

Administrative Data Based Population Estimates using Version 3 Methodology, Scotland (2016-2018)

Statistical Research

Published on 29 August 2023

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

1. Key Findings	3
2. Introduction	5
3. Acknowledgements	6
4. Background	8
Aims of the project	8
Why Statistical Research rather than Official Statistics.....	9
Mid-Year Population Estimate Methodology	9
5. Administrative Data Based Population Estimates Methodology	10
Version 1 and 2 Methodology Summary	10
Version 3 Methodology	11
Comparison of Version 2 and Version 3 Estimates	16
6. Analysis of Administrative Data Based Population Estimates	22
Scotland Overall.....	22
Scotland by age and sex.....	23
Council areas	25
Limitations of analysis.....	27
7. Future Developments	28
8. Background Note	29
Governance	29
Administrative Datasets Timeline.....	29
Quality Assurance of Administrative Data (QAAD).....	30
Stakeholder Engagements.....	30
Revisions	31
Comparison of administrative data based population estimates within the UK .	31
Comparison of administrative data based population estimates outwith the UK	32
NRS Official Population Estimates	32
Supporting Documentation.....	32
9. Glossary	34
10. Notes on statistical publications	35
Statistical Research	35
Information on background and source data.....	35
National Records of Scotland.....	35
Enquiries and suggestions	36

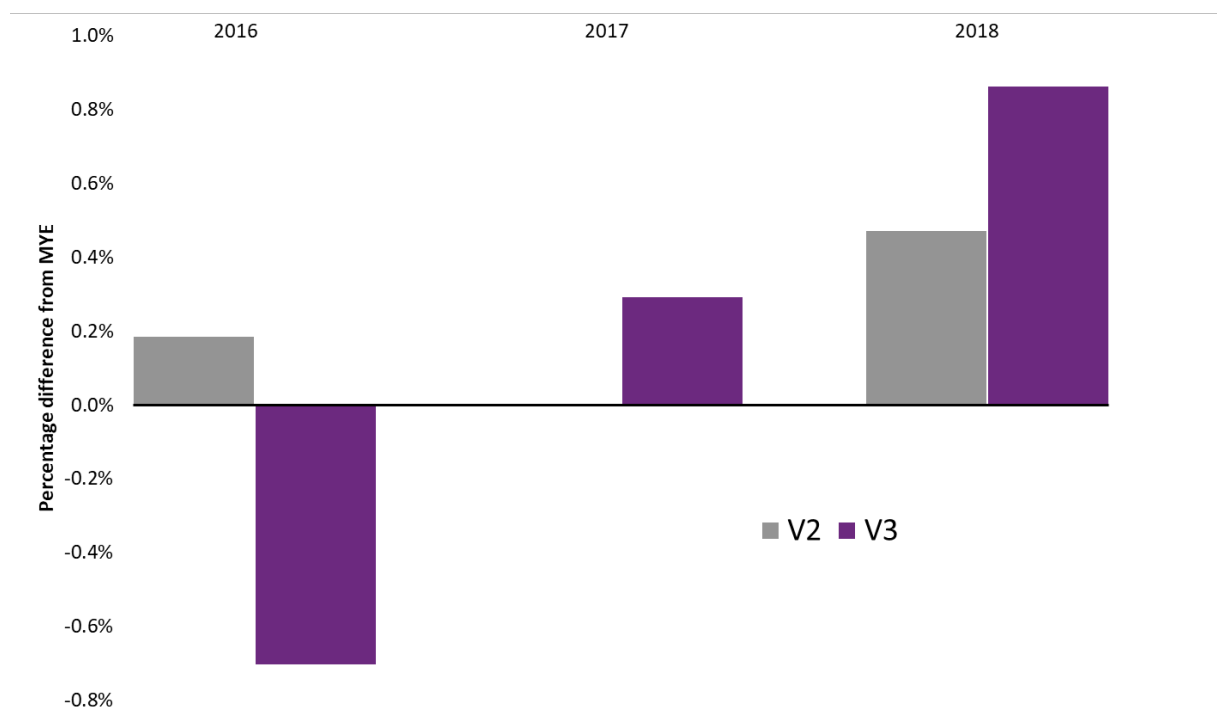
1. Key Findings

The aim of this statistical research is to create population estimates using only a range of administrative data. This research is part of the evidence base to support future developments in Scotland's demographic statistics.

This research is compared with the National Statistics Mid-Year Population Estimates (MYE), which start by using the latest census population estimates and then adjust for future years using data on births, deaths and migration. MYE are published by the National Records of Scotland each year.

The aim of this publication is to present the new version 3 methodology for the Administrative Data Based Population Estimates (ABPE).

Figure 1: Difference in the population estimates between ABPE (versions 2 and 3) and MYE by year.



- The changes in methodology have led to a decrease in the Administrative Data Based Population Estimates for 2016, and an increase for 2017 and 2018 (see Figure 1).

- Figure 1 also shows that the Administrative Data Based Population Estimates remain broadly similar to the Mid-Year Estimates across the three years.
- Changes to data and linking methods increased the estimates, while the estimates were decreased by changes to business rules around which persons to include in the estimates.
- School-aged children are less affected by the methodology changes than adults.

2. Introduction

This statistical research publication presents Administrative Data Based Population Estimates (ABPEs) of Scotland. These are produced from linked administrative datasets and are for to 2016, 2017 and 2018. This publication supersedes any outputs in the earlier statistical research, which were based on versions 1 and 2 of our methodology. The current estimates are based on version 3 of our methodology, where decisions on allotting a Unique Person Identifier (UPID) are modified, and business rules for exclusion/inclusion have been refined, in the light of research.

These estimates **should not** be considered as a replacement for the National Statistics Publication for population estimates. If you require population estimates for any purpose, such as resource allocation, planning of services such as education and health, or to incorporate into other statistics, 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 the results of the research comparing estimates from ABPE version 3 (ABPE_{v3}) with those from ABPE version 2 (ABPE_{v2}). It also compares estimates from ABPE_{v3} with the existing population estimates (MYE). This research is therefore an important step forward in our understanding of how current administrative data might be used to provide key demographic statistics.

3. Acknowledgements

The process of producing this research 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. The suppliers, along with the datasets they provided, are given in Table 1. All the administrative datasets were held securely, and only brought together for linking once they had been de-identified. More information on the governance and datasets can be found in section 8 and section 3 of [ABPE Scotland 2016: Methodology Report \(v1\)](#).

Table 1: Suppliers of the data used in the analysis, and the datasets supplied.

Supplier	Dataset	Description
Electoral Registration Officers (EROs)	Electoral Register (ER)	People registered to vote in Scotland
Higher Education Statistics Authority (HESA)	HESA	Higher education students studying or domiciled in Scotland
National Records of Scotland (NRS)	NHS Central Register (NHSCR)	People who are or have been registered with a GP in Scotland, or whose birth was registered in Scotland
	Vital Events	Births, deaths, marriages and civil partnership registrations
Public Health Scotland (PHS)	Health Activity (HA)	Patients who have interacted with the NHS in the previous three years
Scottish Funding Council (SFC)	Further Education Statistics (FES)	Further education students studying in Scotland
The Scottish Government (SG)	Scottish Pupil Census (SPC)	People enrolled in state funded schools in Scotland

The NRS Admin Data team would also like to thank colleagues at the [Scottish Government](#) and [eDRIS](#) (part of Public Health Scotland (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.

