Administrative Data Based Population, Household and Ethnicity Estimates, Scotland (2019-2022)

Quality Assurance of Administrative Datasets

Published on 27 February 2025

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

Contents

1. Disclaimer

The Administrative Data Based Population, Household and Ethnicity Estimates are statistical research outputs. These estimates should not be considered as a replacement for Scotland's official statistics: Population, migration and households estimates for Scotland.

2. Introduction

This document discusses the quality of the administrative datasets used to create the administrative data based estimates. It provides information on each dataset and how they were quality assured prior to linkage to ensure they were suitable for this project.

We used the Administrative Data Quality Assurance Toolkit which provides a frame work for the Quality Assurance of Administrative Datasets (QAAD), to guide our quality assurance process. This toolkit includes a risk matrix to assess the public interest profile (lower, medium, higher) and the level of risk of data quality concerns (lower, medium, higher). We have completed the risk matrix for each dataset and have assessed the public interest as medium because, while these are not official statistics, there is a significant interest in the viability of Administrative Data-Based Estimates to provide accurate and timely population statistics.

This document supports our compliance with the UK Statistics Authority and the Office for Statistics Regulation's Code of Practice for Statistics. In it 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.

Although the publication this report accompanies are the 2016 – 2022 estimates, This document covers the QAAD for years 2019, 2020,2021 and 2022. The QAAD for previous years can be found below:

Quality Assurance of Administrative Dataset (QAAD) 2016

Quality Assurance of Administrative Dataset (QAAD) 2017 and 2018

3. Overall Quality of the Administrative Datasets

The administrative data based estimates consist of the population, household and ethnicity estimates. These estimates are produced by linking data from a range of administrative sources. This linked data is cut down using inclusion rules and is called Scotland’s Integrated Demographic Dataset (SIDD). The published report provides full details of the SIDD and how each estimates (population, household and ethnicity) are created from it.

The quality of the individual datasets used to create these estimates have been assessed, and are of high quality as they undergo quality assurance by data providers before receiving it. The administrative data team also carries out additional quality assurance checks on the individual datasets. While some minor issues were identified during this process, they had minimal impact on dataset linkage and estimates. These issues and their potential effects are outlined below.

Known data issues

This section contains key data issues that might have an impact on the estimates.

Year	Issue	Implications on estimates
Across all Years	In the National Health Service Central Register (NHSCR) and Health Activity (HA) datasets, a default DOB of 1 January has been used for individuals who did not know or have no evidence of their DOB. This has slightly inflated births on that date by 0.1%.	While this affects the date of birth distribution, the impact is minimal. 1 January appears approximately 0.37% in NHSCR and 0.34% in HA compared to an expected average of 0.27% for other dates. This is expected to have little effect on the population estimates, at most affecting the estimates by 0.1% which is small compared to other sources of uncertainty. Additionally most individuals are expected to link correctly and may just have a slight effect on age distribution shifting some individuals recorded age by up to one year.
2021 and 2022	In the Health activity dataset, there were a small number of cases where date of birth were	This is a small amount of records and also shouldn’t have any effect on the

	missing (52 in the 2021 secondary care dataset and 63 in the 2022 secondary care dataset). This led to the allocation of a default date of 14 Oct 1582 by the R programming language.	estimates because the year of birth from NHSCR data source is prioritised.

4. Source Dataset Information

National Health Service Central Register (NHSCR)

Data Supplier:	National Records of Scotland (NRS)
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	Individual level data
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	Data extract at 30 June 2019, 2020, 2021 and 2022 and on 20 March 2022
Use of Data:	Production of administrative data based estimates.

Data Source Information

The NHSCR is an electronic demographic database for:

- everyone 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 in Scotland (and from across the UK if notified)
- GP Registration (within Scotland) – ‘migration’ into and within Scotland
- GP Registration (within the rest of the UK) – ‘migration’ out of Scotland.

Data supply and communication

The data are provided under the terms of a data sharing agreement and include record level data for a selection of variables as defined in a data sharing agreement for every person on the NHSCR.

The data are 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, with 5% of the manual updates checked daily. These record updates 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 and NHS National Services Scotland which requires updates to have 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 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 data was received, 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.
- 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.

Known Issues

- A default DOB of 1 January has been used for individuals who did not know or have no evidence of their DOB. This has slightly inflated births on that date by 0.1%

Strengths of the data source

- NHSCR is a comprehensive source of record level data that covers the vast majority of the population in Scotland.
- Captures hard to reach population The dataset helps capture internal migration because individuals will typically have to register with a new GP when they relocate.

