National Records of Scotland

Administrative Data Based Population Estimates, Scotland 2017 & 2018

Quality Assurance of Administrative Datasets

Preserving the past | Recording the present | Informing the future

Contents

1. Disclaimer

The Administrative Data Based Population Estimates (ABPE) are statistical research outputs. These estimates **should not** be considered as a replacement for the National Statistics publication: Mid-Year Population Estimates for Scotland.

2. Introduction

This document summarises how the quality of the administrative data based population estimates is affected by the source datasets considering the business rules used to combine them. This document also provides details about how the data sources were quality assured prior to linkage to ensure they were suitable for this project.

This information supports our compliance with the UK Statistics Authority and the Office for Statistics Regulation's Code of Practice for Statistics. In particular this document provides evidence against the first and third principles within the Quality pillar of the Code of Practice which are listed below:

Principle Q1 - Statistics should be based on the most appropriate data to meet intended uses. The impact of any data limitations for use should be assessed, minimised and explained.

Principle Q3 - Producers of statistics and data should explain clearly how they assure themselves that statistics and data are accurate, reliable, coherent and timely.

The quality assurance arrangements for compliance with the Code of Practice were clarified in a regulatory standard issued by the UKSA in January 2015. The information in this standard was supported by an Administrative Data Quality Assurance Toolkit to provide guidance for statistical producers.

Administrative Data Based Population Estimates, Scotland published on 14 December 2021 cover the following year 2016, 2017 and 2018. The 2016 outputs have been revised but this only pertains to the methodology not the underlying data. The Quality Assurance of Administrative Dataset (QAAD) 2016 has already been published.

3. Overall quality of the Administrative Data Based Population Estimates (ABPE)

The ABPE have been produced by linking a variety of datasets. How these datasets are used is dictated by a series of business rules that are used to define which individuals are included in the ABPE. These business rules mean that certain datasets have greater importance to the creation of the ABPE and therefore have a greater potential impact on quality. Full details are described in the Methodology Report.

The dataset that is of greatest importance to the ABPE is the National Health Service Central Register (NHSCR). The business rules for inclusion on ABPE stipulate that all individuals must exist on the NHSCR, aside from those aged zero. Zero year olds will be included on ABPE if they appear in the birth registration data without appearing on the NHSCR. The birth registration data includes all people born by the reference date. The NHSCR data includes all people who were on the NHSCR on the reference date. People who were born before the reference date but had their birth registered after it, would appear on the registrations dataset, but not the NHSCR dataset.

As the NHSCR contains everyone who has registered with a GP in Scotland at any point in time, and everyone born in Scotland since 1939, there are many records for people who are no longer part of Scotland's population. Filter rules are used to reduce the dataset to those who are alive and still appear to be living in Scotland. The quality of the ABPE relies on this subset of NHSCR records (and zero year olds on birth register) having good coverage of Scotland's population.

The NHSCR still has some over-coverage even after filtering. For example this can happen when people move abroad and do not de-register from their GP. The other administrative datasets are used to provide additional evidence for an individual on the NHSCR to be retained in the ABPE, or removed.

Comparisons of the quality of data over 2016, 2017 and 2018

One of NRS's future developments was to try and understand how important the quality and consistency of the underlying datasets was for the creation of ABPE. Over time, the purpose for which administrative datasets are maintained can change for organisations. They may have to collect more data, change underlining guidance or policy (for example, a change in the law) or they may have to reduce data collection. It was important to look at the data over the three years on the same methodology to make sure that there was not a large unexplained fluctuation within a cohort. The chart below shows that the cohorts are rolling forwards as expected. For example, the peak in population of people born in 1947 in the ABPE was 68,000 in 2016 (aged 69); this has reduced to 66,000 in 2018 (aged 71) as the population ages on we would expect a natural decrease due to death rate increasing.

Figure 1: ABPE by age, Scotland 2016 to 2018
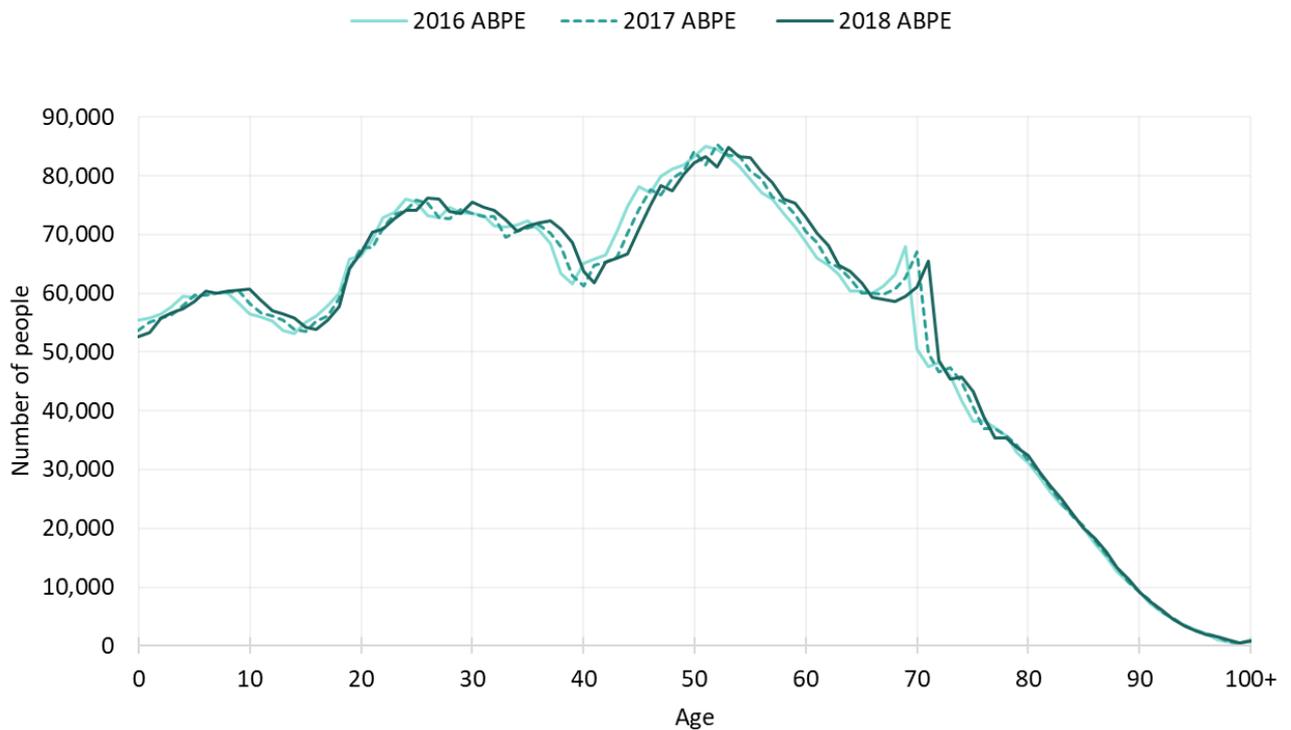
One of the more interesting aspects of those data was that the consistency between the three years of ABPE was still present at lower geographies. We are aware that there are differences between the MYE and ABPE but it was interesting to see whether the increase/decrease observed in council area MYE over the two years was being reflected in the ABPE figures, and if it was at a comparable rate. Figure 2 shows the change in the population between 2016 and 2018 for each council area. Amount two thirds of the ABPE are slightly more than MYE. Only four council areas have the changes in opposite directions in the MYE and ABPE. These four council areas population differences are in the hundreds. For example, the biggest difference in numbers is for Moray with ABPE increasing by 610 persons and MYE decreasing 550. Overall there seems to be reasonable consistency in population trends between the years.

