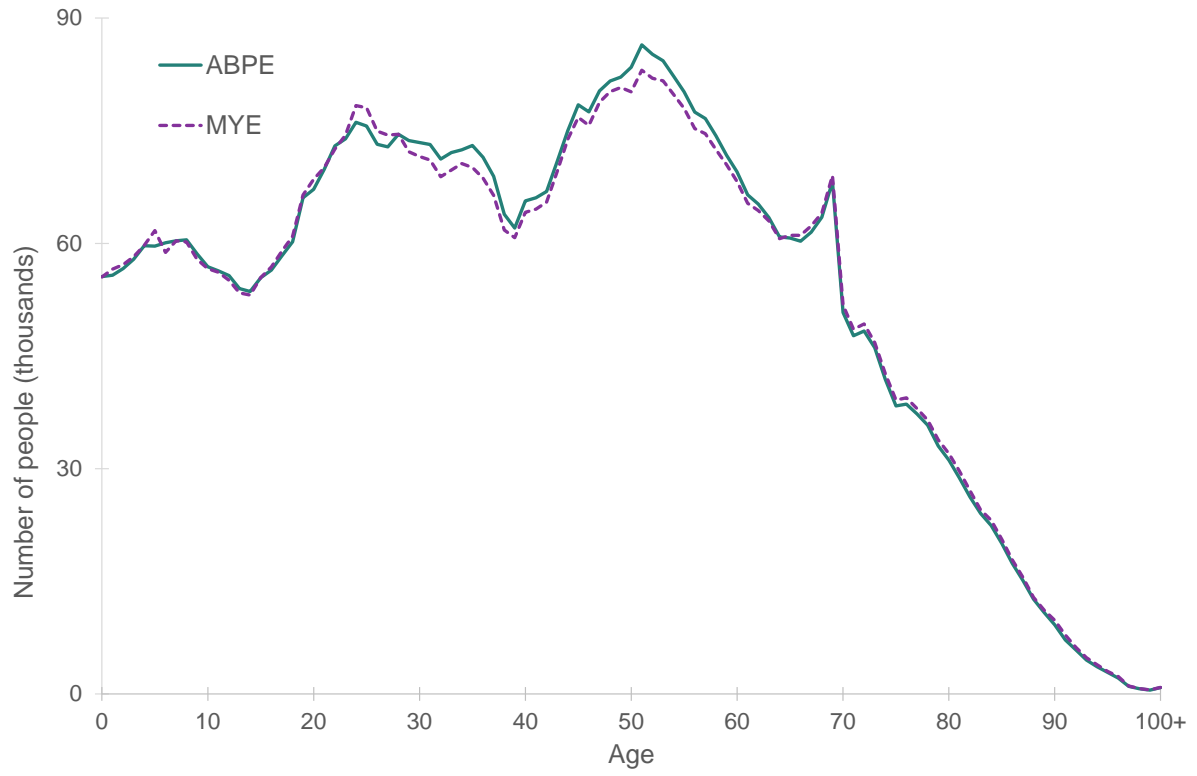
- 1.5 per cent higher for males and 0.1 per cent lower for females
- generally lower for people aged 1–5, 15–27 and 65+
- generally higher for people aged 6–14 and 28–64.

The ABPE also produced population estimates which:

- were higher for most-deprived areas and lower for least-deprived areas (the largest difference was 8.9 per cent for males in SIMD decile 1)
- were lower for small towns and higher for accessible and remote rural areas
- ranged from 3.8 per cent higher to 4.8 per cent lower at council area level, with half of the council areas being within 1.2 per cent.

The conclusion is that the results of the statistical research are encouraging. Future work will now focus on improving the quality of estimates across all age groups and at sub-national geographic aggregations.

**Figure 1: Comparison of ABPE and Mid-Year Estimates for Scotland, 2016**



## 1. Introduction

This statistical research publication presents population estimates produced from the counts of linked administrative datasets to estimate the population of Scotland in 2016. These estimates **should not** be considered as a replacement for the National Statistics Publication: [Mid-Year Population Estimates Scotland, 2016](#). If you require population estimates for any purpose, such as resource allocation, planning of services such as education and health, please use the latest mid-year population estimates available on the [NRS website](#). The figures in this publication should not be used for these purposes.

The purpose of this publication is to report on how the results of the research compare with the existing population estimates. This research is therefore an important step forward in our understanding of how current administrative data might be used to provide key demographic statistics.

## 2. Acknowledgements

The process of creating this first publication on population estimates from administrative sources has involved a number of organisations and individuals. The National Records of Scotland (NRS) Admin Data team would like to thank our data suppliers:

- [Electoral Registration Officers](#) (EROs)
- [Higher Education Statistics Authority](#) (HESA)
- [National Records of Scotland](#) (NRS)
- [Public Health Scotland](#) (PHS)
- [Registers of Scotland](#) (RoS)
- [Scottish Funding Council](#) (SFC)
- [The Scottish Government](#) (SG)

The NRS Admin Data team would also like to thank colleagues at the [Scottish Government](#) and [eDRIS](#) (part of PHS) for their ongoing support with this project. We would also like to thank all the stakeholders and peer groups, who have contributed their expertise and knowledge to support this work.

### 3. Background

#### Aims of the project

The aims of the project are:

- To help inform future recommendations for the census beyond 2022. This includes investigating administrative data collected by public bodies and services which could be used to augment or replace NRS' data collected by a traditional census. This statistical research is part of the administrative data based population and household estimates project, which will provide evidence for the future recommendations.
- To improve the coherence of our population and migration statistics across the UK — working in partnership with the Office for National Statistics (ONS) and the Northern Ireland Statistics and Research Agency (NISRA). This includes work as part of a cross Government Statistical Service programme to transform international migration statistics (one of the key components of population change). We are also collaborating on our respective programmes in NRS, ONS and NISRA to improve how we produce population statistics through greater use of administrative data.
- To support [discussion](#) with data suppliers and stakeholders on the application of this work and receive feedback on these initial proof of concept population estimates to inform future developments.

#### Governance

The aims of the project and the number of datasets being linked meant that the following governance steps were put in place before any research took place.

Scottish Government Analytical Leadership Group approved this project in late 2016.

The NRS Admin Data team investigated a range of datasets to see if they could support the project. Following discussions with colleagues in a number of organisations, the majority of data-sharing agreements were agreed by the end of 2017.

Throughout this period we worked closely with the Public Benefit and Privacy Panel for Health and Social Care (PBPP) on the outline of the project. PBPP is an independent panel which scrutinises research projects/proposals wishing to access NHS data.

We held our first stakeholder consultation on the project in late 2017, and provided further updates at events in 2018 and 2019. We plan to do more events on the findings of these results in 2021.