Limitations of the data source

- 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.
- There might be a lag with patients registering with a new GP. As a result some people will be recorded in the wrong area. Particularly an issue among younger adult males[1].
- 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 could be a delay with new born baby being registered which will delay them appearing on the register.

Risk Matrix

This section contains a risk/profile matrix for the NHSCR. The matrix reflects the levels of risk of data quality concerns and the public interest profile of the statistics. 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 the use of data for this project, the cell highlighted is appropriate:

Level of risk of quality concerns	Public interest profile		
	Low	Medium	High

[1] 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

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 public interest profile has been set to “medium” for the following reasons:

- One of the objectives of this project is to support future recommendations for the census beyond 2022. Therefore, there is a strong interest in evaluating the viability of the estimates to maximise the use of available data sources to provide accurate and timely evidence to measure Scotland’s population.
- Currently administrative data based estimates are experimental statistics 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’.

The risk of quality concerns has been set to “Low” for the following reasons:

- 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 census teams both fall in the Statistical Services division of NRS. 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.

- 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

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. PHS provided the healthcare data used in the production of the administrative data based estimates. Primary care data were provided separately from secondary care data. Each wave of data comprised a confidential data file and a payload data file.
Data type	Individual level data
Data content:	The following variables were included in the confidential files at an individual record level for both the primary care and secondary care data sets: • 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 The payload files for both primary care and secondary care contained the following data variables: • Marital Status • Ethnic Group • Date of last interaction (Day, Month and Year).

	• Transfers Out Status. Identifies patients who are no longer resident within Scotland on the final date of the extract. • Random identifier The key file contained the Random Identifier with its correspondent Row ID to permit matching of the confidential and payload data. The random identifier has a one-to-one mapping with CHI number and is consistent for an individual over time.
Time Period Covered	Data extract at 30 June 2019, 2020, 2021 and 2022 and on 20 March 2022 with the ‘Last Interaction’ variable indicating the date of an individual’s last interaction with selected NHS services over the previous 3 years
Use of Data:	Production of administrative data based estimates.
Data Supplier:	Public Health Scotland (PHS)

Data Source Information

The Community Health Index (CHI) is a register of all patients in NHS Scotland where each patient is assigned a unique 10-digit CHI number. The CHI register exists to ensure that patients can be correctly identified, and that all relevant 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. For the purposes of this project individual CHI numbers were assigned a random identifier, with a one to one mapping between CHI number and random identifier. This allowed data for individuals to be linked without sharing the CHI number itself.

Extracts called the secondary care and primary care datasets (collectively referred to as the Health Activity Datasets) were created for this project by PHS. It is important to note that no individual’s health data was supplied within these datasets, only an activity flag indicating the last date the individual interacted with a NHS service (date of ’Last Interaction’). This variable reports the date of an individual’s last engagement with selected NHS services, providing up-to-date information that can help confirm the administrative based population estimates.

Data supply and communication

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

The health activity data was provided in separate files for primary care and secondary care:

Primary care data covered interactions with Dental Services, Pharmacies and Prescribing, Bowel Screening, and Abdominal Aortic Aneurysm (AAA) screening.

Secondary care data covered interactions with Hospitals, including Outpatients (SMR00), Inpatients and Day cases (SMR01), Maternity (SMR02), Mental Health (SMR04), Cancer Registrations (SMR06), and Accident and Emergency.

Quality Assurance undertaken by data supplier

PHS performed internal quality assurance before sharing the data with NRS. In addition, data quality assurance is routinely undertaken on the source datasets as outlined below.

Secondary care data

Five 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 and SMR06 Cancer Registrations

SMR00, SMR01, SMR02 and SMR04 - Validation

Validation is either carried out locally prior to submission to PHS or centrally once received by PHS. These validation checks may generate:

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

Examples of validation carried out include:

- simple checking - for example, checking that the submitted postcode exists on the UK national postcode directory,
- cross-checking – for example, checking that the listed consultant worked in the provider, location and specialty at the time of admission,
- carrying out additional calculations prior to checking – for example, checking that the patient’s age at admission is consistent with the diagnosis.

Any data errors (missing or invalid information) or queries (information that appears incorrect) are sent back to the NHS board for further investigation.

Further details on SMR data submissions and SMR validation can be found on the PHS website.

In addition to validation checks, Public Health Scotland includes checks on SMR data completeness and timeliness:

SMR00, SMR01, SMR02 and SMR04 - 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. PHS publishes data on SMR data completeness here.