Figure 2: Percentage change for ABPE and MYE by council area, 2016–2018
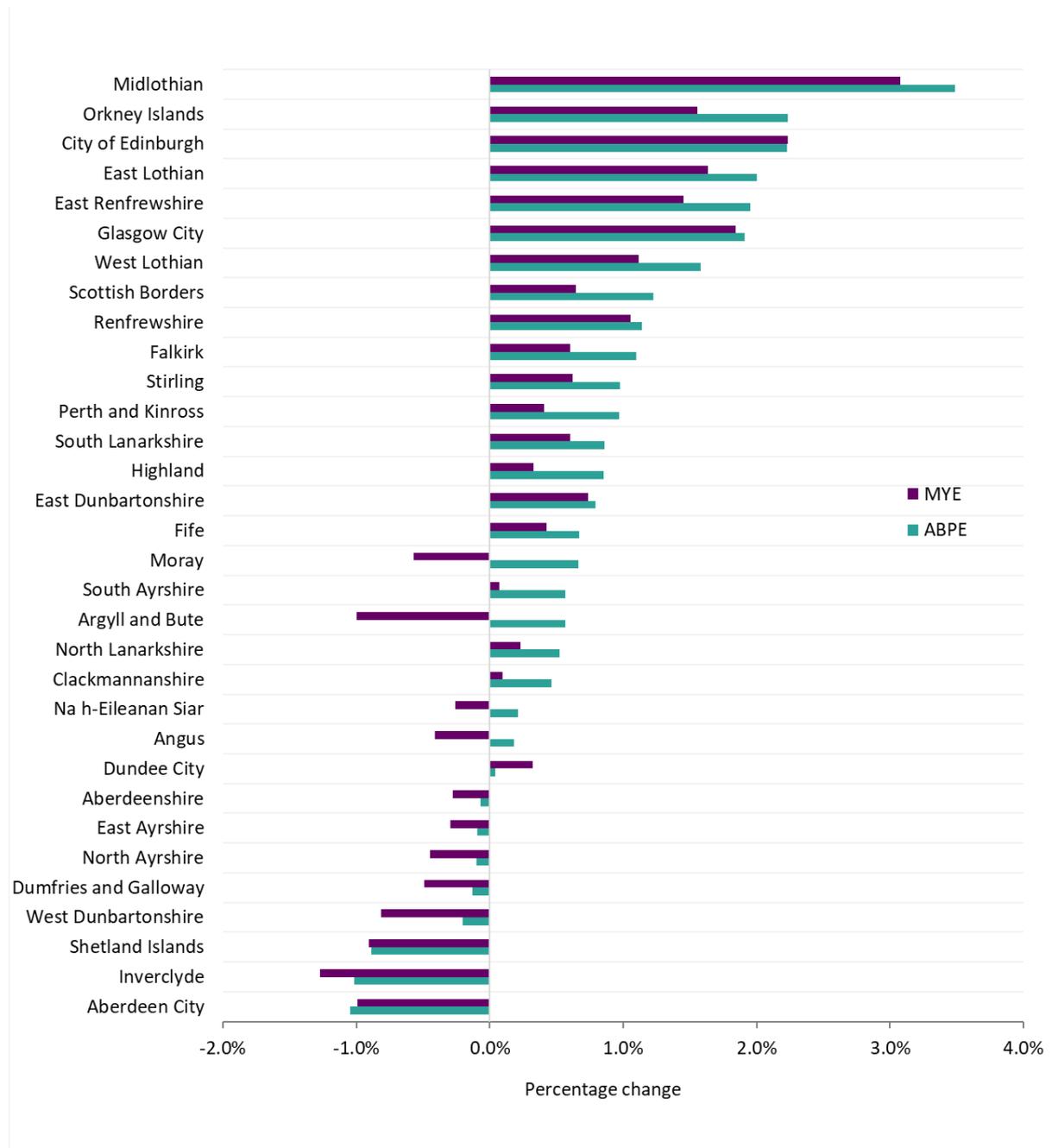
Known data issues in 2017

The electoral registration data for Fife was not included in the 2017 results, due to a clerical error at NRS. The 2017 passwords were accidently overwritten in 2018 and this was not noticed until the data was to be processed this year. NRS investigated this matter with Fife Electoral registration. The data is securely held at NRS but due to staff changes neither organisation could find the original password to un-encrypt it. Subsequent IT changes over the four years meant that Fife was unable to replicate the original extract. NRS decided to process 2017 with this omission, considering it as a good test of how robust the methodology was in coping with missing data.

The findings of this have been very interesting to this statistical research. Fife ABPE for 2016 and 2018 were above MYE by 0.46% and 0.70% respectively. Whereas the 2017 ABPE was below MYE by 2.25%. If the trend was followed we would have expected the 2017 percentage difference between the MYE and ABPE to be approximately 0.50%. This would have been a loss of around 10,200 persons from the analysis. Overall this is 2.75% of Fife's ABPE population for 2017, so the methodology did cope reasonably well with the loss of one dataset but we would clearly not want this to happen in the future years. This should be noted, when looking at Fife's ABPE against other council areas for 2017.

At the moment, NRS can only quality assurance the datasets as we individually process them. It is not until the datasets are de-identified and transferred to the Safe Haven that we can quality assurance them against the data NRS has received from other organisations. A correction of some postcodes for Health Activity 2017 needed to happen. Some people had the same de-identified postcode in Health Activity 2016 and 2018, and NHSCR over the three years, but a different de-identified postcode on Health Activity 2017. For those people the Health Activity de-identified postcode in 2017 was changed to be consistent with the NHSCR and the Health Activity data for the other years.

NRS feel that the quality of the 2017 data is slightly poorer than 2016 and 2018, but as this is only one council area with an estimated undercount of 3%, it would be unwise to omit 2017 data from this statistical research. Had this publication been designated as official statistics this would have had a more serious impact, but as this is statistical research, these issues have highlighted how the quality of administrative datasets can impact ABPE, and is considered an important finding in itself. The methodology used does compensate for some of these issues by using information from other datasets. It highlights two important issues going forward:

- Having additional good quality datasets could help improve the business rules in the future to deal with issues like this.
- The possible use of a statistical methodology like Dual System Estimation from administrative survey or other population administrative datasets could

improve the ABPEs as that would provide better population estimates than counts

The quality assurance of the 2017 and 2018 ABPE have the following strengths and limitations:

Strengths

- By requiring that individuals appear in more than one dataset, some over-coverage of the NHSCR is mitigated.
- For datasets other than the NHSCR and birth registrations, the impact of any potential over-coverage will be reduced as this will be mitigated by the person also having to appear on the NHSCR as alive and resident in Scotland.
- Estimates are produced from a dataset of de-identified data at individual level rather than being produced from aggregate counts. This has the potential to allow accurate migration information to be produced by linking data across different years.
- More investigation into the Health Activity datasets has allowed NRS to adjust the timeframe on interaction for different ages
- A better understanding of the dataset interactions now there are three year's worth of data, thus improving the business rules for the methodology.

Limitations

There are several reasons why someone who is part of Scotland's population may be missing from the ABPE. These include:

- Any individuals who have not registered with a GP in Scotland and were not born in Scotland will be excluded as they will not be part of the filtered subset of NHSCR records.
- Some individuals who are part of Scotland's population will appear on the NHSCR, but will not be present in any of the other datasets. These people will be removed through the application of the business rules.
- Linkage is not perfect, and therefore inconsistencies with how an individual's data is recorded between datasets will mean that some links are missed. These inconsistencies could be caused by errors during data collection, or by the individual providing different information for each data collection. This could lead to a person being wrongly excluded from the ABPE where they appear on the NHSCR but are not uniquely located on any of the other datasets due to a missing link when in fact they are present on one of the datasets.
- Individuals may be included in the ABPE when they should be removed. For example, if an individual has not informed their GP that they have moved out of Scotland, they could still be included. If GP records are not updated, individuals may still appear on the NHSCR as living in Scotland. They may also appear on other datasets such as the Health Activity dataset if they had

8

used health services within the reporting period, but prior to their departure date, and therefore are wrongly included in the ABPE.

- If data is not available it may have an adverse impact of the ABPE unless the methodology can compensate.
- Differences in the reference period for each dataset, shown in Figure 3, will lead to some inconsistencies in the data. This will mean links between datasets could be missed if the information about a person changes during those times, for example if they move home or change their name.

Figure 3: Source dataset reference periods
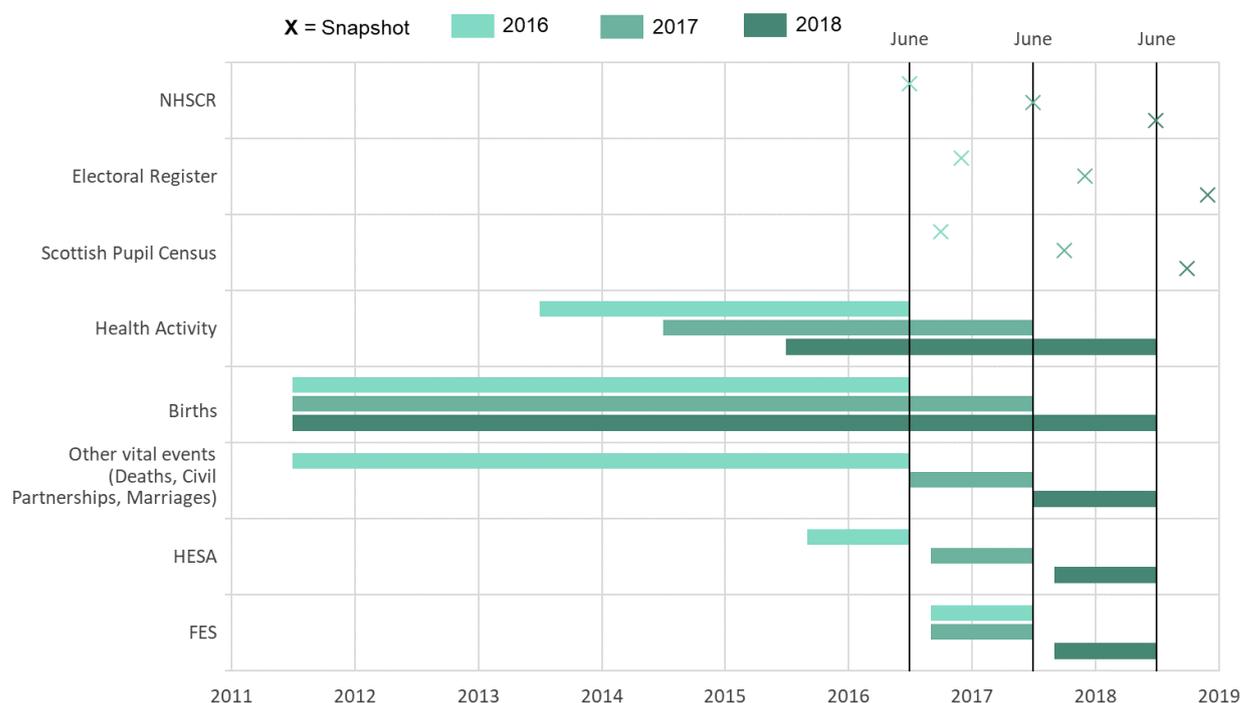
Risk/Profile Assessment