4. Background

Aims of the project

The aims of the project are:

- To help inform future recommendations for the census and demographic statistics beyond 2022. This includes investigating administrative data collected by public bodies and services, which could be used to augment, complement 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 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 population estimates to inform future developments.

Although it was not an aim of the project, this work was instrumental in developing the NRS response to the lower than expected census return rate. The research into the different administrative data sources available, linking methodologies, and experience with the governance systems that manage different administrative data sources meant that we could respond quickly to produce the administrative data spine that was used in our census estimation methodology, and secure the relevant approvals.

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 that is at an early stage of its development and would not meet the requirements for official or experimental statistics. By using this term, NRS is able to formally publish material that can support further discussion and development. This publication presents the third iteration of our methodology and is therefore still regarded as statistical research. As with our earlier publications on Administrative Data Based Population Estimates, 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.

Mid-Year Population Estimate Methodology

In Section 6, the results are compared with respective [Mid-Year Population Estimates](#) (MYE). Figures from the [Centenarian publication](#) 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 uses the 2011 Census as a base and uses 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, in particular, members of the armed forces and prisoners.

5. Administrative Data Based Population Estimates Methodology

Version 1 and 2 Methodology Summary

The methodology behind the production of ABPEs generated thus far is fully documented in previous publications and methodology reports available on National Records of Scotland website. Version 1, which was used in the first ABPE publication on 2016 data, is described in the accompanying methodology report at: [Administrative Data Based Population Estimates, Scotland 2016 - Statistical Research | National Records of Scotland \(nrscotland.gov.uk\)](#). Version 2 was used in the second ABPE publication on 2016–2018 data. Differences between versions 1 and 2 are described at: [Administrative Data Based Population Estimates, Scotland 2016 - 2018 - Statistical Research | National Records of Scotland \(nrscotland.gov.uk\)](#).

In summary, linking variables were derived from four pieces of information: name, sex, date of birth and postcode. All personal data and all variables derived from them are de-identified to make it impossible to recover the original information. This means we cannot identify individuals. The de-identified versions of the datasets listed in Table 1 are sent to the National Safe Haven, and then linked together using the de-identified linking variables.

Once the de-identified records have been linked together the links are analysed. Records that appear to represent the same person are pulled together into a group, with unlinked records being placed in distinct groups. Each group is given a UPID (Unique Person Identifier). Some UPIDs will represent persons who are not in the Scottish population on the reference date. Therefore, to make the list of UPIDs closer to the list of persons living in Scotland on the reference date, business rules are applied. The business rules trim down the list of UPIDs, for example removing persons who appear on the death registrations dataset. There are some exceptions, but in general UPIDs will be included if the person appears on the NHSCR with an active posting, and on at least one other dataset. Changes to the business rules will change the final list of UPIDs and hence the population estimates.

For each person an age, sex and (de-identified) postcode is assigned. A lookup from de-identified postcode to various levels of geography (from local authority down to data zone) is used to place each person in the appropriate area. NRS use administrative datasets where sex or gender is self-declared. The ABPE use the term sex for reporting but we are aware that this variable may contain sex and/or gender from the underlying dataset. For most UPIDs the sex information will be derived from the Health Activity dataset, which is itself an amalgamation of various health datasets.

Once completed, the trimmed UPID list constitutes Scotland's Integrated Demographic Dataset (SIDDD). This statistical dataset includes a record for each person estimated to be in the Scotland population on the reference date. For each record it includes the age, sex and location of that person down to data zone level. However, details such as name, date of birth and postcode are not known as they have been de-identified. The SIDDD is the dataset used to estimate the population. The population estimates (ABPE) are counts of records on the respective SIDDD for each year.

Version 3 Methodology

This publication now presents version 3 of the ABPE methodology, which has been applied to 2016, 2017 and 2018 data to produce new estimates. Changes to the methodology are driven by what makes sense for each individual person that appears on the data. In particular, individuals who appear and who have evidence to suggest that they are in the Scotland population on the reference date will be included on the SIDDD, while those without such evidence, or where there is evidence against their being in the population, will not be included. In this way the methodology attempts to optimize the list of persons on the SIDDD, making it as similar as possible to the list of persons genuinely in the population. The methodology changes are not directly intended to bring the overall population estimates closer to the MYEs.

The differences between versions 2 and 3 fall into three categories: data, linking, and business rules. These will be discussed in turn below.

Data

It was reported in the previous publication that there had been a problem with the Fife electoral register dataset that resulted in the estimate for Fife being lower than the corresponding official mid-year estimate. This was because any individual who appears on NHSCR and ER, but not any other dataset (or where their last NHS interaction did not meet the threshold for their age) would not be included in the SIDD.

In order to mitigate this, a Fife 2017 ER dataset was constructed from the Fife ER datasets for 2016 and 2018. These datasets were linked together using name and postcode. If an exact match was found then it was assumed that the person had remained in the same location over this two-year period, and hence was likely to be at the same location in 2017. The constructed 2017 Fife dataset therefore included all, and only, those people who appear in the 2016 and 2018 ER datasets at the same location. In this way, most of the people who would have appeared on the true ER in 2017 at Fife could still be included. This would not be as good as the true ER dataset, but it was considered a substantial improvement over no data at all.

Linking

For version 3 there are two changes to the overall linking process. The first is that the linking now takes account of Unique Property Reference Number (UPRN) when assigning a strength to the link. If two records appear at the same unique property, then that gives greater confidence that they represent the same individual, than if they only appear at the same postcode. Therefore, links between records that are at the same address have a smaller penalty than records that are only at the same postcode. That difference can be enough to make some links strong enough to be considered in the final analysis. This is especially the case for records where name or date of birth is missing (mainly for school pupil census and electoral register respectively). This revision can result in more people being included in the SIDD.

The second change to the linking process concerns the step that takes the records and links, and assigns each record a Unique Person Identifier (UPID). In previous versions the guiding principle was that if two records had not been linked then they

should be considered different persons, and be assigned different UPIDs. However, this could result in some cases being assigned incorrectly.

Table 2: UPID Allocation (Example 1)

Record	First	Last	DoB	Postcode	Source	UPID _{v2}	UPID _{v3}
1	Sarah-Jane	Smith	1/1/1970	AB1 1AA	NHSCR	1	1
2	Sarah-Jane	Smith	1/1/1970	EH10 1AA	HA	2	1
3	Sarah-Jane	Smith	Missing	EH10 1AA	ER	2	1

In the fictional example in Table 2 above, it is likely that record 1 and record 2 represent the same individual (as name and date of birth are identical), and these records would be successfully linked. It is also likely that record 2 and record 3 represent the same individual (as name and location are identical). This suggests that record 1 and record 3 are also the same person. However, it may be the case that record 1 and record 3 do not link directly. According to version 2 methodology record 1 and record 3 would be given distinct UPIDs. Record 2 would then be assigned the same UPID as whichever of record 1 or 3 it had linked most strongly to, which here would be record 3. Both of these UPIDs would then have failed the business rules (UPID 1 only appears on NHSCR and UPID 2 does not appear on NHSCR). However, if we believe that these records all represent the same individual who appears on three sources, then they should be counted.