In the period covered by these analyses, SMR completeness was between 97 to 100 per cent nationally across Scotland. The majority of Health Boards had completion rates of around average, although there were some outliers. For example, NHS Dumfries and Galloway's completion rate for new outpatients (SMR00 New) was 89 per cent and NHS Orkney's completion rate for SMR02 Deliveries was 54 per cent. Return rates at NHS Orkney for maternity deliveries (SMR02) were notably low, as low as 50% in the 2022 calendar year.

SMR00, SMR01, SMR02 and SMR04 - 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. PHS publishes data on SMR data completeness here.

SMR06

This dataset includes information on all new diagnoses of cancer occurring within Scotland. Unlike the SMRs listed above, SMR06 is an ongoing register recording information on treatments and treatment dates. Source records are provided as provisional records, with confirmed registrations created by Public Health Scotland following internal quality assurance checks. SMR06 data is collected annually on 'completion' of registrations for the year,12-16months after the year end of the incidence years.

Primary care data

Dentistry

Dental data are collected through electronic payment claim submissions submitted by dentists to claim payments for dental services provided[2]. PHS only provide treatment contacts in the more recent data extracts; prior to 2019, where no

[2] Electronic payment submissions replaced GP17 forms for general dentistry claims in October 2018 and orthodontic claims in January 2020.

treatment claims had been made for a patient, registration data was provided as a proxy for contact.

COVID-19 measures impacted on dental services when dental treatment was restricted to emergency care only. General Dental Service (GDS) activity reduced during the first national lockdown and although not yet fully recovered, some parts of this service have returned to pre-pandemic levels. It was only from 1 April 2022 that dentists were allowed to de-escalate their infection prevention and control measures in line with national guidance to alleviate system pressures and allow an increase in patient throughput. See ‘The impact of COVID-19 on NHS dental services and oral health in Scotland: Annual Report’ for fuller details on the impact of COVID measures. Although this report states “Users should therefore be aware of the aspects of data quality and caveats surrounding these data, all of which are listed in this document”, there is no explicit reference to quality assurance measures within the annual report.

Community Pharmacist and Dispensing Contractors

Information on 100% of NHS Scotland prescriptions dispensed within the community and claimed for payment by a pharmacy contractor (i.e., pharmacy, dispensing doctor or appliance supplier) is held on the Prescribing Information System. Routine monthly checks are carried out by Practitioner Services on a random sample of approximately 5% of prescription payments. These check all data captured for payment and the payment calculation accuracy with a target accuracy of 98%. Outputs and metadata for contractor activity is published on the PHS website, but results of the random checks cited above are not retained or published. Those checks are primarily for internal confirmation to make sure that the correct data has been used.

Quality Assurance undertaken by the Admin Data team within NRS

Once the NRS Administrative Data team receive the data from PHS, a number of data consistency and validation checks are performed on Health Activity datasets 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), Unique Property Reference Number (UPRN) derived from address variable and postcodes.
- Sense checking the number of records by single year of age, reviewing distribution of date of birth and age of persons on each dataset.
- Checking the completeness of payload variables (Ethnic Group, Marital Status.

These checks provide additional information to NRS team when linking data to produce estimates.

Known issues

- There were a small number of cases where date of birth was missing (52 in the 2021 secondary care dataset and 63 in the 2022 secondary care dataset).
- There is a standing convention that the first of January is used as the default day or month of birth where patients registering for the first time have no known date of birth or culturally have no day or month of birth. This can potentially inflate births on that date.

Strengths of the data source

- There were a small number of cases where date of birth was missing (52 in the 2021 secondary care dataset and 63 in the 2022 secondary care dataset).
- There is a standing convention that the first of January is used as the default. day or month of birth where patients registering for the first time have no known date of birth or culturally have no day or month of birth. This can potentially inflate births on that date.

Limitations of the data source

- Moves within Scotland cannot be picked up until the patient registers with a new GP. As a result, some people will not be recorded in their current area of residence. Particularly an issue among younger adult males[3].
- No data linkage process is 100% accurate, therefore, due to the number of records being used to create this dataset, there may be a small percentage that are not linked correctly.
- The secondary care dataset is generally more complete than the primary care dataset for both identifiable and payload variables.

Risk/Profile Matrix

This section contains a risk/profile matrix for the Health Activity data. The matrix reflects the levels of risk of data quality concerns and the public interest profile of the statistics. These have been determined by a review undertaken by the NRS Admin Data team using the information contained within the Office for Statistics

[3] 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

Regulation's Administrative Data Quality Assurance Toolkit. For the use of data for this project, the cell highlighted is appropriate:

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 Public interest profile has been set to “Medium” for the following reasons:

- One of the objectives of this project is to support future recommendations for the census beyond 2022. Therefore, there is a strong interest in evaluating the viability of the estimates to maximise the use of available data sources to provide accurate and timely evidence to measure Scotland’s population.
- Currently administrative data based estimates are experimental statistics 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’.