The matrix below reflects the levels of risk of data quality concerns and the public interest profile of the ABPE. These have been determined by a review undertaken by the NRS Administrative Data team using the information contained within the Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit.

Level of risk of quality concerns	Public interest profile		
	Low	**Medium**	**High**
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit.

The Public Interest profile has been set as "medium" for the following reasons:

- One of the objectives of the ABPE is to support future recommendations for the census beyond 2022.

- There is a strong interest in the viability of ABPE to maximise the use of all available data sources to provide accurate and timely evidence to measure our population.

The risk of quality concerns has been set to "medium" for the following reasons:

- The ABPE have produced figures that are broadly comparable at Scotland level with the official mid-year population estimates. These results are encouraging however we are aware that future improvements to the methodology and possible additional datasets are required to further improve the quality of the estimates. This is discussed in the [main publication](#) and the [methodology report](#).

- Several administrative datasets are provided by external data suppliers. This means that the data could be subject to change from year to year depending on requirements of the data for that supplier. We will continue to communicate with data suppliers to understand the data they provide and how any changes could impact this project.

4. Source dataset information

National Health Service Central Register (NHSCR)

Data Supplier:	National Records of Scotland (NHSCR)
Supplier info:	National Records of Scotland (NRS) is a Non Ministerial Office of the Scottish Government. The purpose of NRS is to collect, preserve and produce information about Scotland's people and history and make it available to inform current and future generations. The NHSCR branch of NRS is responsible for maintaining the NHSCR, an electronic demographic database of all people born in Scotland, died in Scotland and those who have ever registered with a GP in Scotland.
Data type	Unit records
Data Content:	The following variables are included at an individual record level: • First name • Middle name • Last name • Previous names • Sex • Birthdate • Birth country • Death date • NHS Number (Scottish, England/Wales and Northern Irish numbers) • Person ID • Postcode • Date postcode was recorded • Posting (indicates which health board the person has registered to a GP in)
Time Period Covered	Extracts as at 30 June 2017 and 30 June 2018
Use of Data:	Production of administrative data based population estimates as statistical research

Data Source Information

The NHSCR is an electronic index for:

- every patient registered, now or in the past, with a Scottish general medical practitioner (GP);
- everyone born in Scotland since 30 September 1939, who have not been registered with a Scottish GP;
- patients formerly registered with a Scottish GP, who died after 29 September 1939.

The main purpose of the register is to permit the efficient movement of patient's medical record envelopes when they:

- transfer between Scottish Health Boards and health authorities in the rest of the UK;
- leave the country;
- join the Armed Forces (or are dependants of Armed Forces personnel).

The key inputs into the NHSCR are:

- Births (in Scotland);
- Deaths (from across the UK);
- GP Registration (within Scotland) – 'migration' into Scotland;
- GP Registration (within the rest of the UK) – 'migration' out of Scotland.

Data supply and communication

The data provided is done so annually under the terms of a data sharing agreement and includes record level data for a selection of variables as defined in a data sharing agreement for every person on the NHSCR.

The data is sent to the admin data team by the NHSCR team (who receive the extract from Atos) via approved NRS data transfer procedures as agreed in a data sharing agreement.

Quality Assurance undertaken by data supplier

The data entered by staff is regularly scrutinised. Supervisors check 5% of the work undertaken by staff each day to identify any potential training issues. These records are randomly selected based on subject matter, taking into account new areas of work, trends or concerns previously identified. This also helps the NHSCR to meet its service level agreement with the Scottish Government, NHS National Services Scotland which requires an accuracy level of 97%, which is currently being achieved.

As well as this, the NHSCR team undertake a variety of data quality initiatives on an annual/bi-annual basis where staff investigate the population of different variables in the register and to correct duplicates. These initiatives are carried out relatively frequently as they target areas of known concern and the findings are generally kept internal to the NHSCR team. These data quality initiatives include:

- investigating records where no death has been recorded for a person aged over 110 years old. In the majority of cases a death is traced (these are usually deaths that were missed at the time, usually from the 1970s or 1980s before the NHSCR was computerised) and the record is updated to reflect this.
- checking records where the postings variable is blank. This allows us to be confident that all records that should have a posting do. Where no posting exists it is usually for persons who are born in Scotland but they never registered with a Scottish NHS GP.
- populating records that do not have a Community Health Index (CHI) number[1] either with the CHI number if one exists or with a flag to show that there is not a CHI number for that record.

Extracts of the NHSCR are used by various statistical teams across the National Records of Scotland for a variety of purposes. NHSCR also collects feedback from these users of the NHSCR extracts where anomalies are identified and investigates these anomalies so a resolution or explanation can be found.

Quality Assurance undertaken by the admin data team within NRS

Once the admin data team receive the data, a number of data consistency and validation checks are performed, including:

- Checking the proportion of missing values for variables;
- Checking the validity of postcodes;
- Checking the distribution of the population across different council areas and comparing this to previous years and/or existing population estimates;
- Checking the distribution of the day and month elements of dates of birth;
- Checking the age distribution of the population;
- Checking that variables that should be unique are unique.

These checks are largely programmed with the output flagging up any anomalies, although analysts do also look at a small sample of records to spot any issues.

---

[1] https://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?ID=128&Title=CHI%20Number

If these checks suggest the data may need to be amended/adjusted then the potential issues are communicated with the data supplier so the register can be amended if appropriate. However in this case these checks did not identify any issues with the data so this was not required.

Strengths and Limitations of the NHSCR data source

Strengths	Limitations
• NHSCR is a comprehensive source of record level data that covers the vast majority of the population in Scotland. • The data contains all of the variables used to link with other data sources (name, date of birth, postcode and sex)	• Generally does not include address information beyond postcode. There is a Unique Property Reference Number (UPRN) variable, however this variable is completed for less than 25 per cent of records. • It does not pick-up people who leave the UK (unless they informed their GP) leading to some inflation in the register • Moves within Scotland cannot be picked up until the patient registers with a new GP. As a result some people will be recorded in the wrong area. Particularly an issue among younger adult males[2]. • There will be a lag in recent migrants into Scotland appearing on the NHSCR as they will only appear when registering with a GP. • There is a delay in new born babies appearing in NHSCR with a postcode (and posting) until they are registered with a GP.

[2] Page 18 of the Mid-Year Population Estimates Methodology guide: "It is acknowledged that NHSCR flows undercount the number of migratory moves for young men in particular, due to General Practitioner (GP) registration behaviour in different groups."

https://www.nrscotland.gov.uk/files//statistics/population-estimates/mid-19/mid-year-pop-est-19-methodology.pdf

Health Activity

Data Supplier:	Public Health Scotland (PHS)
Supplier info:	Public Health Scotland is Scotland's lead national agency for improving and protecting the health and wellbeing of all of Scotland's people. PHS's vision is of a Scotland where everybody thrives. PHS's focus is on increasing healthy life expectancy and reducing premature mortality. To do this, they use data, intelligence and a place-based approach to lead and deliver Scotland's public health priorities.
Data type (counts or unit records)	Unit records
Data Content:	The following variables are included at an individual record level: • Unique ID • Surname • First Forename • Second Forename • Previous Surname • Date of Birth • Sex (Gender) • Patient Structured Address • Full Patient Postcode • General Practitioner Practice Postcode • Row ID Additionally, PHS send a Last Interaction variable along with unique linking identifiers to the National Safe Haven. Identifiers allow linking of Primary and Secondary data files, with RowID above, serving as the linking variable for Last Interaction.
Time Period Covered	Data extract at 30 June 2017& 2018, with 'Last Interaction' variable covering previous 3 years
Supply Schedule:	Annually
Use of Data:	Production of statistical research on administrative data based population estimates.