It is not always the case that records linked indirectly should be considered the same individual.

Table 3: UPID Allocation (Example 2)

Record	First	Last	DoB	Postcode	Source
1	Sarah-Jane	Smith	1/1/2000	AB1 1AA	NHSCR
2	Sarah-Jane	Smith	5/5/1970	AB1 1AA	NHSCR
3	Sarah-Jane	Smith	1/1/2000	AB1 1AA	HA
4	Sarah-Jane	Smith	5/5/1970	AB1 1AA	HA
5	Sarah-Jane	Smith	Missing	AB1 1AA	ER

In the example in Table 3 above record 5 would link to all the other records. This would mean that all records were placed in the same cluster (where a cluster is a collection of records that all link to each other directly or indirectly via another record). Despite this we can see that there are two individuals here. We therefore do

not want to assign each record the same UPID, which was why the original rule assigned unlinked records distinct UPIDs.

In order to correctly handle both these cases the UPID allocation process was amended. There is now an initial check that counts the number of records from each source dataset in the cluster. If none of the sources have more than one record in the cluster then all the records are assigned the same UPID. This would be applied in Example 1 (Table 2), correctly assigning all three records to the same person, which would then pass the business rules.

If there are multiple records from any source then the UPID allocation reverts back to the version 2 method, where records that are not linked are given different UPIDs. This would mean that in Example 2 (Table 3), the records would be assigned to one of two UPIDs. Both of these would pass the business rules, and so this would be correctly counted as two persons.

Business rules

In version 3 there are also changes to the business rules.

Health Activity

The health activity datasets have a variable called 'transfer-out'. This can give an indication that it is believed that the person is no longer resident in Scotland. The business rules were therefore amended to exclude anyone who has this flag on either the primary or secondary care datasets.

HESA

The HESA dataset gives information on where students are domiciled, where they are during term time and where their permanent address is. The dataset includes all HESA students who are domiciled in Scotland, or who are studying in Scotland. In version 2 if it was shown that the person is studying outwith Scotland then the HESA record was not counted as evidence of their presence. However, this can be taken as evidence that they are not in Scotland. The business rule has therefore been amended to exclude anyone who was studying outwith Scotland in the academic year leading up to the June reference date.

For people who have an active posting on the NHSCR, HESA is now only taken as evidence if the person has a term-time address in Scotland in the academic year leading up to the reference date.

Analysis showed that there were many records on HESA that did not appear on the NHSCR. It was considered that what could be happening here is that students who are domiciled in other parts of the UK could already be registered with an NHS GP at their home address. Given this, they may elect to not register with a Scottish GP, and so would not get an active posting on the NHSCR. As these people will still be part of the Scotland population the version 3 methodology includes some HESA records even if the person does not have an active NHSCR posting. In particular it will include persons:

- domiciled in the rest of the UK (home address not in Scotland) who were studying in Scotland in the academic year leading to the reference date, on a course that has not finished, even if: 1) they were not on the NHSCR, or 2) they were showing as having moved elsewhere on the NHSCR
- domiciled in Scotland who were studying in Scotland in the academic year leading up to the reference date, on a course that has not finished, even if they were showing as having moved elsewhere on the NHSCR.

Persons studying in Scotland but domiciled outwith the UK will appear on HESA, but will also need to appear on the NHSCR in order to be included in the SIDD. This distinction between how overseas students are treated from students from the rest of the UK, is because it was felt less plausible that overseas students would already be registered with an NHS GP, and hence less plausible that they would not register with a NHS GP in Scotland.

Vital Events

In version 2, evidence was taken from vital events if the person had been involved in any of the registrations on the datasets. As these went back to 2011 this means that the later years included more VE data than earlier years. This meant that the amount of evidence available differed between the years. It also meant that some of this evidence was quite old by the time of the reference date.

Version 3 therefore changed the business rules to make the vital events evidence more consistent across the years, and also to be more stringent on inclusion. It did this by taking advantage of the fact that vital events registrations implicitly link separate individuals. For births the mother, father and child are all connected to the same registration. Marriage and civil partnership registrations connect the two partners with a single registration. The version 3 business rules therefore count evidence from vital events if at least one of the other people linked via the registration have already met the business rule.

Table 4: Business Rules (Example 3)

First	Last	Source
Sarah	Smith	NHSCR
Sarah	Smith	HA
Sarah	Smith	MGS
John	Smith	NHSCR
John	Smith	MGS

In the example in Table 4 above we have a record that Sarah and John were married in the five years prior to the reference date (MGS). Sarah satisfies the business rules in virtue of appearing on NHSCR and health activity (with a sufficiently recent interaction). Aside from the marriage registration, John only appears on NHSCR, which would otherwise not be sufficient to include him in the SIDD. However, John’s connection with Sarah means that (as Sarah already meets the business rules) he does satisfy the business rules.

Comparison of Version 2 and Version 3 Estimates

Figures 2 to 4 show the percentage differences between the version 2 ABPEs and version 3 for each of the three years by age. These charts show the effect on the estimates of the data and linking revisions (solid line), all the changes excluding the transfer-out change (dashed line), and the full version 3 (dotted line), all relative to $ABPE_{V2}$.

Figure 2: Percentage differences between ABPE_{V2} and ABPE_{V3} (2016)

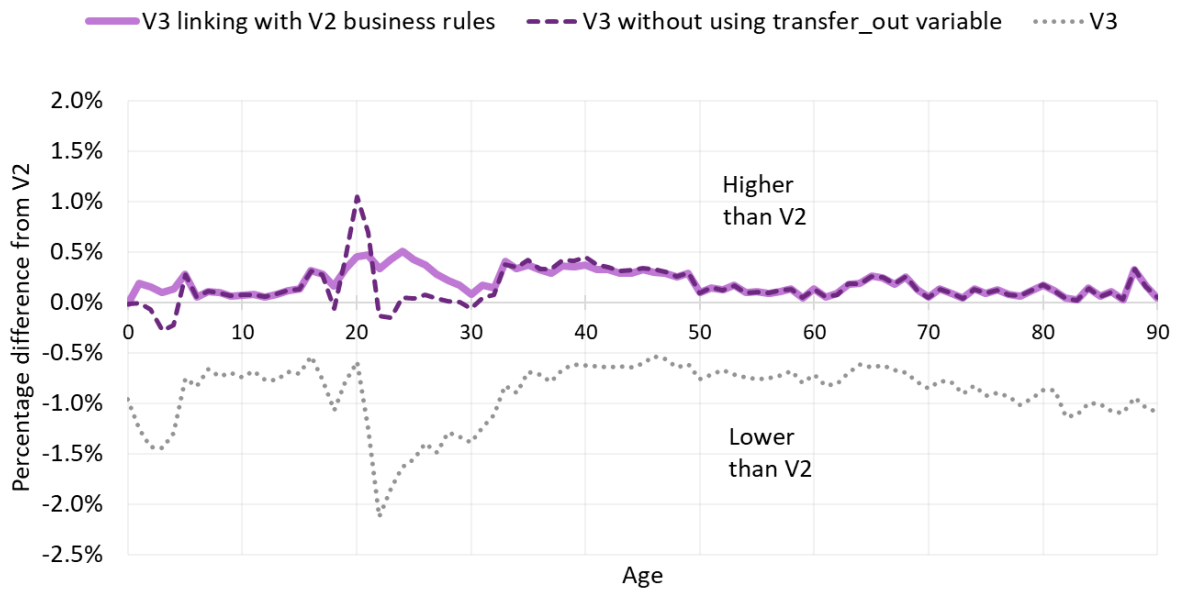


Figure 3: Percentage differences between ABPE_{V2} and ABPE_{V3} (2017)