A [Data Protection Impact Assessment \(DPIA\)](#) was completed for the project and was published on the NRS website in 2019. This document outlines the process designed to systematically analyse, identify and minimise the data protection risks of this project and sets out our accountability under the General Data Protection Regulation (GDPR). This document is a live document and is reviewed and updated on a regular basis.

This project has taken four years to put together due to the complexities and governance around linking so many datasets. The presentation of the conclusions from the research will provide valuable information to support ongoing engagement with users as we further develop datasets and methodologies.

### Why Statistical Research rather than Official Statistics

For producers of official statistics, such as NRS, the term ‘Statistical Research’ is used to refer to research which is at a very early stage of its development and wouldn’t meet the requirements for official or experimental statistics. By using this term, NRS is able to formally publish material which can support further discussion and development.

While this publication is statistical research, NRS have provided [a voluntary adoption statement](#) to show how the principles of the [Code of Practice for Statistics](#) have been followed for this publication.

### Overview of the project: flow of data

A significant element of the complexity of this project was due to the need to bring together data sources from a number of different suppliers, all relating to slightly different time periods, though centred on 2016.

The datasets used in this publication are:

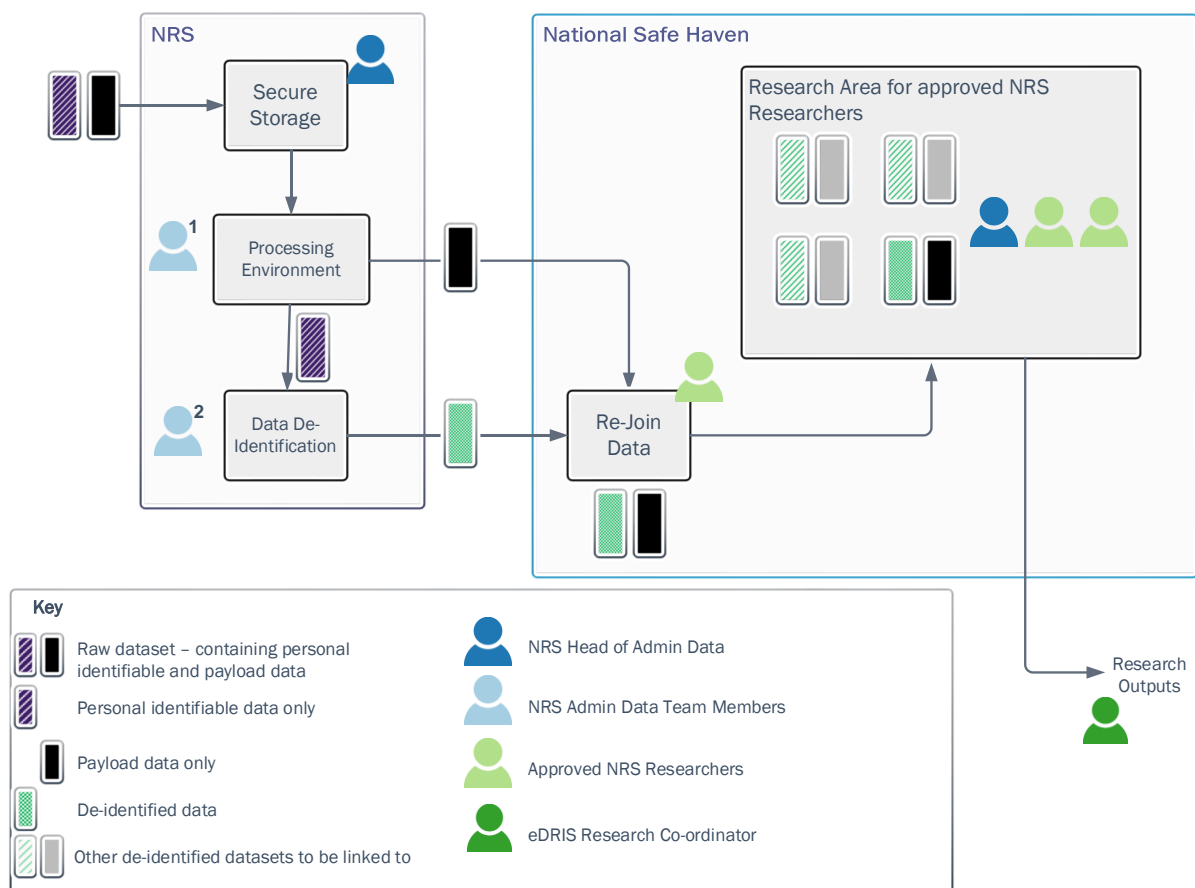
- Further Education Statistics (SFC)
- Health Activity (PHS)
- Higher Education Statistics Agency (HESA)
- National Health Service Central Register – NHSCR (NRS)
- Register of Electors (EROs)
- Residential Sales (RoS)
- Scottish Pupil Census (SG)
- Vital Events: Birth Registrations (NRS)
- Vital Events: Death Registrations (NRS)
- Vital Events: Marriage Registrations (NRS)
- Vital Events: Civil Partnership Registrations (NRS)

Datasets are sent securely to NRS from data suppliers. They are transferred one at a time to a secure area on the NRS servers to be processed. Once in the secure area the data are quality assured and then de-identified, that is, altered in such a

way that the original values cannot be recovered from the result thus protecting the identity of the data subjects.

Each dataset is separated into two parts: personal and payload. The personal information contains names, date of birth, postcode and/or address. This goes through the de-identification process. All other information is placed in the payload dataset. The payload may contain variables such as sex, ethnicity, disability, religion and date of last interaction with a service. While there are two health datasets in this project, they do not include individuals' medical information.

**Figure 2: Overview of the flow of data**



A statistician accesses the administrative dataset in the NRS secure IT environment and separates it into personal and payload data. The payload data is securely sent to the [National Safe Haven](#)<sup>1</sup>. The personal data is transferred to another secure area and the de-identification process is done by a different statistician. The de-identified personal data is then transferred to the National Safe Haven. The datasets are rejoined and analysed by another statistician, who was not involved in their preparation. This method of separating the processes and limiting the staff who see

<sup>1</sup> The National Safe Haven is used for the analysis as part of a trusted third-party model.



the dataset at different stages increases the security in processing individual's data. Figure 2 shows this process.

Once the data sources have been processed they are linked as described in the methodology section.

## Methodology

As part of the rationale for publishing as statistical research, the NRS Admin Data team expects further refinements in methodology to occur in future releases. This may be due to the addition of more datasets, improvements in data linkage techniques, changes in who is included or through the introduction of estimation techniques.

The [Administrative Data Based Population Estimates Scotland 2016: Methodology Report](#) provides more detail on the methodology. We have summarised the main parts of the methodology for this publication. The main difference between the ABPE and the MYE is that the ABPE does not depend on there being an earlier population estimate, whereas the MYE takes the latest Census Population estimates and makes adjustments using the latest data on births, deaths and migration.