Data Source Information

The Community Health Index (CHI) is the main linking key which is used in Scotland for health care purposes. The register exists to ensure that patients can be uniquely identified, and that all information pertaining to a patient's health is available to providers of care. No single body has responsibility for CHI; the data controllers for CHI are the 14 National Health Service (NHS) Boards. An extract called the Health Activity Dataset was created for this project by PHS. No individual health data was supplied, only an activity flag of last time they used a NHS service.

The variable of interest for project is 'Last Interaction'. This variable reports date of an individual's last engagement with a health practitioner (General Practitioner, Accident & Emergency, Day Case and Outpatient Hospital appointment, Dentist, Community Pharmacist and Dispensing Contractors delivering primary care across Scotland), providing an up-to-date population register that can help confirm population estimates in any time period. This variable is sent directly to our secure processing site (National Safe Haven) by PHS, with unique identifiable key for subsequent linking as per their data processing agreement.

Data supply and communication

Under the terms of a data sharing agreement, the data is provided annually and transferred securely to NRS.

The health activity data is provided in separate files for primary care and secondary care, with a smaller supplementary secondary file in 2017 (batch 2.1). Primary Care data covers interactions with Dental Services, Pharmacies and Prescribing, Bowel Screening, and Abdominal Aortic Aneurysm (AAA) screening. Secondary Care data covers interactions with Hospitals, including Outpatients (SMR00), Inpatients and Day cases (SMR01), Maternity (SMR02), Mental Health (SMR04), Cancer Registrations (SMR06), and Accident and Emergency. The number of records in each time period are noted in Table 1, where Heath Activity 2017 covers the three-year period from 30 June 2014 to 1 July 2017 and Health Activity 2018 covers 30 June 2015 to 1 July 2018.

Table 1: Number of records in each Health Activity data file processed by NRS

Dataset	Number of records* Health Activity 2017	Number of records Health Activity 2018
Primary	5,627,000	5,658,000
Secondary	3,727,000	3,747,000
Secondary (2.1)	7,000	---

* rounded to the nearest thousand

Quality Assurance undertaken by data supplier

PHS perform internal quality assurance processes before sharing data. General data management includes checks on completeness and timeliness, with dataset specific checks as set out in their publication on Quality Assurance Process at About our statistics - Data & intelligence from PHS (isdscotland.org)

Completeness – NHS data providers will know how complete their Scottish Morbidity Record (SMR) datasets are and the extent of any backlog. SMR data is expected to be received by PHS 6 weeks following the end of the month of discharge or clinic date. In this period the target has been achieved with a national return of 99 per cent or higher as sourced at https://www.isdscotland.org/products-and-Services/Data-Support-and-Monitoring/SMR-Completeness/ although some Health Boards obviously fall short of that average with only 97 per cent completion rates.

Timeliness –The Scottish Government target for SMR submission to PHS is 6 weeks (42 days) following discharge/transfer/death or clinic attendance. PHS calculates timeliness as data received 6 weeks following the end of month of discharge/transfer/death or clinic attendance, tracking any backlog as well as highlighting number of records that were submitted after the 6-week target. https://www.isdscotland.org/products-and-Services/Data-Support-and-Monitoring/SMR-Timeliness/

Four main entries from the Scottish Morbidity Record (SMR) datasets feed into the Health Activity dataset, namely:

   – SMR00 Outpatients
   – SMR01 General Acute Inpatients/Day Cases
   – SMR02 Maternity Inpatients/Day Cases
   – SMR04 Mental Health Inpatients/Day Cases

Validation is either carried out locally and prior to submission to PHS or centrally at PHS. A set of validation rules is carried out by the data provider, where checks may generate:

   – Errors where the information recorded is missing, invalid or fails to conform to a logical sequence of events, or
   – Queries where the information recorded appears to be infeasible but is found to be to be correct.

Automatic checks are made to see if a record already exists with the same or similar DOB, Name, Gender, and Address. Validation on address is performed by looking up Quick Address Software (QAS). PHS rely on users who have update access to enter address information correctly, with address changes triggered by patients through GP system or added by hospitals for new patients not yet registered with a GP. The National Health Service Central Register (NHSCR) is used to update the main PHS records on changes/embarks from Scottish Health Boards, but NHSCR is not involved in addressing of PHS records i.e. they are independent of one another insofar as data entry is concerned.

Quality assurance measures are in place for data that is sourced from other Primary care providers:

- Dentistry Annual Report available at [Primary Care Dentistry in Scotland • Annual Report 2017/18 (isdscotland.org)](#) but there is no link to QA measures within that data collection
- Community Pharmacist and Dispensing Contractors – see Metadata link on https://www.isdscotland.org/Health-Topics/Prescribing-and-Medicines/

Accessing precise practices and data for 2017 and 2018 proved problematic as those years bridged the transition from ISD to the new Public Health Scotland website. The older ISD website is no longer maintained and even though this information was published at the time, a number of the newer PHS pages have broken links to the previously published information.

Speaking with our colleagues in PHS, they were not aware of any quality assurance issues that would impact this 2017 and 2018 data. Going forwards, the NHS Performs platform and PHS Data and Intelligence website will provide a source of data quality assurance.

Quality Assurance undertaken by the Admin Data team within NRS

Once the Admin Data team receive the data, a number of data consistency and validation checks are performed on Health Activity dataset data prior to standardising variables, de-identification and transfer to safe haven. Those checks include:

- Checking the proportion of missing values for variables.
- Checking the validity of names (First, Middle, Last, Previous Last), UPRN and postcodes.
- Sense checking the number of records by single year of age compared to published information (Mid-Year Estimates for 2017 and 2018).

These checks provide additional information to NRS team when linking data to produce population estimates in the safe haven.

Strengths and Limitations of the data source

Strengths	Limitations
• Health Activity dataset is a comprehensive source of record level data that covers the vast majority of Scotland's population • High quality data administered by PHS. Also able to use an active flag that gives us a time indication for interaction with the Health service.	• Moves within Scotland cannot be picked up until the patient registers with a new GP. As a result some people will be recorded in the wrong area. Particularly an issue among younger adult males. • Due to the number of datasets being used to create this dataset there may be a small percentage that are not linked correctly.

Scottish Pupil Census (SPC) 2017 and 2018

Data Supplier:	Scottish Government: Education Analytical Services (EAS)
Supplier info:	EAS provides data on school pupils through an annual pupil census that captures characteristics of pupils. This QAAD is based on the data from the censuses that took place in September 2017 and 2018. The SPC forms part of 'Summary statistics for schools in Scotland', an annual publication that describes the education system in terms of the number of schools and pupils, the types and sizes of schools and classes they learn in, and some characteristics of the pupils.
Data type (counts or unit records)	Unit records
Data Content:	The Pupil Census covers all publicly funded schools in Scotland (local authority and grant-aided). Pupils in this census are those recorded by a Local Authority (LA) as being on the roll of the school, except those in full time education at another institution. The following variables are included at an individual pupil record level : • Scottish Candidate Number (SCN) • Home postcode • Sex • Date of Birth • Ethnic background (self-identified from categories used in 2011 Census) • School SEED code (Identifier)
Time Period Covered:	2017 and 2018 Scottish Pupil Censuses
Use of Data:	Production of administrative data based population estimates as statistical research

Background Information

Data is collected from all Local Authority and Grant-aided schools and school centres. All local authorities use the same management information system called SEEMiS. This makes it easier to ensure consistency across local authorities in how they record information. There are checks done before data is submitted to ScotXEd by SEEMiS, then futher validation checks are done by ScotXEd before statisticians in Learning Analysis do more detailed checks. This rigorous process ensures the data is as accurate as it can be given the time constraints of the collection.

The data collected is included in the National Statistics publication 'Summary statistics for schools in Scotland' for each year:

- 2017: https://www.gov.scot/publications/summary-statistics-schools-scotland-8-2017-edition/
- 2018: https://www.gov.scot/publications/summary-statistics-schools-scotland-9-2018/

Data supply and communication

The data is provided to NRS by EAS annually under the terms of data sharing agreement and includes record level data for a selection of variables as defined in the data sharing agreement for every pupil based on unique identifiers of SCN and SEEMiS Student ID.

Quality Assurance undertaken by data supplier

The data collected by EAS is primarily taken from local authority management systems. The fact that the information collected is that actually used by LAs in local management of the education system has proven to be a strong driver in ensuring that data are correct.

Local authorities supplying data have built in validation checks in SEEMiS and the procXed Data Collection System; validation checks agreed with data providers are regularly updated, and Head Teachers sign off summary tables that are used.

Scottish Government has a wider set of built in validation checks so that errors or queries can be identified as early as possible. The validation checks have usually been agreed on consultation with data providers and are regularly updated.