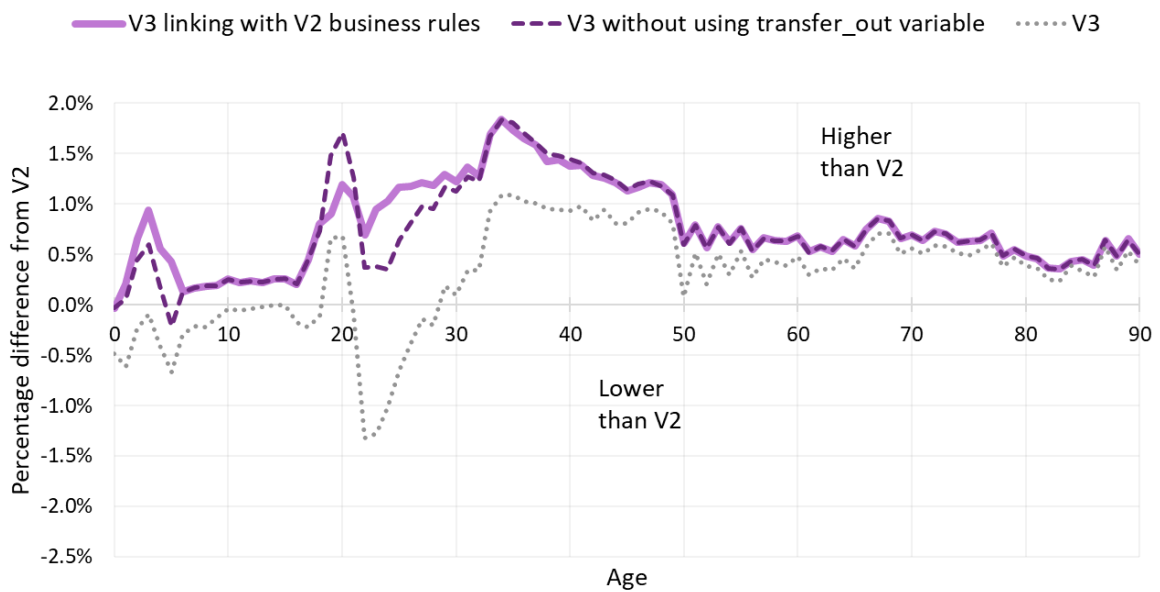
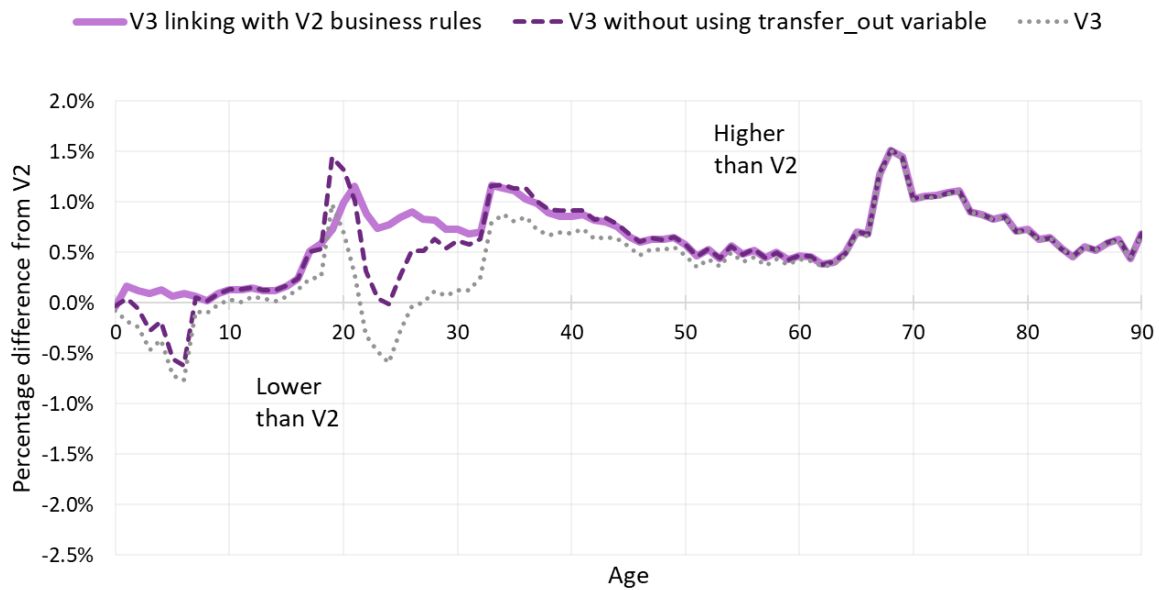


Figure 4: Percentage differences between ABPE_{V2} and ABPE_{V3} (2018)



The data linking lines (solid) are always positive, showing that the data and linking changes add more individuals. This makes sense as the data changes add in more records (for Fife in 2017). Strengthening links where the UPRN agrees can also mean that more people meet the business rules if it ensures that, for example, their NHSCR and ER records are now given the same UPID. The changes to UPID allocation can also mean that more people will meet the business rules because, for example, it avoids the situation in Example 1 where their NHSCR record gets one UPID and their health activity and ER records get a different UPID.

In 2017, these changes have a large effect, especially for age groups where ER has a substantial impact (mainly middle ages). It can also be seen that these changes have a larger effect on the 2018 estimates than the 2016 estimates. Table 5 shows how often a person appears at a different location on their ER record than on their NHSCR record. It can be seen that in 2018 there are more cases where people appear at different locations on the NHSCR and ER, especially for cases where health activity is also at a different location from the NHSCR. As these can be the cases that would be affected by the UPID allocation changes, this contributes to the larger impact on 2018 than 2016. The different patterns in location observed in 2016

and 2018 could be due to the EU referendum in 2016, which meant that more people would update their ER records.

Table 5: Difference in Health Activity and NHSCR address for UPIDs where NHSCR and Electoral Register are at different locations¹

HA-NHSCR Address	2016	2017	2018
Different	36,900	100,200	56,900
Same	108,100	104,800	144,800
Total	144,900	205,000	201,700

The dashed lines in figures 2 to 4 are close to the data linking (solid) lines, showing that most of the business rule changes have a small impact on the estimates. For late teens and 20s the effect is larger, showing that the changes for HE students have an effect. On most of this range the net effect of these business rule changes is to reduce the estimates, suggesting that the change to only taking evidence from people studying in Scotland in the academic year leading to the reference date is having the dominant effect. However, for some ages around age 20, these business rules have a positive net effect on the estimates, suggesting that the inclusion of students domiciled in the rest of the UK even if they are not on NHSCR is having the dominant effect there.