The data linkage approach used in this methodology uses seven variables to link records: forename, surname, sex, day of date of birth, month of date of birth, year of date of birth and postcode. As an example, consider the name Aisha Khan.

The de-identification process applied to the personal data is important to make the data more secure. This turns Aisha Khan into a series of scrambled letters and numbers. It preserves the uniqueness of each name without revealing its contents. For example, Aisha Khan is now represented by the following:

Forename = c175447679ae2047, Surname = 2fb538526102fec9

The process is consistent, allowing Aisha Khan to link from one dataset to another. Instead of seeing Aisha Khan, the statistician sees c175447679ae2047 2fb538526102fec9. As this is a one-way process, it is impossible for anyone to derive the original name, thus making it impossible to identify individuals in this project.

The de-identified datasets are sent to the National Safe Haven, and then linked together using the de-identified linking variables. Some links will have exact agreement on the de-identified variables (forename, surname, postcode and date of birth). For others some information may be different, so we need to rely on a different combination of variables, for example name and postcode. In other cases the information may be recorded differently, and so derived linking variables, such as nicknames or name parts, are used.

A score is then assigned to each link, based on how well the two records agree. The score is also amended if the combination of variable values are not rare. For

example it is more likely that there are two different John Smiths with the same date of birth, than two Sarah-Jane Watt-Maxwells. The rareness of variable values is identified using the frequency of each de-identified value in the dataset.

Using the links found, and the scores for each of them, the records from the different datasets are grouped together in such a way that each group is expected to represent a unique person. However, some of the datasets contain records for people who are no longer there, in particular, deceased or emigrated. Therefore, the groups are trimmed down using the following business rules that exclude them and only include others if they:

- Appear on the NHSCR without a flag to say that they have died or moved elsewhere, and appear on one of the other datasets
- Appear on birth registrations and are aged below one.

Full details on these business rules and how they were derived is available in the [methodology report](#).

NHSCR is prioritised as it includes the greatest number of people in the population, and also indicates where people are no longer living in the Scotland (either because they have died or have left Scotland).

For each individual an age, sex and (de-identified) postcode is assigned. Where there is conflict between different datasets on these values NHSCR information is prioritised. The variable used to indicate sex is derived mainly from NHSCR and birth registration. If it is missing there then it is taken from other datasets. (This variable has been named gender in the code but it contains sex and gender.)

From the de-identified postcode a lookup table is used to assign the council area, [Scottish Index of Multiple Deprivation](#) (SIMD) decile<sup>2</sup> and [Urban–Rural Classification](#)<sup>3</sup>. We are then able to produce an age–sex breakdown by council area, SIMD decile or Urban-Rural classification. Once completed, this constitutes the population spine, Scotland's Integrated Demographic Dataset (SIDD).

The 2016 ABPE are the counts from the SIDD. NRS plan to explore alternative ways to produce the estimates from the SIDD. Possibilities include using dual-system estimation, a method used by Scotland's Census<sup>4</sup> to account for under-coverage. That method requires two distinct linked datasets. One option would be to produce two datasets from different administrative data sources, and link them. Another option would be to build a SIDD similarly to done in 2016 (perhaps with stricter business rules to ensure under-coverage), and link it to data from a population coverage survey.

---

<sup>2</sup> Based on 2016 SIMD deciles

<sup>3</sup> 2013/14 version

<sup>4</sup> See [www.scotlandscensus.gov.uk/estimation-and-adjustment](http://www.scotlandscensus.gov.uk/estimation-and-adjustment) for more information on how dual-system estimation is used in Scotland's Census.

In Section 4 the results are compared with the published Mid-Year Population Estimates (MYE) 2016. Figures from the '[Population Estimates for Scottish Centenarians](#)' have also been used to provide a further breakdown of those aged 90 and over. A full description of the [methodology](#) used for the MYE is published on the NRS website but a summary is provided here.

The MYE use the 2011 Census as a base and use a standard demographic method called the cohort component method. The cohort component method can be summarised as follows:

- Take the previous mid-year resident population and age-on by one year.
- Then estimate the population change between 1 July and 30 June by:
  - adding births occurring during the year;
  - removing deaths occurring during the year;
  - allowing for migration to and from the area.

Adjustments are also made for changes in some population groups that are not captured by the internal or international migration estimates: members of the armed forces and prisoners.

### **Quality Assurance of Administrative Data (QAAD)**

Administrative datasets are collected for a primary purpose for the organisation who own them. In this project, the data was being used in a different way from the primary purpose. The NRS Admin Data team worked closely with the data suppliers to gain a clear understanding of the strengths and limitations of each dataset in relation to this project. In addition, the completeness and quality of the data was reviewed when it was received from the data suppliers. More detailed information on each dataset has been provided in the [QAAD report](#).

This information helped to support the choice of business rules described in the methodology section.

## 4. Analysis of Administrative Data Based Population Estimates, 2016

After the production of the SIDD, administrative data based population estimates (ABPE) of the following have been produced:

- Scotland by single year of age and sex
- Council area by 5-year age bands and sex
- Urban-Rural classification by 5-year age bands and sex
- SIMD deciles by 5-year age bands and sex

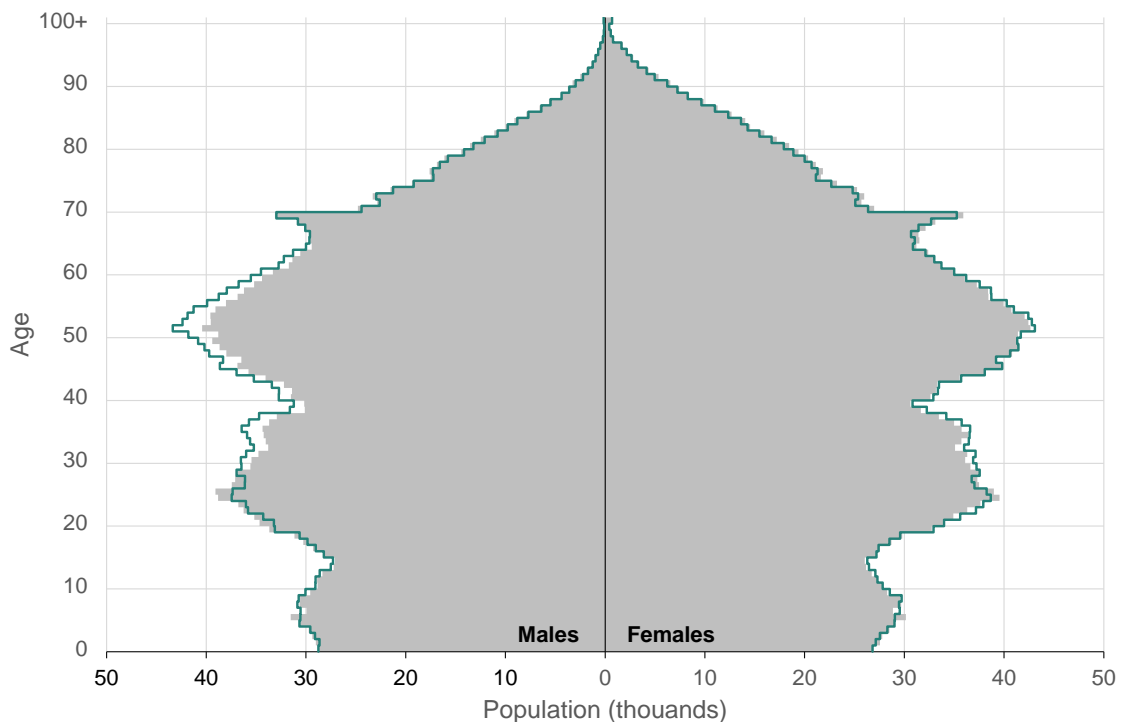
These outputs have been compared with two National Statistics Publications covering 2016:

- [Mid-Year Population Estimates \(MYE\), 2016](#)
- [Centenarians in Scotland, 2009 to 2019](#)

### Scotland by single year of age and sex, 2016

Figure 3 shows the age distribution of the ABPE compared with the MYE by sex. It can be seen that the population by age using the ABPE is approximately similar to that of the MYE.

**Figure 3: Comparison of ABPE (line) and MYE (bar) by age and sex**



**Figure 4: Percentage difference between ABPE and MYE by age and sex**

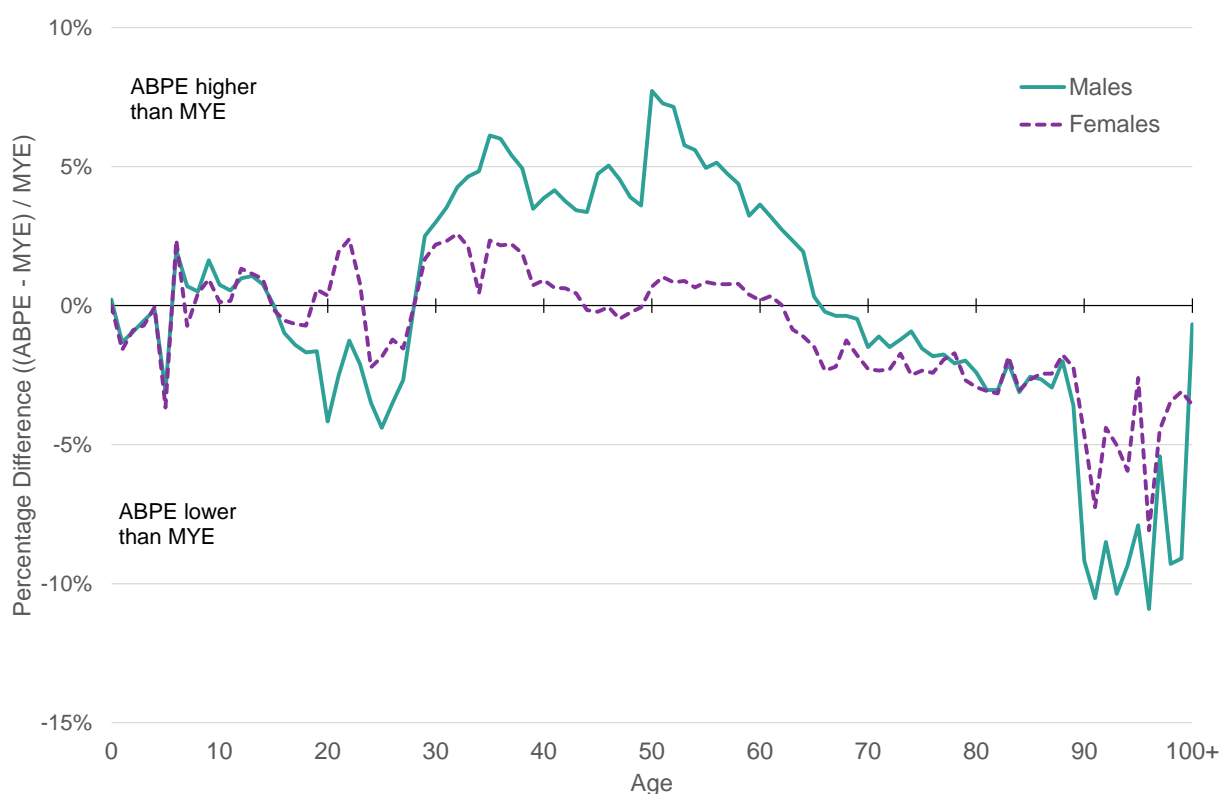


Figure 4 shows the percentage difference between the two sets of population estimates by age and sex. This shows more clearly the fluctuations by age and the patterns in the differences. For people aged around 30–60 the ABPE is higher than the MYE (especially for males). For older ages the ABPE is lower than the MYE. Note that, although the relative differences<sup>5</sup> are largest for ages 90+, the total difference for this group is 2,636, less than the single year difference for some ages, such as age 50 (3,277).

More specifically:

- The ABPE is generally lower than the MYE for people aged:
  - 1–5: (females by 1.4 per cent (2,007), males by 1.2 per cent (1,837))
  - 15–27: (females by 0.2 per cent (1,078), males by 2.4 per cent (10,725))
  - 65+: (females by 2.4 per cent (13,054), males by 1.5 per cent (6,685))

<sup>5</sup> Note, each MYE number used in the graph for ages 90 and over, is taken from the Centenarians in Scotland, 2009 to 2019 publication and so are rounded to the nearest 10, The Mid-2016 Population Estimates Scotland publication only has single year of age to age 89, with older ages being grouped together, as 90+. Making use of the Centenarians in Scotland, 2009 to 2019 publication allows the graph to be extended using single year of age up to age 99. When reporting the overall MYE total and other age ranges the Mid-2016 Population Estimates Scotland figures are used.

- The ABPE is generally higher than the MYE for people aged:
  - 6–14: (females by 0.7 per cent (1,849), males by 1.0 per cent (2,599))
  - 28–64: (females by 0.8 per cent (10,263), males by 4.4 per cent (56,395))
- The ABPE is 3.4 per cent (2,074) lower than the MYE for 5-year-olds, but 2.2 per cent (1,278) higher for 6-year-olds.
- For males there is a marked increase in the difference between the ABPE and MYE at age 50

The reasons for these differences require further investigation. Possible avenues for these investigations include:

- The business rules could be reviewed for children too young to appear on the school pupil census (primary 1 starts at age 4 or 5), who are likely to only be captured by the health activity and birth vital events datasets.
- There are suggestions that the variations by age are influenced by the frequency with which people interact with the NHS. For example, the increase at age 50 for males coincides with the age at which the national bowel screening programme commences. For subsequent publications, it is planned that the time since the last interaction (rather than just that there was an interaction in the previous three years) will be available. The threshold for inclusion could then be adjusted by age to account for variations in frequency of interaction by age.