Once automated validation checks and queries have been finalised, further sense-checks are completed by statisticians and other colleagues with knowledge of the sector.

Quality Assurance undertaken by National Records of Scotland (NRS) Admin Data team

Once the admin data team receive the data, a number of data consistency and validation checks are performed, including:

- Checking the proportion of missing values for variables
- Checking that variables are in the expected formats and values
- Checking the validity of postcodes
- Comparing the data with similar data received in previous years and investigating when there appear to be significant changes.
- Checking the distribution of the day and month elements of dates of birth
- Checking the age distribution of the population.
- Removing duplicate records where identical information is recorded

Strengths and Limitations

Strengths	Limitations
• SPC data is a comprehensive source of record level data that covers the vast majority of school age population. • High quality data administered by LA through ScotXed and EAS division of Scottish Government. • Data includes home postcode making SPC a good dataset for creating/confirming or validating administrative household estimates. • SPC is an annual data collection that the Scottish Government has run for decades and it is classified as a National Statistics publication.	• Name is not collected by EAS and linking methodology in the project is modified to reflect this. • Full address information is not collected by EAS –only having postcode may limit linking exercise. • For the data discussed in this document, the extracts requested by NRS did not account for pupils who attend more than one school. Therefore a limitation of these extracts is that it is not possible to identify a pupil's main school for these pupils. However the full dataset does contain a variable that allows the main school to be identified and will be included in future data extracts supplied to NRS from 2021. • No information on independent sector, home schooling etc. as out of the scope of this data collection.

22

Higher Education Statistics Agency (HESA)

Data Supplier:	Higher Education Statistics Agency (HESA)
Supplier info:	HESA are the experts in UK higher education data. They collect, assure and disseminate data about higher education (HE) in the UK on behalf of their Statutory Customers. HESA works with HE providers in each of the four nations of the United Kingdom, collaborating with them to collect and curate one of the world's leading HE data sources.
Data type	Unit records
Data Content:	The following variables are included at an individual record level : • Forename(s) • Surname • Surname at 16 if different from above • Sex • Birthdate • Nationality • Term-time postcode • Unique Identifiers (Unique Learner Number, Scottish Candidate Number, HESA Unique Student Identifier) • Postcode of permanent home address • Date studies started • Date studies ended • UKPRN (UK Provider Reference Number - for establishment registered at) • Expected Length of study • Year of student instance • Year of course • Location of study • Suspension of active study flag The population covered in this data is all students studying at Scottish higher education providers (including The Open University) and Scottish domiciled students studying at higher education providers in England, Wales and Northern Ireland.
Time range covered	2016/17 and 2017/18 academic years

	HEIs collect data in period August to July that is returned to HESA online by 01 October of end year, covering all enrolments during the entire academic sessions e.g. 01 Aug 2017 to 31 July 2018, reported on 01 October 2018
Use of Data:	Production of administrative data based population estimates as statistical research

Data Source Information

The HESA Student record has been collected since 1994/95 from subscribing Higher Education Providers (HEPs) throughout the devolved administrations of the United Kingdom. The data collected as part of the Student record is used extensively by various stakeholders and is fundamental in the formulation of:

- – Funding
- – Publications (including UNISTATS & Performance Indicators)
- – League tables

The aggregated figures from this data are used by HESA in their annual National Statistics publication 'Higher Education Student Statistics: UK', links for the relevant years for the data used here are provided below:

2016/17
https://www.hesa.ac.uk/news/11-01-2018/sfr247-higher-education-student-statistics

2017/18
https://www.hesa.ac.uk/news/17-01-2019/sb252-higher-education-student-statistics

2018/19
https://www.hesa.ac.uk/news/16-01-2020/sb255-higher-education-student-statistics

HESA's Quality Report (link below) provides some additional information on uses of student data in the 'Relevance' section.
https://www.hesa.ac.uk/about/regulation/official-statistics/quality-report

For the years covered in this report, the Student record collects individualised data about students active during the reporting period. The reporting period is from 01 August year 1 to 31 July year 2, for example, the 2017/18 Student record was collected in respect of the activity which took place between 01 August 2017 and 31 July 2018.

Data supply and communication

The data is supplied by Higher Education providers to HESA via a secure web-based transfer system created and maintained by HESA. The data supplied are subject to an extensive quality assurance process.

The data provided to NRS by HESA is shared under the terms of a data sharing agreement. The data includes record level data for a selection of variables for all students studying at Scottish higher education providers (including The Open University) and Scottish domiciled students studying at higher education providers in England, Wales and Northern Ireland.

HESA publish extensive information about the collection of the data, the validation process used and any known issues with the data on their website.

For 2016/17 this information is found at:
https://www.hesa.ac.uk/collection/c16051
https://www.hesa.ac.uk/collection/c16051/support-guides

For 2017/18 this information is found at:
https://www.hesa.ac.uk/collection/c17051
https://www.hesa.ac.uk/collection/c17051/support-guides

Quality Assurance undertaken by data supplier

HESA produce a student record quality report[3] that explains how they assure themselves that the data is accurate, reliable, coherent and timely.

As mentioned in the 'Data supply and communication' section, HESA has developed extensive quality assurance procedures and runs a range of automated validation checks (quality rules) against all submissions from data providers. When submitting final data the provider must pass various rules that ensure the data is in the correct format and does not trigger any validation errors. In the situation that correct data still triggers these validation errors, the provider must contact HESA to provide an explanation.

These rules4 include, but are not limited to:

- checking unique identifiers are valid by using a checksum

---

[3] HESA's Quality Report https://www.hesa.ac.uk/about/regulation/official-statistics/quality-report
[4] Quality rules for:
  2016/17: https://www.hesa.ac.uk/collection/c16051/quality-rules
  2017/18: https://www.hesa.ac.uk/collection/c17051/quality-rules
  2018/19: https://www.hesa.ac.uk/collection/c18051/quality-rules

- providing a warning when personal information submitted for a student does not match the previously sent information for the student.
- only allowing dates of birth to be in a certain range if date of birth is provided
- showing an error if it appears that forename and surname have been transposed compared to the last year's submission.
- warning if more than 2% of students have 'other' recorded for sex in case this is due to a systematic error.
- error if all students have been returned with the same sex code as a range of codes is expected
- warning or error if the number of students have the same term-time postcode without being marked as living in provider maintained property or halls of residence exceeds specified thresholds
- a postcode must be recorded for all UK domiciled students

Data Quality Analysts at HESA then examine the data to ensure the submission is credible. This is an iterative process during which providers may need to submit and review several times before signing off the data to ensure the final submission is credible.

Quality Assurance undertaken by the admin data team within NRS

Once the admin data team receive the data, a number of data consistency and validation checks are performed, including:

- Checking the proportion of missing values for variables
- Checking that variables are in the expected formats and values
- Checking the validity of postcodes
- Comparing the data with similar data received in previous years and with published data about students in Scotland to check that trends and patterns appear to be correct.
- Checking the distribution of the day and month elements of dates of birth
- Checking the age distribution of the population.
- Removing duplicate records where identical information is recorded (this can occur if an individual enrols on multiple courses in the academic year).

If these checks suggest the data may need to be amended/adjusted then the potential issues are communicated with the data supplier so the data can be amended if appropriate. However, this was not required after checking these data sets.

Strengths and Limitations of the HESA data source

Strengths	Limitations
• A considerable proportion of records in this data are for young adults who can be difficult to identify in other datasets. This dataset should therefore be particularly valuable in improving estimates of young adults. • As the data includes term-time and home postcode, it may be able to resolve issues where postcodes differ for one individual in other datasets. • Contains some previous surname information so have an improved chance of making links where surname has changed. • Extensive validation process by the data supplier and HESA to make the data as complete as possible.	• There is a lag in being able to receive the data. For example 2017/18 data is only available in early 2019. This could therefore impact on when the most up-to-date population estimates could be published but was not an issue for this publication. • Only provides data on a specific subset of the population. Even in the age groups where this data will be most beneficial (i.e. young adults) there will be a considerable proportion of the population that will not appear here if they did not attend higher education.

27

Further Education Statistics (FES)

Data Supplier:	Scottish Funding Council (SFC)
Supplier info:	The SFC is a Non-Departmental Public Body of the Scottish Government. The SFC invests around £1.9 billion a year in Scotland's 19 universities and 26 colleges (within 13 college regions) for learning and teaching, skills development, research and innovation, staff, buildings and equipment.
Data type	Unit records
Data Content:	The following variables are included at an individual record level: • Forename(s) • Surname • Sex • Birthdate • Nationality • Religion • Ethnicity • Does the student have a disability • Pre-study domicile • Postcode of permanent home location (pre-study domicile of student) • Student Matriculation Number • Date studies started • Date studies ended • College attended • Mode of attendance
Time period covered	2016/17 and 2017/18 academic years. FES data is returned to SFC via FES online by 01 October of end year, covering all enrolments during the entire academic session e.g. 1 Aug 2017 to 31 July 2018, reported on 01 October 2018
Use of Data:	Production of administrative data based population estimates as statistical research

28

Data Source Information

The SFC collect data about students on Further Education programmes and the students enrolled on them in order to allocate funding and assess the performance of colleges against the outcome agreements.