The other notable impact of these business rules is for children. Only considering five years' worth of vital events data means that birth registrations are only considered for children under five. This has the greatest impact on the 2018 estimates, where this reduces the estimates for five and six year olds. Younger ages could also be affected if their parents do not meet the business rules.

The final version 3 lines (dotted) show the estimates that include the transfer-out flag business rule, and so differ from the dashed lines by taking account of the transfer-out change. This has one of the larger impacts, especially for 2016. Comparing the number of UPIDs with transfer-out flags in the different years showed that 2016 has more persons with the transfer-out flag than other years. In addition, the later years

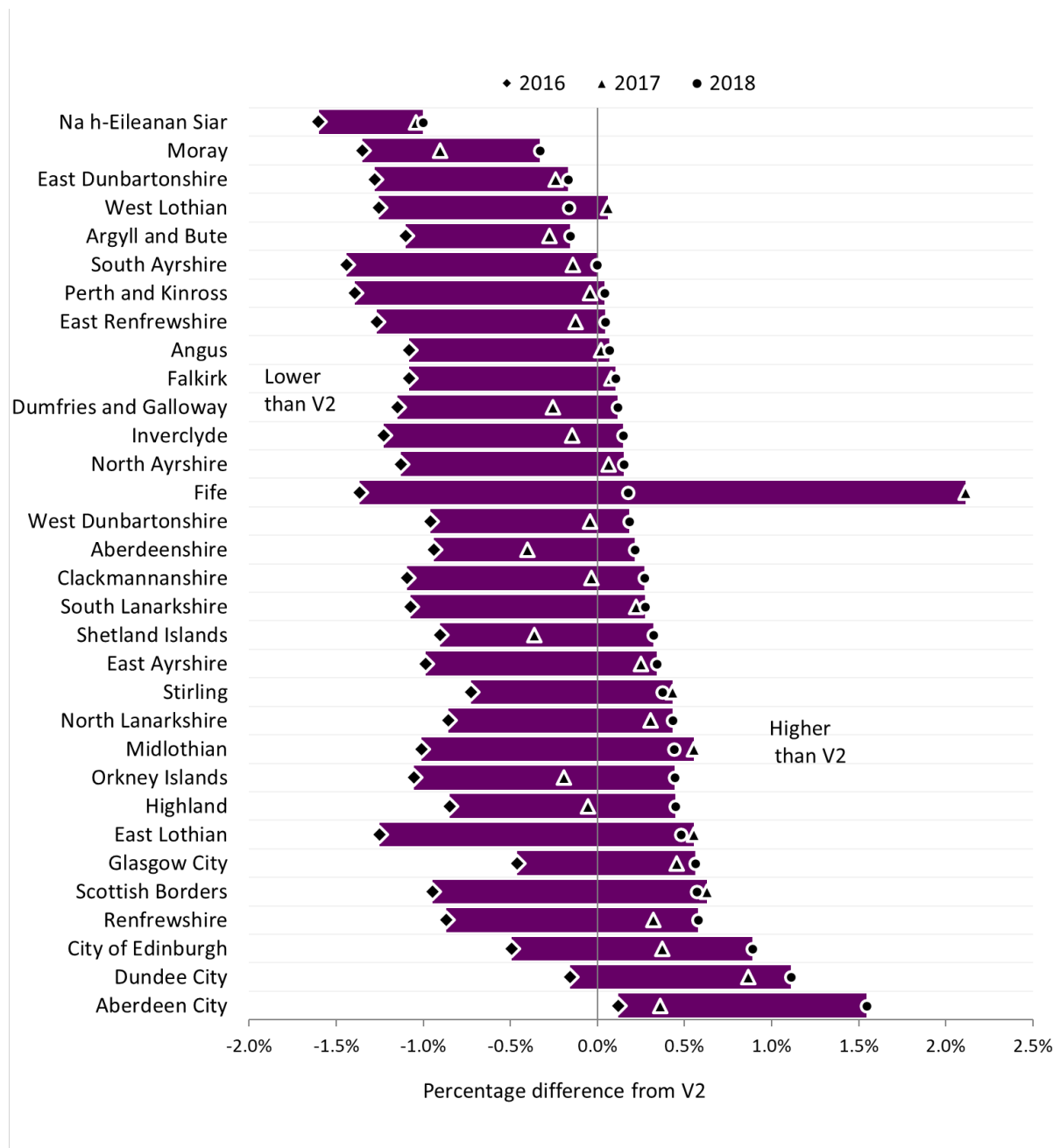
¹ Rounded to the nearest 100

have more overlap between the persons with a transfer-out flag and persons with inactive postings on the NHSCR. This will also reduce the impact of the transfer-out business rule as the people it would affect would already have been excluded from the SIDD by the NHSCR posting.

Overall, the differences between $ABPE_{V2}$ and $ABPE_{V3}$ tend to be smaller for most school ages than for adults.

Figure 5 shows the differences between $ABPE_{V2}$ and $ABPE_{V3}$ for each local authority. Most local authorities show similar patterns to the overall difference, with $ABPE_{V3}$ lower than $ABPE_{V2}$ in 2016, and about the same or higher in other years. Fife stands out as an exception, with the 2017 $ABPE_{V3}$ notably higher than $ABPE_{V2}$, reflecting the change regarding the Fife ER data.

Figure 5: Differences between ABPE_{V2} and ABPE_{V3} by Local Authority



6. Analysis of Administrative Data Based Population Estimates

Administrative data based population estimates using version 3 methodology (ABPE_{V3}) are available as [external tables](#) for:

- Scotland by single year of age and sex
- Council area by sex and single year of age and 5-year age bands
- SIMD deciles by sex and single year of age and 5-year age bands
- Urban-Rural classification by sex and single year of age and 5-year age bands
- SIMD deciles by 10-year age bands and sex and council area

These outputs have been compared with the associated National Statistics Publications covering 2016 to 2018:

- [Mid-Year Population Estimates \(MYE\), 2016 to 2018](#)
- [Centenarians in Scotland, 2010 to 2020](#)

Scotland Overall

This section presents comparisons between the ABPE_{V3} and the MYE. Table 6 shows that each version of the ABPEs are broadly comparable with the MYEs. However, in each year the ABPE_{V2} were closer to MYE than ABPE_{V3}. Recall that the aim for version 3 was to improve the accuracy of the SIDD at person level (as discussed in the Version 3 Methodology section), rather than focusing on getting the estimates closer to the MYEs. The ABPE_{V3} for 2017 and 2018 are higher than ABPE_{V2} and increasingly higher than the MYE.

Table 6: Difference between ABPE (V2 and V3) and MYE, Scotland 2016, 2017 and 2018.

Year	Difference (ABPE – MYE)	
	V2	V3
2016	+0.19%	-0.70%
2017	+0.00%	+0.29%
2018	+0.47%	+0.87%

The figures that follow all show to ABPE_{v3}. Although the overall population estimates have changed, many of the same patterns reported in the [previous publication](#) remain in place.

Figure 6: Comparison of ABPE_{v3} (2018) and Mid-2018 Population Estimates for Scotland, by age.