### Comparison of administrative based data population estimates within the UK

Each devolved administration are responsible for published their own population estimates. NISRA produced population estimates based on data gathered from statistical censuses and surveys, and data extracted from its own and other organisations' administrative or management systems.

ONS have produced administrative data based population estimates for England and Wales as statistical research (that is, creating population estimates not based on a census). In June 2019, ONS published a comparison of the third version of their [Admin-Based Population Estimates, England and Wales: 2011 and 2016](#).

The second version of their methodology has a similar pattern to Scotland of under/over-coverage for males to their national mid-year estimates for 2016, in particular, over-coverage amongst males from the late 20s to early 60s and under-coverage for older age groups. ONS have under-coverage for late 20s to early 60s by their third version of their methodology.

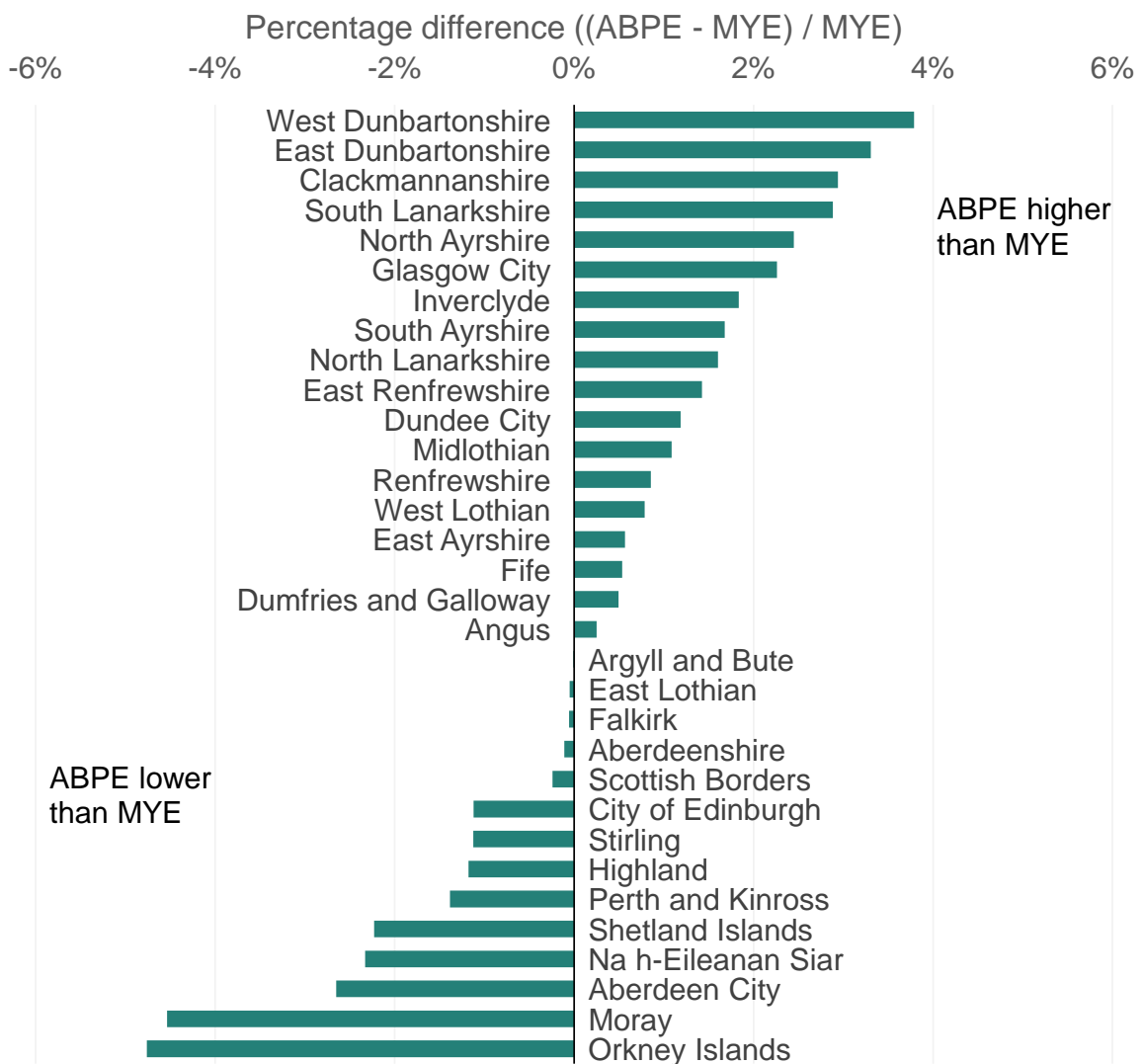
For females, ONS and NRS have different patterns for under/overage coverage when comparing to their mid-year estimates. ONS have changed their business rules between versions 2 and 3, and also incorporated more datasets.

## Council areas, 2016

There were 6,391 people (0.12 per cent) for whom a location could not be identified, so they have not been assigned to any council area (or other geographic breakdown) but are included in Scotland figure. This category are included in the [output tables](#).

The comparison between the ABPE and the MYE, broken down by council area is shown in figures 5 and 6. Relative differences between the ABPE and the MYE by council area are shown in Figure 5.