The FES dataset contains information about the students enrolled on college programmes. Full student FES details are required for all SFC fundable programmes and non-fundable Employability Fund programmes as long as the student has attended at least once. Skills Development Scotland (SDS) administers and manages the Employability Fund on behalf of the Scottish Government. Individuals may appear in this dataset multiple times as a record is submitted for each programme that a person is enrolled on

Data supply and communication

The data provided is done so annually under the terms of a data sharing agreement.

When data is received any queries regarding the data are discussed so that the Admin Data team have a full understanding of the data and if there are any reasons for changes from previous year's data.

Quality Assurance undertaken by data supplier

There are three Management Information System (MIS) software suppliers in the college sector (Capita, Tribal and Civica) and they annually update college management information systems (MIS) to the latest FES guidance published by SFC5. They in turn will mirror many of the code lists within FES in to the college MIS and build in internal validation and error checks prior to files being uploaded to SFCs FES Data Portal.

The student records are submitted by colleges to SFC via the Further Education Statistics (FES) system (the Data Portal). This is an automated and 'live' data capture and record system which encompasses around 300 built-in iterative validation checks to ensure the data is correct and credible. Only when the data has passed will SFC permit the data to be used for analysis. In addition to checks performed by SFC, every college Principal must also sign off the data as a true and accurate record for their college. The SFC analytical team also conducts data quality visits to ensure the student records submitted by colleges are accurate and

---

[5] Guidance Notes for FES can be found at: http://www.sfc.ac.uk/publications-statistics/statistics/statistics-colleges/college-data-collections/college-data-collections.aspx

comparable across the sector. Aggregations of the FES data are then used to produce National Statistics publication 'College Performance Indicators', for:

- 2016/17: http://www.sfc.ac.uk/publications-statistics/statistical-publications/statistical-publications-2018/SFCST022018.aspx
- 2017/18: http://www.sfc.ac.uk/publications-statistics/statistical-publications/2019/SFCST022019.aspx

In producing population estimates, the variables used to link the datasets are of particular importance. Extra information about the validation of these variables, beyond checking they are valid values, from the data suppliers is provided below:

Names – There are no specific validation steps to check that individual names are correct. However any errors will usually be corrected by students throughout their time studying at a college. It is possible that names will differ from official names, e.g. Jim instead of James, however this can be accounted for to some extent in linkage methodology used in the overall project.

Postcodes – A significant proportion of students provide postcode information at application stage where applicants enter the postcode and then choose their address from a list. This will minimise errors in postcodes entered, however generally no proof of postcode is required.

Date of Birth – If a student applies for student funding the date of birth is checked when the funding application is being processed. Otherwise the date of birth provided by the student is taken on trust.

Sex – Colleges receive this information from students. In some cases colleges are finding that it is becoming slightly more common for students to provide different sex (and name) information than what they had recorded at school. However there is no suggestion that this is an error.

Quality Assurance undertaken by the admin data team within NRS

Once the admin data team receive the data, a number of data consistency and validation checks are performed, including:

- Checking the proportion of missing values for variables
- Checking that variables are in the expected formats and values
- Checking the validity of postcodes
- Comparing the data with similar data received in previous years and investigating when there appear to be significant changes.
- Checking the distribution of the day and month elements of dates of birth
- Checking the age distribution of the population.
- Removing duplicate records where identical information is recorded

If these checks suggest the data may need to be amended/adjusted then the potential issues are communicated with the data supplier so the data can be amended as necessary.

Strengths and Limitations of the FES data source

Strengths	Limitations
- Could be useful data source for young adults who are less likely to update their personal information in other data sources. - Validation processes performed by colleges and the SFC, so data is credible. - Students unlikely to be missed as colleges will want to receive the correct funding allocation. - Data feeds into a National Statistics publication. - Contains all the variables used when linking to other datasets.	- There is a lag in being able to receive the data. For example 2016/17 data is only available in early 2018. - Only provides data on a specific subset of the population. Even in the age groups where this data will be most beneficial (i.e. young adults) there will be a considerable proportion of the population that will not appear here. - Postcode information can be from pre-study, so may not match other datasets where a student may have provided a postcode for their term-time address.

Vital Events – Births, Deaths, Marriages and Civil Partnerships

Data Supplier:	National Records of Scotland (Vital Events)
Supplier info:	National Records of Scotland (NRS) is a Non Ministerial Office of the Scottish Government. The purpose of NRS is to collect, preserve and produce information about Scotland's people and history and make it available to inform current and future generations. The Vital Events branch of NRS produces statistics about the births, deaths, marriages and civil partnerships that are registered in Scotland.
Data type (counts or unit records)	Unit records
Data Content:	Birth, death, marriage and civil partnership registration records at individual level. Variables included: **Birth registration data** First name, Last name, Date of Birth, Sex, Address, Postcode, Date of Registration, Father's name, Father's date of birth, Father's address and postcode, Mother's name, Mother's date of birth, Mother's address and postcode. **Death registration data** Deceased's name, deceased's date of birth, deceased's sex, deceased's usual residence address and postcode, deceased's date of death, date of registration. Informant's name, informant's relationship to deceased, informant's address and postcode. **Marriage and Civil Partnership registration data** Date of marriage/civil partnership, date of registration. For each party: Name, Date of Birth, Country of Birth, Country of Residence, Previous Marital status, sex, usual address and postcode.
Time Period Covered	Births: 27 March 2011 to 30 June of reporting period e.g. 27 March 2011 to 30 June 2018 Deaths, Marriages and Civil Partnerships: 1 July to 30 June of reporting period e.g. 01 July 2017 to 30 June 2018

32

Use of Data:	Production of administrative data based population estimates as statistical research

Data Source Information

Every birth, death, marriage and civil partnership that occurs in Scotland must be registered by law.[6,7,8]

For a birth or death to be registered the registrar must be satisfied that the event has occurred. For births, evidence of the event usually takes the form of the informant (usually the mother) providing a card issued by the hospital or midwife who was present at the birth. For deaths this usually takes the form of a Medical Certificate of Cause of Death completed by the medical practitioner who certified the death, this certificate is usually given to the deceased's family. These documents are retained by the registrar upon registration of the events to prevent the birth or death being registered again.

Registrars are asked to take all possible measures to ensure no births or deaths fail to be registered. To do this registrars work with local medical establishments, midwives and funeral directors to identify any missed events. When it becomes known that a birth or death has not been registered in the prescribed time for registering these events, there are processes in place to rectify this.

For marriages and civil partnerships the registration of the event is an essential step in a legal marriage or civil partnership taking place. Therefore it is not possible for these events to occur without being registered. This also removes the risk of these events being registered multiple times.

The data collected is usually input directly to the NRS Forward Electronic Register (FER) computer system as the registrar asks the informant(s) a standard sequence of questions. The computer system will warn the registrar of errors or apparent omissions and warn them of this. The informant(s) and the registrar then read through a printed copy of the record which should pick up any typing errors.

The record is then locked, however corrections can be made if an error is discovered in the future. In every year since 2007, around 97% of records have been created error free, so for individual variables the error rate will be even lower.[9]

---

[6] Registration of Births, Deaths and Marriages (Scotland) Act 1965
[7] Marriage (Scotland) Act 1977
[8] Civil Partnership Act 2004
[9] Page 50 of the Registrar General's Annual Review of Demographic Trends - 2018:
https://www.nrscotland.gov.uk/files/statistics/rgar/2018/rgar18.pdf

There is further scrutiny from NRS examiners who check the information that NRS knows from experience is most likely to contain errors. And corrections are made if necessary. More details of this process are provided on the NRS website: https://www.nrscotland.gov.uk/files//statistics/vital-events/quality-data-obtained-from-registration-of-ve.pdf

Data supply and communication

The data provided is done so annually under the terms of a data sharing agreement and includes record level data for a selection of variables as defined in a data sharing agreement for every registered birth, death, marriage or civil partnership in the previous year. The data is sent by the Vital Events team to the administrative data team via approved NRS data transfer procedures as agreed in a data sharing agreement.

The Administrative Data team have close links with the Vital Events team as they are both in the same organisation and work within the same building. NRS Vital Events have close links with the NRS Registration team, who in turn have close links with registration offices across Scotland. These close working relationships mean that any data quality issues, or planned changes in data collection, are considered in advance and any issues can be considered before the data is used. All parties involved in collecting and processing the data sit within NRS. The Administrative Data team and the Vital Events team all sit within the Statistical Services area of NRS. This means one person has oversight of both areas which further improves the already good links between the teams.