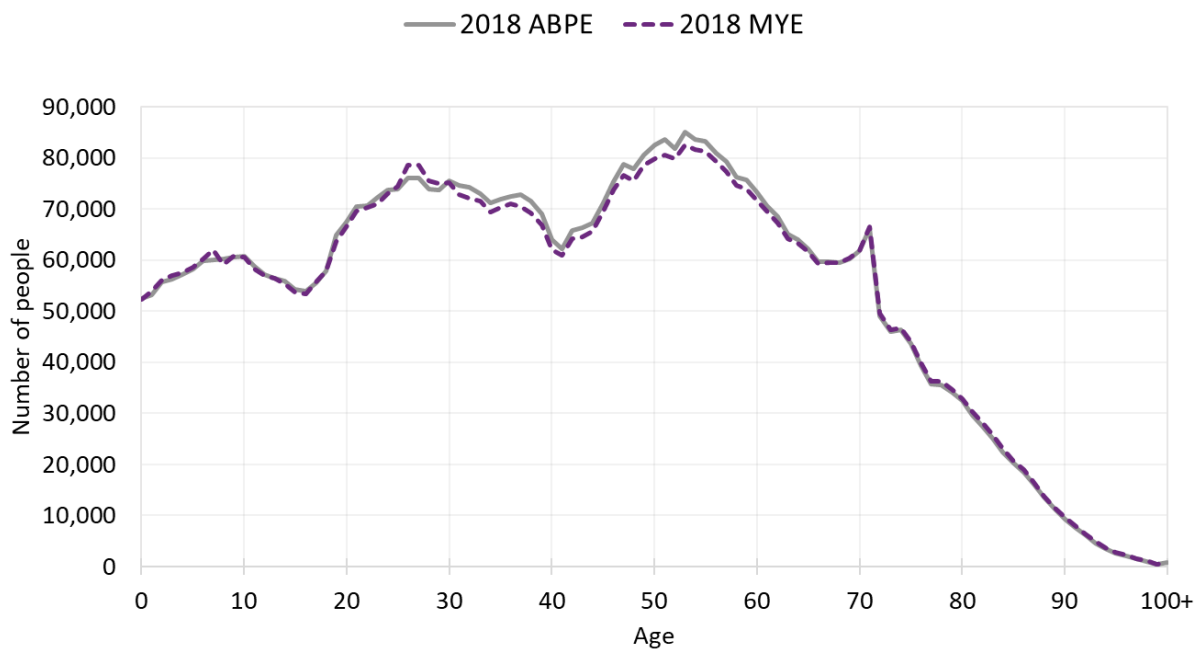


Figure 6 shows that the ABPE_{v3} and MYE for 2018 are mostly similar for all age groups.

Scotland by age and sex

Figure 7 and Figure 8 show the percentage difference between ABPE_{v3} and MYE for the three years of data by age and sex.

Although Figure 6 showed broad similarity between the ABPE_{v3} and MYE across ages, this somewhat masks differences when comparing the ABPE and MYE for males and females separately. The patterns of peaks and troughs for each sex are mirrored across all three years, much as reported under version 2 methodology. The most notable difference is in the 30 to 65 age range, where the percentage difference is relatively small for females but is consistently higher for males. The

pronounced peak at age 50 for males, is believed to be associated with bowel screening that starts at this age.

Figure 7: Percentage difference between ABPE_{v3} and MYE by age for females.

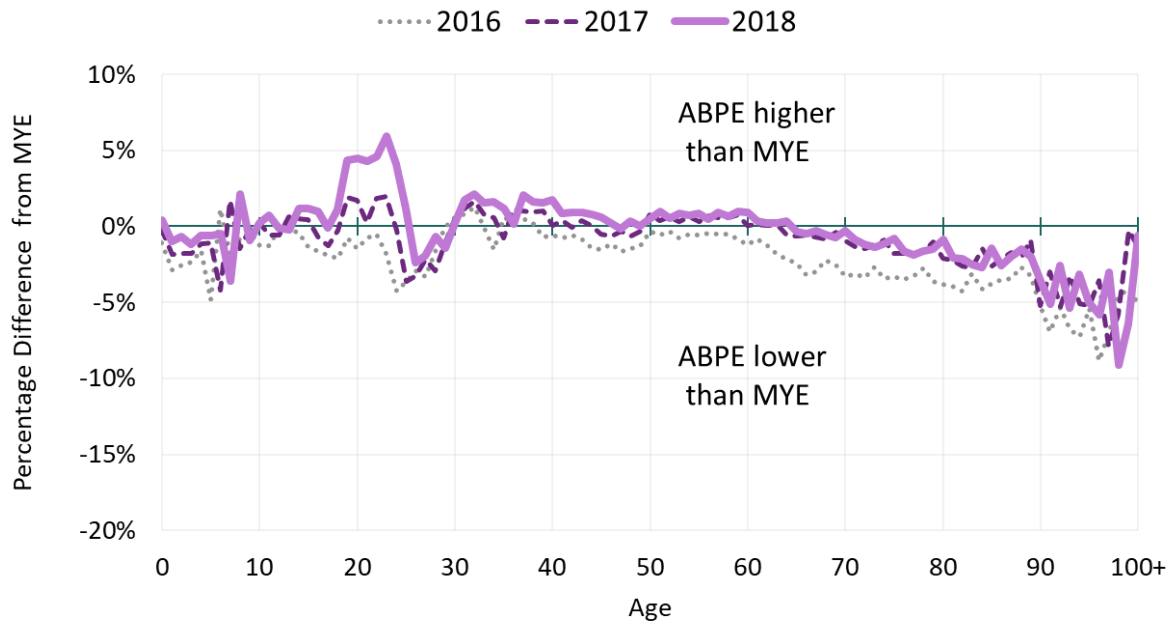
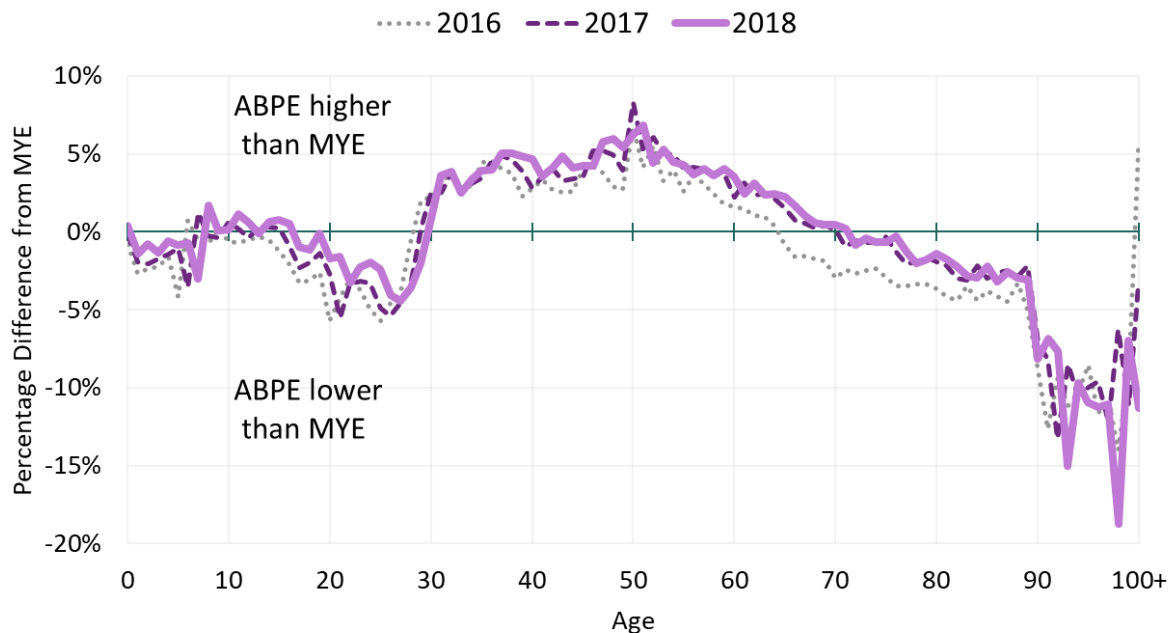


Figure 8: Percentage difference between ABPE_{v3} and MYE by age for males.