**Figure 5: Percentage difference between ABPE and MYE by council area.**



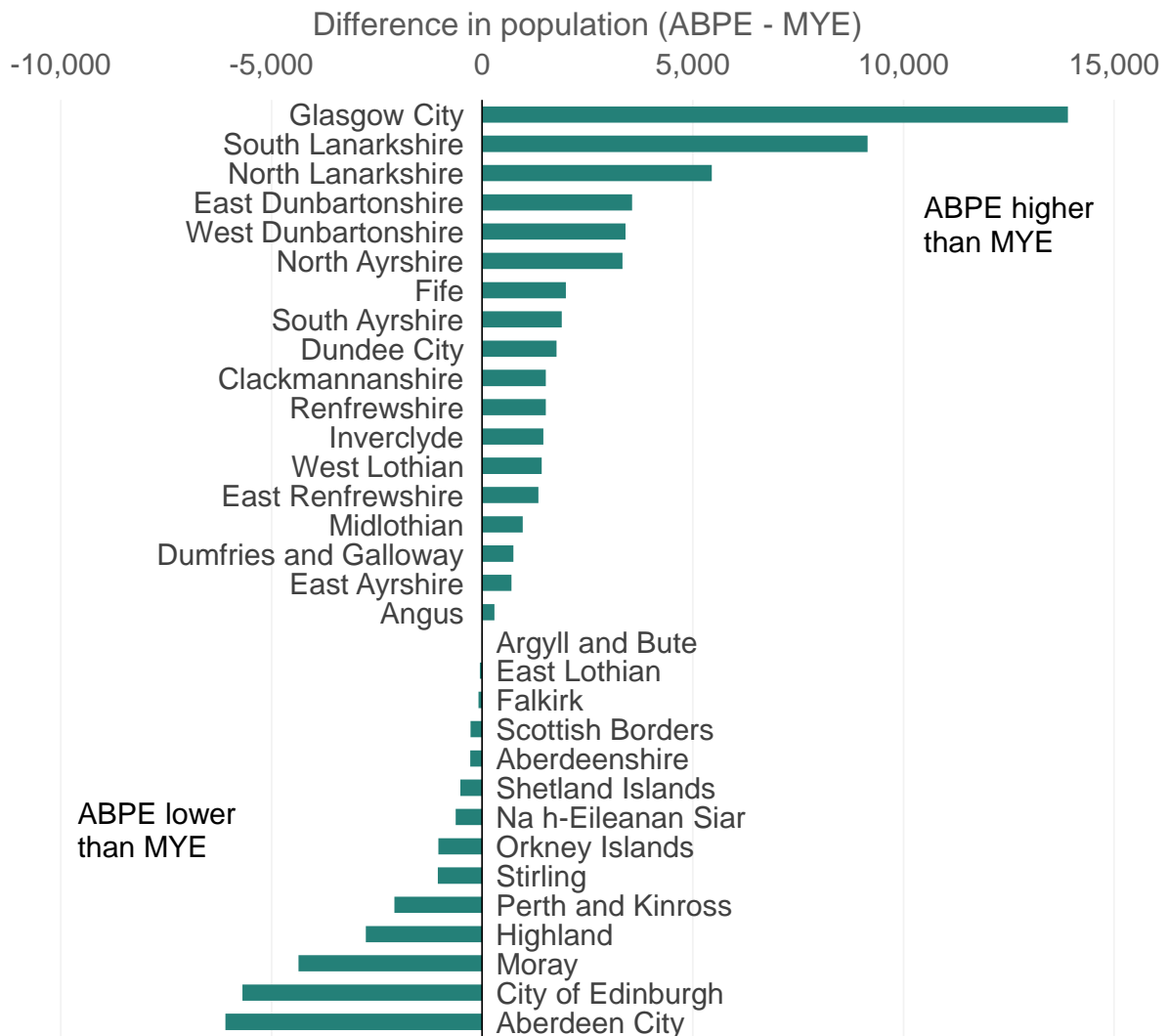
Findings on the differences between the ABPE and the MYE by council area include:

- The percentage difference between ABPE and MYE for council areas range from the ABPE being 4.8 per cent lower (for Orkney Islands) to 3.8 per cent higher (for West Dunbartonshire).

- For half of the council areas the ABPE is within  $\pm 1.2$  per cent of the MYE
- Regarding the four main city council areas, the ABPE is higher than the MYE for Glasgow City and Dundee City, and lower for City of Edinburgh and Aberdeen City.
- The ABPE is higher than the MYE for many council areas that are close to large urban areas.
- The ABPE is lower than the MYE for particularly remote council areas.

Figure 6 shows the differences in population by council area, with the largest council areas tending to show the largest differences.

**Figure 6: Difference between ABPE and MYE by council area**





Some of the patterns in differences between ABPE and the MYE at Scotland level are also found in most council areas. For example:

- ABPE has a lower number of females aged 65 and over than the MYE in every council area. For males, this was aged 70 and over.
- ABPE is higher for males aged 30–59 in the majority of council areas.

In council areas where ABPE is higher than the MYE, the differences generally mirror the differences found at Scotland level. However, in the council areas where ABPE is lower than the MYE there is much more variation in where the differences are found by age and sex.

A set of [interactive charts](#) for council areas have been published as part of the statistical outputs.

Note, that we have not presented results at geographies lower than council area. However, we plan to investigate aggregations of lower-level geographies to further understand the differences seen across council areas. The following two subsections provide an alternative view of geographic differences in the comparisons, by breaking down by urban–rural classification and SIMD decile.

## Urban-Rural Classification by sex, 2016

The comparison between the ABPE and the MYE, broken down by the [2013/14 8-fold Urban-Rural classification](#) is shown in Figure 7.

**Figure 7: Percentage difference between ABPE and MYE by 8-fold Urban–Rural classification and sex**



The findings on the percentage differences between the ABPE and the MYE include:

- In urban areas the ABPE is 0.7 per cent (12,721) lower than the MYE for females, but 1.2 per cent (22,556) higher for males.
- The ABPE is lower than the MYE for small towns (females by 2.3 per cent (8,152), males by 0.5 per cent (1,547)).
- The ABPE is lower than the MYE for very remote rural areas (females by 2.4 per cent (1,806), males by 1.2 per cent (898)).
- The ABPE is higher than the MYE for accessible and remote rural areas (females by 4.0 per cent (15,783), males by 4.2 per cent (16,180)).