Quality Assurance undertaken by data supplier

At registration, data are provided by the parents, or other qualified people and entered by registrars into the national electronic registration system (FER), where data validation takes place. The system is electronic for the vast majority of offices but there are a few manual offices where data arrives in FER after a couple of days delay. The data from the FER system is passed to the NRS Vital Events statistical database. Here the Vital Events team do further checks on the data. These checks include:

- Looking for any differences in the number of events in the statistical database and the FER. Where there are differences this is investigated to identify
    a) records that are missing from the statistical database and
    b) records that should be deleted from the statistical database.
  These are corrected in the database following the investigation.
- In FER, codes are allocated by the registrar for certain variables such as country of residence. The Vital Events computer system highlights and corrects errors in these codes, and Vital Events staff also aim to identify and correct any anomalies. In addition, quality checks are carried out on records by the Vital Events branch staff supervisor.

More details of this process are provided on the NRS website:
https://www.nrscotland.gov.uk/files//statistics/vital-events/checking-quality-nrs-statistical-data-on-ve.pdf

Quality Assurance undertaken by the Admin Data team within NRS

Once the Admin Data team receive the data, a number of data consistency and validation checks are performed, including:

- Checking the proportion of missing values for variables.
- Checking the validity of postcodes
- Sense checking the number of records compared to previous years and published information.
- Checking that variables are in expected formats and value ranges.

If these checks raise any questions then this is discussed with the Vital Events team to find an explanation or a solution.

Strengths and Limitations of the Vital Events data source

Strengths	Limitations
• Near complete coverage of these vital events occurring in Scotland due to the legal requirement of registration and the steps taken to get full coverage. • Well-defined process for collecting and quality assuring data which will minimise errors. • These datasets are the data source for National Statistics publications published by the National Records of Scotland.	• Events including residents of other countries are included if the event occurs in Scotland. This could lead to additional people being included in the population estimates if not identified. • Events involving residents of Scotland that occur outside of Scotland are not included in the data.

Register of Electors

Data Supplier:	Electoral Registration Officers in Scotland
Supplier info:	The Electoral Registration Officer (ERO) is an official appointed by the local authority to prepare and maintain the Register of Electors.
Data type	Unit records
Data Content:	The following variables are included at an individual record level: • Forename(s) • Surname • Date of Attainment (Date someone turns 18 if they are under 18). • Address and Postcode • UPRN • Elector Number (A unique identifier in the dataset) • Franchise (used to show which list of electors the person is registered on e.g. parliamentary, local government, European parliament. Also indicates where someone is an overseas voter)
Time period covered	Electoral Register as at 1 December 2017 and 1 December 2018
Use of Data:	Production of administrative data based population estimates as statistical research

Data Source Information

The Register of Electors contains details of everyone who has registered to vote in Scotland. It is used to determine who can vote at elections while the Register is in force. A new Register is published at least once a year[10], normally no later than 1st December. Publication of the Register can be delayed to no later than 1 February if there is an election during the annual canvass period. A revised version may be

---

[10] Details of 14 & 15 year olds who are attainers on the local government register in Scotland are not published and are therefore not in the data set provided to NRS

published at other times if, for example, major changes are made to the Register in the course of the year.

Individuals are able to be added to the register at any time and are encouraged to do so throughout the year, with EROs having a legal requirement to invite anyone who is not registered to register to vote. Any non-responses to an Invitation to Register must be followed up with two reminders and a personal visit. There are no personal visits to anyone under the age of 16.

The EROs also have a legal requirement[11] to run an annual canvass where forms are sent out to every household to help identify any changes that need to be made to the Register. There is also a legal requirement to take specified steps to follow up any non-response to the annual canvass, including issuing two reminders and a personal visit.[12] EROs are also pro-active through the year in reviewing any electors they believe are no longer eligible to be registered at an address and removing them from the Register.

By law, a person who is requested for information during the annual canvass must provide the information. In Scotland, there is a criminal penalty of up to £1,000 for failing to provide the requested information, or £5,000 for providing false information.

Another factor that affects the coverage of the data are upcoming elections, as they act as a prompt for people who want to vote to update their details.

There was a UK General Election in June 2017 which would have helped to encourage people to ensure their details are up to date so they would be able to vote at that time. Scottish local elections were also held in May 2017, although turn-out for this was lower than the general election so is likely to have had a smaller impact on quality (66.8% turn-out for the general election compared to 46.9% in the Scottish local elections[13]) . There were no elections in the remainder of 2017 or in 2018, so the public may not have been as prompt in updating their details if they have moved address. This could have a slight impact on the accuracy of the data despite the efforts made to maintain the register.

---

[11] Representation of the People Regulations (Scotland) 2001
[12] Section 9A(2) Representation of the People Act 1983 and Regulation 32ZB 2001 Regulations, Representation of the People Regulations (Scotland) 2001
[13] https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/elections-and-referendums/past-elections-and-referendums

Data supply and communication

The data provided is done so annually under the terms of a data sharing agreement. All data was provided for 2017 and 2018, however due to a clerical error by NRS the password for Fife's encrypted file was unavailable, so Fife was omitted from 2017 quality assurance.

When data is received any queries regarding the data are discussed so that the Admin Data team have a full understanding of the data.

Quality Assurance undertaken by data supplier

For the data covered by this report, the Register was updated monthly between January and September to add new electors and to deal with address changes etc. This procedure was suspended thereafter to allow the annual canvass of households to take place and time for preparation of the new Register. Forms were issued to each household, requesting details of eligible residents. The information obtained during the canvass helped EROs to identify changes that need to be followed up.

The sections below give some detail of checks performed when updating the register to add, amend or remove an individual from the register.

Checks for new applications

When the ERO receives an application from someone to be added to the register there are a variety of checks. Of greatest relevance for the purposes of producing population estimates are the checks on someone's identity and their address.

– Verification of identify - to verify someone's identify the information they provide is compared to DWP records. If the person's identity cannot be verified against DWP records then local data sources may be used instead. If they still cannot be verified then the application enters an exception process where the individual is asked to provide documentary evidence such as a passport or driving licence. If they cannot provide this information then they must get their application attested.
– Residence - among the other requirements to be registered, the ERO must be satisfied that that the individual is resident at the address in the application. If the ERO is not satisfied they can ask for further information and put the application on hold until this is provided.

Amendments to name on existing records

Electors can apply to change their name when already registered. To do so they must provide documentary evidence of the name change. If unable to do so they must provide their date of birth and National Insurance number as part of the application.

Deletions from the register

As well as adding new people to the register, someone who is no longer eligible must be removed to prevent inflation of the register. A person who is registered stays registered unless and until the ERO determines that:

- the person was not entitled to be registered in respect of the address
- the person has ceased to be resident at the address or has otherwise ceased to satisfy the conditions for registration
- the person was registered as the result of an application for registration made by someone else or the person's entry has been altered as the result of an application for a change of name made by someone else.

Examples of when a record is deleted are if the ERO receives a death certificate for an individual or receives notification from two different sources that the elector is no longer eligible.

Records are also deleted when an ERO is notified that someone has made an application to join the Electoral Register in another area, which has been allowed by the ERO in that area, and there is information to indicate that the individual no longer resides at the original address.

Address database

The EROs also have to ensure that their address database is up-to-date, particularly prior to the annual canvass. There is guidance to support EROs in how to do this, however each ERO will have differing procedures depending on the systems they have access to and to handle issues that are particular to their area. Generally the address information comes from the relevant Assessor's Council Tax Valuation List (CTVL) or local authority Corporate Address Gazetteer (CAG) and updated on a regular bases (weekly/monthly).

These updates occur when the CTVL or CAG are updated with properties being added, amended or removed. If the ERO receives information to suggest that an address could be incorrect in some way, it is checked against the Assessor's records or CAG and then amended if necessary.

Published Quality Assurance by other organisations

The Electoral Commission conduct a study which considers the accuracy and completeness[14] of the Electoral Registers[15]. There was not a study of the 2017 Register but in the results for Scotland in 2018 were:

- Parliamentary registers were 84% complete and 87% accurate
- Local government registers were 83% complete and 86% accurate

The findings lead to an estimate of:

- between 630,000 and 890,000 people in Scotland who were eligible to be on the local government registers but were not correctly registered
- between 400,000 and 745,000 inaccurate entries on the local government registers in December 2018

Completeness was lowest for private renters (49%) and those who have only lived at their address for up to one year (32%). Completeness is also lower for younger people with 68% completeness for those aged 18-34 compared to 87% of 35-54 year olds and 92% of those aged 55+.