There are divergent patterns observed in the 18 to 25 age range, showing the number of females in ABPE_{v3} are higher than the MYE, compared with the male [24](#)

population estimate that is generally lower than MYE. For 25–30 year olds, the $ABPE_{V3}$ is lower than the MYE for both males and females.

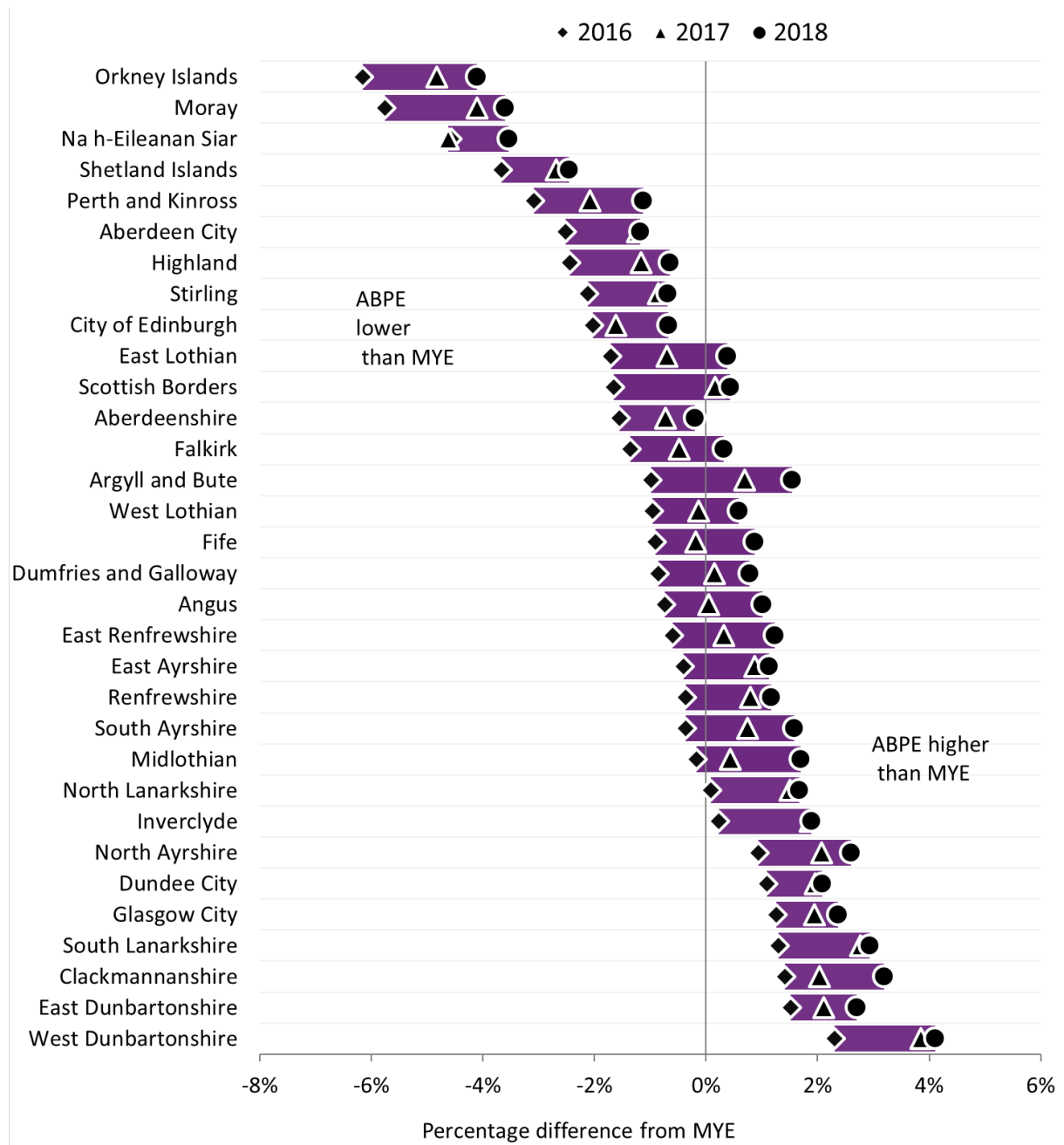
Although the percentage difference between the $ABPE_{V3}$ and MYE is very large for those aged 90 and above, there are relatively few people in those age groups, so this represents a small difference between the two estimates. For example, the near 20 per cent difference between males aged 98 is a difference of 45 people.

The data for 2016 generally tracks below that for 2017 and 2018 for both sexes. Overall, the $ABPE_{V3}$ continues to capture more males and fewer females than the MYE, similar to the $ABPE_{V2}$.

Council areas

Figure 9 shows how similar the $ABPE_{V3}$ and MYE are for the council areas. The percentage difference between the ABPE and the MYE for the majority of council areas ranges between around +/- 4%, aside from three council areas where the $ABPE_{V3}$ is much lower than MYE. It is also worth noting that those three council areas and the next lowest are predominantly rural communities: Orkney, Moray, Na h-Eileanan Siar and Shetland. It can also be seen that the large change for Fife in 2017 between $ABPE_{V2}$ and $ABPE_{V3}$ brings this close in line with the MYE, and with a similar pattern as other areas.

Figure 9: Percentage difference between ABPE_{v3} and MYE, 2016–2018 by council area, ordered by 2016 difference.



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

Limitations of analysis

These outputs are statistical research and the comparisons between them and other estimates will continue to be investigated. Although not all 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. NRS will be looking into securing new administrative datasets that could help with the coverage of Scotland's population. It is only when the ABPE is compared with the corresponding census, that we may be able to investigate further some of those differences.

Currently the estimates are counts of persons who appear on the SIDD. More-sophisticated estimates could make use of the SIDD data, but not be a direct count. A study has started to examine the possibility of using estimation techniques. These techniques include dual system estimation and weighted class estimates. This may require an extensive literature review of the survey landscape of Scotland, and an understanding of UK and international comparisons, and will be considered as part of future research in this area.

7. Future Developments

This project will continue to produce ABPEs for years beyond 2018 using version 3 methodology. In addition, there are two related statistics that will be explored using the SIDD data:

- migration between years
- administrative data based estimates on occupied dwellings.