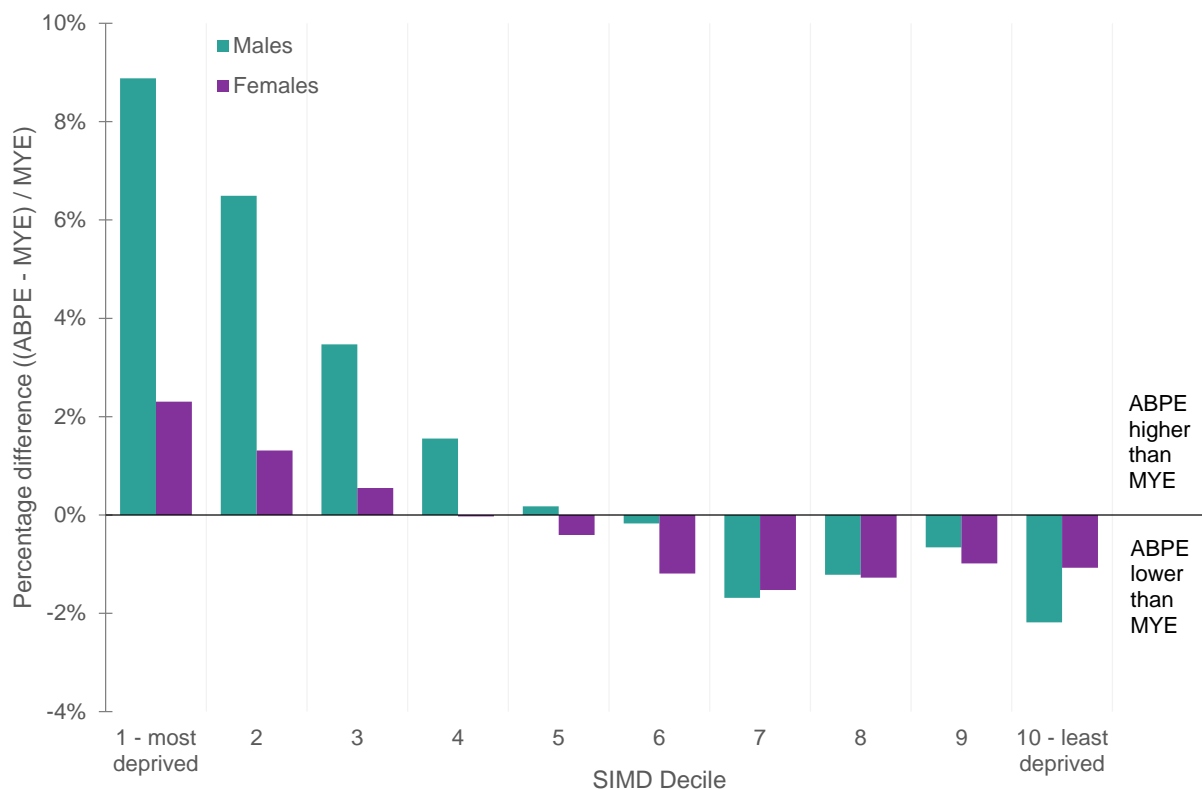
- The ABPE for females varies between 3.3 per cent lower than MYE to 4.4 per cent higher. The range for males is smaller, being 2.3 per cent lower to 4.3 per cent higher than MYE.

It is not clear what is causing these differences. The pattern of differences by urban–rural classification may differ by age group. The effect of urban-rural is larger for females than males. Exploring these might provide insights as to the source of the differences between the ABPE and MYE for different classifications.

### Scottish Index of Multiple Deprivation (SIMD) by sex, 2016

The comparison between the ABPE and the MYE, broken down by 2016 SIMD deciles is shown in Figure 8. The ABPE tends to be higher than the MYE for most deprived areas (especially for males), and lower for least deprived areas.

**Figure 8: Percentage difference between ABPE and MYE by SIMD decile and sex**



Findings on the difference between the ABPE and the MYE include:

- The ABPE is higher than the MYE for the most deprived decile (males by 8.9 per cent (22,550), females by 2.3 per cent (6,210)).
- The ABPE is lower than the MYE for the least deprived decile (males by 2.2 per cent (6,069), females by 1.1 per cent (3,100)).

- The ABPE for females varies between 1.5 per cent lower to 2.3 per cent higher than MYE. The range for males is larger, with the ABPE being 2.2 per cent lower to 8.9 per cent higher.

Unlike for the urban-rural comparisons, the effect of SIMD is larger for males than females. Again it is not clear what is causing differences by SIMD. Areas for possible exploration could include looking at differences in how often people interact with the NHS by SIMD band.

### Limitations of analysis

As noted, these statistics are statistical research and the methods will be reviewed for future publications. Although not all the people who appear on the datasets used are included in the SIDD, it is likely that there are still people living in Scotland who do not appear on any of the datasets used, as they do not interact with these organisations. Conversely, not everyone who appears on the SIDD will be living in Scotland on the reference date.

The methodology of this project will be refined in the future. It is only when the ABPE is compared with the corresponding census, that we may be able to investigate further some of these underlying differences.

## 5. Future Developments

This publication was produced to share our findings and methodology with users of Scotland's population statistics to help focus discussion on how administrative data could support population statistics in the future. The project will be running until 2022 and it is our hope that administrative based population estimates will be produced for each year from 2016 to 2022.

To address over-coverage the business rules could be made more stringent. This could involve making different use of the administrative datasets currently being used, such as trimming down the health activity dataset to only people who interacted more recently. If other administrative datasets could be used then the business rules could be amended to only include people who also appeared on those new datasets. The NRS Admin Data team will continue to explore further administrative datasets as they become available. Also we may make business rules more specific in future, such as having different rules for different age groups or different geographic areas.

To address under-coverage, changes could also be made to the business rules, including for specific groups or using extra data sources. Alternatively the estimates could be calculated from the SIDD, but not directly using the counts of people in the SIDD. Scotland's Census uses dual-system estimation to account for under-coverage in the census. To use such a technique for the ABPE would require two linked datasets. NRS will explore doing this using two datasets build from administrative data, and using the SIDD as one dataset and linking it to data from a population coverage survey. The population coverage survey dataset could be constructed from data from existing government surveys.

NRS also plan to develop administrative data based estimates on occupied dwellings. The research for this has begun and will continue into 2021.

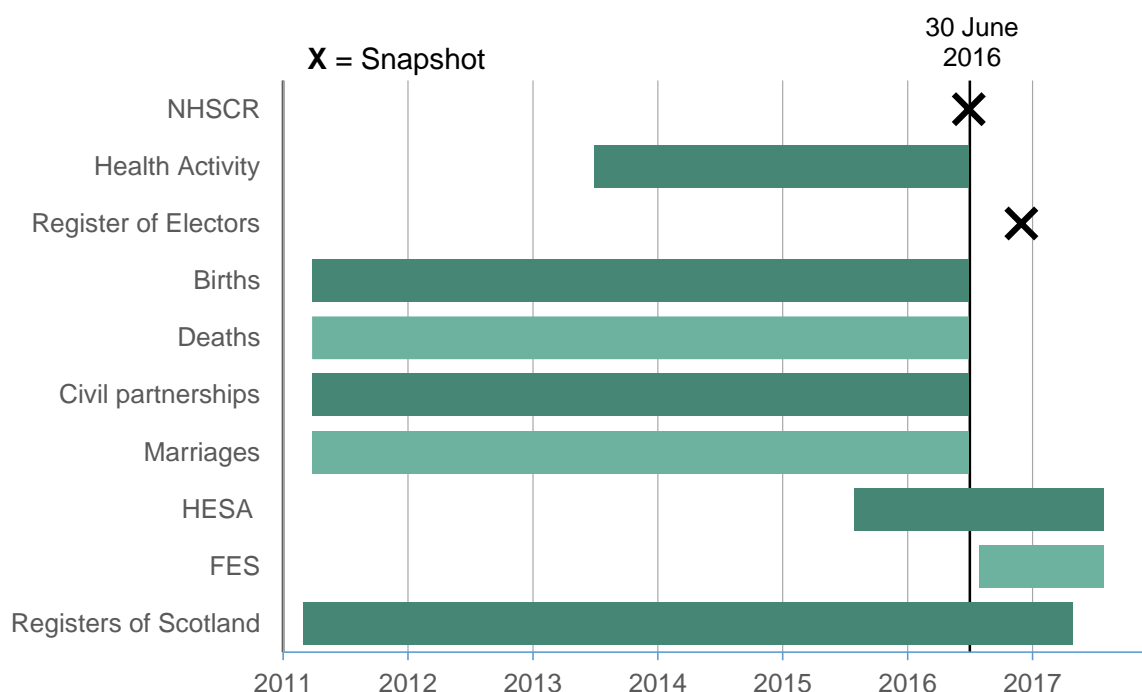
It is our hope that future stakeholder events will add to a greater understanding of how this work could be refined and its possible uses in the future.