Quality Assurance undertaken by the admin data team within NRS

Once the admin data team receive the data, a number of data consistency and validation checks are performed, including:

- Checking the proportion of missing values for variables
- Checking that variables are in the expected formats and values
- Checking the validity of postcodes
- Comparing the data with similar data received in previous years and investigating when there appear to be significant changes.
- Checking the distribution of the day and month elements of dates of birth
- Checking the age distribution of the population.
- Removing duplicate records where identical information is recorded

If these checks suggest the data may need to be amended/adjusted then the potential issues are communicated with the data supplier so the data can be amended if appropriate.

---

[14] Accuracy looks at the number of false entries on the electoral registers and completeness measures whether those eligible to be registered are on the registers.

[15] https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-views-and-research/our-research/accuracy-and-completeness-electoral-registers

Following these checks some small amendments were made to improve the data for the purpose of producing experimental population estimates, however these did not require the involvement of the data supplier.

Where possible, Unique Property Reference Numbers (UPRN) were added to each record from the address information if the UPRN had not been provided.

Strengths and Limitations of the Register of Electors data source

Strengths	Limitations
• A large proportion of the adult population in Scotland will be included in the data. The Electoral Commission estimate of completeness in 2018 was 84% for the parliamentary registers and 83% for the local government registers. • Identity is verified when applying to be on the register, minimising false entries. • Data provider has legal requirements to meet regarding how the data is maintained and updated. • The risk of receiving a fine for not providing the information, or providing false information, should improve data quality received from individuals. • The data also captures some information on people who have moved abroad, but are registered as overseas voters. This movement may not have been captured elsewhere. • The Unique Property Reference Number (UPRN) is provided on the Electoral Register for some areas, and in all cases full address information is provided. Meaning 92.9% of records in the 2017 Register are assigned a UPRN. In the 2018 Register this increase to 95.8%. • There were local government and a UK parliamentary election in 2017 which will encourage people to update their details.	• The Registers were published at 1 December while our estimates are mid-year. There will be a mismatch in where some individuals are due to this time difference. • The Register does not include sex for any records, and date of birth can only be derived for a small number of records where someone is yet to turn 18. • Unable to identify where someone is born on 29th February 2000 as there is not a 29th February 2018 for them to turn 18 on. • No coverage on children as they are not eligible to vote. • There are some subsets of the population where there is an increased probability of not appearing on the register. These include young adults, homeless, private renters and those who have not lived at their current address for more than one year. • There were no elections in 2018 to provide an additional incentive for people to update their details. • Not all residents were eligible to register to vote in 2017 & 2018 e.g. residents with a non-EU or non-Commonwealth citizenships or convicted prisoners

5. Risk/Profile matrix for source datasets

This section contains a risk/profile matrix for each data source. The matrix reflects the levels of risk of data quality concerns in using these datasets for this work and the public interest profile of the ABPE. These have been determined by a review undertaken by the NRS Admin Data team using the information contained within the [Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit](#).

For each data source the Public Interest profile has been set to a default value of "medium" for the following reasons:

- One of the objectives of the ABPE is to support future recommendations for the census beyond 2022.
- There is a strong interest in the viability of ABPE to maximise the use of all available data sources to provide accurate and timely evidence to measure our population.
- Currently administrative population estimates are statistical research and are not the official estimate for Scotland's population. Therefore will not be used in calculations to allocate government funds or as the denominator in per capita statistics which would justify a Public Interest score of 'High'.

National Health Service Central Register (NHSCR)

Level of risk of quality concerns	Public interest profile		
	Low	**Medium**	**High**
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit.

Justification for Risk of Quality Concerns score

The risk of quality concerns has been set to "low" for the following reasons:

– There are issues that cannot be avoided due to the nature of the data collection. For example, when people leave Scotland but do not inform their GP they will remain on the NHSCR and recent migrants will not appear on the register until they register with a GP. However as these are known issues they can be considered when using the data.
– The risk of quality concerns is reduced due to the service level agreement to have at least 97% accuracy that is being met.
– This is further reduced as the NHSCR team have a variety of data quality initiatives that are undertaken on a regular basis to mitigate these data quality issues.
– The NHSCR team and the Admin Data team both fall in the Statistical Services division of NRS and both report to the same Director. This means that there is an increased awareness of issues each other may be facing and the impact this may could have on the other party. We can therefore be confident that we will be made aware of any changes that would have an impact on how this data is used.

Health Activity

Level of risk of quality concerns	Public interest profile		
	Low	**Medium**	**High**
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit.

Justification for Matrix Score

The Risk of quality concerns has been set to "Medium" for the following reasons:

– While there are some limitations to the data, knowing where under- and over-coverage needs to be addressed means it can be accounted for when using the data.
– The complex nature of Health Activity Dataset that is dependent on multiple sources with varying levels of internal quality assurance measures.

Scottish Pupil Census (SPC)

Level of risk of quality concerns	Public interest profile		
	Low	**Medium**	**High**
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit.

Justification for Risk of Quality Concerns score

The risk of quality concerns has been set to "low" for the following reasons:

- The data has been judged to be suitable for use in a National Statistics publication.
- There is a clear agreement about what data will be provided, when, how, and by whom. The producers adhere to quality standards and meet the statistical needs for this judgement to be of low risk.

Higher Education Statistics Agency (HESA)

Level of risk of quality concerns	Public interest profile		
	Low	**Medium**	**High**
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit.

Justification for Risk of Quality Concerns score

The risk of quality concerns has been set to "low" for the following reasons:

- There is a well-documented validation process used by HESA to maximise data quality.
- The quality of the variables that are most important to us for the admin mid-year population estimates is likely to be high as students will be motivated to ensure that the provider holds the correct information for them.
- It is unlikely that higher education students are missing from the data as the data providers will benefit from having full coverage of their students as this data is used for funding purposes. Many students will also receive student loans where there is a requirement for them to be registered with their HE provider.

Further Education Statistics (FES)

Level of risk of quality concerns	Public interest profile		
	Low	**Medium**	**High**
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit.

Justification for Risk of Quality Concerns score

The risk of quality concerns has been set to "low" for the following reasons:

- The data is used as art of a National Statistics publication so has already been judged to be of sufficient quality for that.
- There are numerous validation checks performed by both the colleges and the SFC to ensure the data is credible.
- The quality of the name variables are likely to be high as students will be motivated to ensure that the provider holds the correct information for them and there was nothing to indicate an issue with these variables.
- It is unlikely that many students are missing as the data providers benefit from having full coverage given this data is used for funding purposes.
- For a small proportion of the data default dates of birth and postcodes appear to have been used. However there is not a clear way of identifying if this is the case or not. This will make it more difficult to confidently link these records to other datasets increasing the chance of us missing links. However as this dataset is not the primary evidence that they are in Scotland, the quality risk remains low.

Vital Events

Level of risk of quality concerns	Public interest profile		
	Low	**Medium**	**High**
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit.

Justification for Risk of Quality Concerns score

The risk of quality concerns has been set to "low" for the following reasons:

- there is a legal requirement to register these vital events and generally people will want them to be recorded accurately
- there are very robust processes set up for collection and quality assurance of this data
- the data is used as the data source for National Statistics publications so has been judged to be of sufficient quality for those publications.

Register of Electors

Level of risk of quality concerns	Public interest profile		
	Low	**Medium**	**High**
Low	Statistics of low quality concern and low public interest. [A1]	Statistics of low quality concern and medium public interest. [A1/A2]	Statistics of low quality concern and high public interest. [A1/A2]
Medium	Statistics of medium data quality concern and low public interest. [A1/A2]	Statistics of medium quality concern and medium public interest. [A2]	Statistics of medium quality concern and high public interest. [A2/A3]
High	Statistics of high data quality concern and low public interest. [A1/A2/A3]	Statistics of high quality concern and medium public interest. [A3]	Statistics of high quality concern and high public interest. [A3]

*A1/A2/A3 – definitions supplied Office for Statistics Regulation's Administrative Data Quality Assurance Toolkit.

Justification for Risk of Quality Concerns score

The risk of quality concerns has been set to "medium" for the following reasons:

- There are well defined procedures for verifying the identity of individuals on the register. Due to this, along with the potential legal ramifications of providing false information, the vast majority of records can be expected to be correct.
- The annual canvass, along with procedures for removing records, should minimise inflation of the register.
- While children are not included, other data sources can be used to identify these.
- There are subsets of adult population that appear to be less likely to appear in the Electoral Register but as this information is being combined with other information it provided a very good indication of recent address.

6. Background notes

Background

This document supports the Statistical Research publication [Administrative Data Based Population Estimates, Scotland 2016-2018](#).

Methodology

[The Administrative Data Based Population Estimates v2, Scotland 2016–2018: Methodology Report](#) provides more detail on the methodology, as well as information on the quality of the data and known uses of the data.

Future developments

We intend to continue developing the methodology for producing administrative data based population estimates based on the learnings from producing these estimates.

Following this publication, NRS 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 demography at http://www.gov.scot/scotstat .

7. 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.

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

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