In addition, further work will be done to investigate differences between the ABPEs and MYEs.

8. Background Note

Governance

The governance arrangements for this project were covered in Section 3 in the [previous publication](#). There have been some updates during 2023.

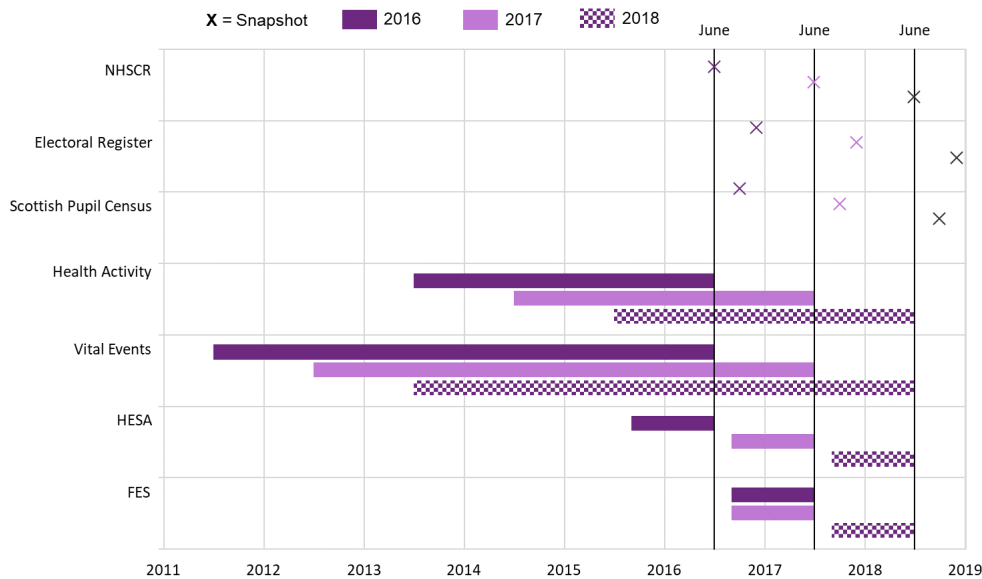
A number of updates to this project have been submitted to the Public Benefit and Privacy Panel for Health and Social Care (PBPP-HSC) and the Statistics Public Benefit and Privacy Panel (SPBPP) covering changes in personnel. The related documentation, such as the data sharing agreements, DPIA, training certificates, and so on, were also reviewed.

As part of our annual review process an updated [Data Protection Impact Assessment](#) (DPIA) has been published with this report, and has more detail on governance.

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 10.

Figure 10: Time period covered by each dataset.



Quality Assurance of Administrative Data (QAAD)

There are two QAAD reports covering this statistical research publication. The data used in this research is the same as that used for the previous publication. Whilst the underlying data has not changed, the methodology and business rules used to create the population estimates have changed.

- [Quality Assurance of Administrative Datasets, 2016](#)
- [Quality Assurance of Administrative Datasets, 2017 and 2018](#)

Stakeholder Engagements

We have been continuing to sharing the findings of the previously published statistical research with data users. Presentations have been given to various users; academics, government statistical producers and the main demographic users in Scotland.

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

Revisions

This publication presents the third version of the ABPE methodology. It is planned that all subsequent ABPEs, up to the end of the current project, will use this version of the methodology.

Methodology version 1 – [Administrative Data Based Population Estimates Scotland 2016](#)

Methodology version 2 – [Administrative Data Based Population Estimates v2 Scotland 2016–2018](#)

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.

Comparison of administrative data based population estimates within the UK

Similar approaches to producing population estimates using administrative data are being conducted by other statistical agencies nationally and internationally. Though the approaches are similar, there are differences in the different types of administrative data that are available to each organisation.

In the UK, the [Office for National Statistics](#) (ONS) have a programme of work called the [Transformation of Population and Migration Statistics](#). This work is aiming to put administrative data at the core of what they do. In 2021 ONS published [admin-based population estimates](#) for 2016 to 2020 based on counts similar to the NRS method. These ABPEs have since been re-branded as Statistical Population Datasets (SPD). ONS have now published [ABPEs based on a dynamic population model](#), which uses the SPDs alongside other data sources to produce coherent and timely estimates from administrative data.

The [Northern Ireland Statistics and Research Agency](#) (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. NISRA reviewed 'The use of administrative data in population estimates' and published their initial findings for [Northern Ireland](#) in 2014.

Comparison of administrative data based population estimates outwith the UK

Internationally, New Zealand's [Stats NZ](#) have developed their own research environment, the [Integrated Data Infrastructure](#) and have used it in producing an [Administrative Population Census](#). This is part of Stats NZ's [census transformation programme](#).

The [Australian Bureau of Statistics](#) have released an [Administrative Data Snapshot of the population](#) for June 2021, which includes information on population, income and movement within Australia.

[Statistics Canada](#) have also been looking at [creating the Canadian Statistical Demographic Database](#), initially using data from 2011 administrative data sources. Statistics Canada use administrative data sources in the production of official statistics including [population and demography](#). They also have a [census program transformation project](#), which includes exploring the use of administrative data.

Ireland's [Central Statistics Office](#) have produced [experimental statistics](#) that are based on administrative records. Their most recent [Irish Population Estimates from Administrative Data](#) attempts to estimate the population of Ireland in April 2021.

The [United States Census Bureau](#) have also started producing population estimates using administrative data. Their publication [Real-Time 2020 Administrative Record Census Simulation](#) describes their process of creating a census of the U.S. population in 2020 using administrative data.

NRS Official Population Estimates

The NRS National Statistics publications used for benchmarking purposes are:

- [Mid-Year Population Estimates \(MYE\)](#)
- [Centenarians in Scotland, 2010 to 2020](#)

These are the sources that should be used when doing any research or analysis using population statistics, and the outputs 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:

- [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 to 2018 – Tables](#)
- [Administrative Data Based Population Estimates, Scotland 2016 to 2018 – Charts](#)
- [Administrative Data Based Population Estimates, Scotland 2016 to 2018 – Interactive Charts](#)

9. Glossary

The table below provides a description of the abbreviations used in this document.

Abbreviation	Description
ABPE	Administrative Data Based Population Estimates
DPIA	Data-Protection Impact Assessment
eDRIS	electronic Data Research and Innovation Service
ERO	Electoral Register Officer
FES	Further Education Statistics
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 [UK]
PBPP-HSC	Public Benefit and Privacy Panel for Health and Social Care
PHS	Public Health Scotland
QAAD	Quality Assessment of Administrative Data
SFC	Scottish Funding Council
SG	Scottish Government
SIDD	Scotland's Integrated Demographic Dataset
SIMD	Scottish Index of Multiple Deprivation
SPBPP	Statistics Public Benefit and Privacy Panel
SPD	Statistical Population Dataset
UPID	Unique Person Identifier

10. 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 8.

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: David Rowley

For media enquiries, please contact scotlandscensus@nrscotland.gov.uk

For all other enquiries, please contact statisticscustomerservices@nrscotland.gov.uk

© Crown Copyright

You may use or re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. Further information is available within the Copyright and Disclaimer section of the National Records of Scotland website