## 6. Background Note

### Administrative Datasets Timeline

Several administrative data sources have been used in the production of these population estimates. While the estimates use 30 June as a reference data, the data sources in this publication cover different time periods and have a degree of lag between the time period covered and when they are available. The time periods covered by each dataset are shown in Figure 9.

**Figure 9: Time period covered by each dataset**



The timeliness of data is an issue. Although most of the datasets were received before the end of 2016, the datasets covering students (HESA and FES) were not available until spring 2018. The de-identification process takes six months, limiting publication to 1.5 years after the end of the reference year at the earliest. Considering the administrative data based population estimates for 2020 as an example, it would be approximately autumn/winter 2022 before all the datasets were available to analyse. NRS will work with our data suppliers to explore ways to improve this.

### Stakeholder Engagements

Stakeholder engagement has always been important to this project; the NRS Admin Data team in 2017 embarked on a wide ranging process of [stakeholder engagement](#).

The [Population And Migration Statistics Committee](#) (PAMS) is one of the main demographic users groups in Scotland. Any substantial progress was supplied to this group.

Following this publication, the NRS Admin Data team wish to discuss the findings of this research with as many users as possible. If you have any comments or would like to be involved in stakeholder events, then please register your interest under the topic title: demography at <http://www.gov.scot/scotstat>.

## Revisions

This statistical research is expected to be revised in subsequent publications due to changes in methodology. Those subsequent publications will supersede this publication.

Revisions and corrections to previously published statistics are dealt with in accordance with the Scottish Government Statistician Group [corporate policy statement](#) on revisions and corrections.

## Links to related statistics

Similar approaches to producing population estimates using administrative data are being conducted by other statistical agencies nationally and internationally. Though the approaches are similar, the major differences are the different types of administrative data that are available to each organisation. Therefore, apart from with ONS, no direct comparisons are made.

In the UK, the Office for National Statistics (ONS) have produced [administrative data based population estimates](#) for England and Wales, for 2011 and 2013 to 2016. This is part of their Census and Data Collection Transformation Programme. Northern Ireland Statistics and Research Agency (NISRA), which looks at the use of administrative data in population estimates, have published their initial findings for [Northern Ireland](#).

Internationally, New Zealand have developed their own research environment, the [Integrated Data Infrastructure](#) and have used it to produce various [administrative data based population estimates](#).

[Statistics Canada](#) have also been looking at creating a population spine, initially using data from 2011 administrative data sources.

Ireland's Central Statistics Office (CSO) have produced a Statistical Population Dataset for 2011, from which they have [produced population estimates](#).

The National Statistics publication used for benchmarking purposes are the [Mid-year Population Estimates 2016](#). This is the source that should be used when doing any research or analysis using population statistics and the results presented here **should not** be used as an alternative source

## Supporting Documentation

This publication has a number of supporting documents, these can be found on our website:

- [Administrative Data Based Population Estimates, Scotland 2016 – Methodology Report](#)
- [Administrative Data Based Population Estimates, Scotland 2016 – Quality Assurance of Administrative Dataset](#)
- [Data Protection Impact Assessment \(DPIA\) - Administrative Data Based Population and Household Estimates Project](#)
- [Voluntary Adopter of Code of Official Statistics Statement](#)
- [Administrative Data Based Population Estimates, Scotland 2016 – Tables](#)
- [Administrative Data Based Population Estimates, Scotland 2016 – Charts](#)
- [Administrative Data Based Population Estimates, Scotland 2016 – Interactive Charts](#)

## 7. Glossary

Table 1 provides a description of the abbreviations used in this document.

**Table 1: Description of abbreviations used**

Abbreviation	Description
ABPE	Administrative Data Based Population Estimate
CSO	Central Statistics Office
DPIA	Data-Protection Impact Assessment
ERO	Electoral Register Officer
FES	Further Education Statistics
GDPR	General Data Protection Regulation
HESA	Higher Education Statistics Agency
MYE	Mid-Year Estimate
NHSCR	National Health Service Central Register
NISRA	Northern Ireland Statistics and Research Agency
NRS	National Records of Scotland
ONS	Office for National Statistics
PAMS	Population and Migration Statistics Committee
PBPP	Public Benefit and Privacy Panel for Health and Social Care
PHS	Public Health Scotland
QAAD	Quality Assessment of Administrative Data
RoS	Registers of Scotland
SFC	Scottish Funding Council
SG	Scottish Government
SIDD	Scotland's Integrated Demographic Dataset
SIMD	Scottish Index of Multiple Deprivation



## 8. Notes on statistical publications

### Statistical Research

This publication presents statistical research and the methodology is still under development. We welcome any feedback from users on ways in which the methodology or data sources may be developed to improve the quality of these statistics in future years.

### Information on background and source data

Further supporting documentation is published alongside this publication on the NRS website. The hyperlinks link can be found at the end of [Section 6](#).

### National Records of Scotland

We, the National Records of Scotland, are a non-ministerial department of the devolved Scottish Administration. Our aim is to provide relevant and reliable information, analysis and advice that meets the needs of government, business and the people of Scotland. We do this as follows:

Preserving the past – We look after Scotland’s national archives so that they are available for current and future generations, and we make available important information for family history.

Recording the present – At our network of local offices, we register births, marriages, civil partnerships, deaths, divorces and adoptions in Scotland.

Informing the future – We are responsible for the Census of Population in Scotland which we use, with other sources of information, to produce statistics on the population and households.

You can get other detailed statistics that we have produced from the [Statistics](#) section of our website. Scottish Census statistics are available on the [Scotland’s Census](#) website.

We also provide information about [future publications](#) on our website. If you would like us to tell you about future statistical publications, you can register your interest on the Scottish Government [ScotStat website](#).

You can also follow us on twitter @NatRecordsScot

## **Enquiries and suggestions**

Please get in touch if you need any further information, or have any suggestions for improvement.

Lead Statistician: Lindsay Bennison

Statistics Customer Services telephone: (0131) 314 4299

E-mail: [statisticscustomerservices@nrscotland.gov.uk](mailto:statisticscustomerservices@nrscotland.gov.uk)

For media enquiries, please contact: [scotlandscensus@nrscotland.gov.uk](mailto:scotlandscensus@nrscotland.gov.uk)