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

- While there are some limitations to the data, knowing them means it can be accounted for when using the data.
- The complex nature of Health Activity Dataset being dependent on multiple sources means it has varying levels of internal quality assurance measures.

Scottish Pupil Census (SPC)

Data Supplier:	Scottish Government: Education Analytical Services (EAS)
Supplier info:	Education Analytical Services (EAS) is part of the Scottish Government's Learning Directorate. EAS collects and produces statistics on education and other children's services in Scotland. The Pupil Census is a collection of data on publicly funded schools and their pupils. The data gathered in the pupil census is drawn from management information held by schools and local authorities for the purposes of administering education. The information published is therefore a reflection of the information provided by school staff and pupils' parents/guardians.
Data type	Individual level data
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:	2019/20, 2020/21, 2021/22 and 2022/23 School Pupil Censuses
Use of Data:	Production of administrative data based estimates.

Background Information

The Pupil Census refers to the pupil population as at mid-September each year, with numbers of pupils by age as at the end of the following February. Data are collected from all Local Authority and Grant aided schools and school centres. Pupil Census data are published in Schools in Scotland and associated supplementary statistical tables. Schools in Scotland is an Accredited Official Statistics publication; the publication has been assessed by the UK Statistics Authority.

The Pupil Census data are usually published the March following collection, i.e. six months after September collection. The data are used for statistical analysis and to support evidence-based policy making. The data are collected electronically and largely sourced from school management information systems. Further details are available here: Scottish Exchange of Data: school-pupil census - gov.scot (www.gov.scot).

In terms of accuracy, the Pupil Census collection is drawn from management information held by schools and local authorities for the purposes of administering education. The information published is therefore a reflection of the information provided by school staff and pupils' parents/guardians.

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 of the data source

- 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 an Accredited Official Statistics publication.

Limitations of the data source

- 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.
- No information on independent sector, home schooling etc. as out of the scope of this data collection.

Risk/Profile Matrix

This section contains a risk/profile matrix for the SPC dataset. The matrix reflects the levels of risk of data quality concerns and the public interest profile of the statistics. 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 the use of data for this project, the cell highlighted is appropriate:

	Public interest profile

Level of risk of quality concerns	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 public interest profile has been set to “medium” for the following reasons:

- One of the objectives of this project is to support future recommendations for the census beyond 2022. Therefore, there is a strong interest in evaluating the viability of the estimates to maximise the use of available data sources to provide accurate and timely evidence to measure Scotland’s population.
- Currently administrative data based estimates are experimental statistics 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’.

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 an Accredited Official 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.

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

Higher Education Statistics Agency (HESA)

Data Supplier:	Higher Education Statistics Agency (HESA), part of Jisc.
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	Individual level data
Data Content:	The following variables are included at an individual record level: • First name • Middle name • Last name • Previous surname • Sex ID • Birthdate • Person ID • Postcode (term time) • Postcode (home) • Country of domicile • Mode of study • End date • Length of course • Year of course
Time period covered	Data covers 2019/20, 2020/21 and 2021/22 academic year
Use of Data:	Production of administrative data based estimates

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)

The aggregated figures from these data are used by HESA in their annual accredited official statistics publication ‘Higher Education Student Statistics: UK’, links to the published data are here:

Higher Education Student Statistics: UK, 2019/20

Higher Education Student Statistics: UK, 2020/21

Higher Education Student Statistics: UK, 2021/22

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

Data supply and communication

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

The data is provided to SG by HESA under the terms of a data sharing agreement. The data include 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:

HESA Collections | HESA Student 2019/20: Support guides | HESA

HESA Collections | HESA Student 2020/21: Support guides | HESA

HESA Collections | HESA Student 2021/22: Support guides | HESA

Quality Assurance undertaken by data supplier

HESA produce a student record quality report[4] 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 rules[5] include, but are not limited to:

- checking unique identifiers are valid by using a checksum
- 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

A number of data consistency and validation checks are performed, including:

- Checking the proportion of missing values for requested variables
- Checking that variables are in the expected formats and values
- Checking the validity of postcodes

[4] https://www.hesa.ac.uk/about/regulation/official-statistics/quality-report

[5] 20/21: Quality Rules Directory for C20051 | HESA

- 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.

Strengths of the source

- The estimated student population of an area includes all those usually resident there, regardless of nationality.
- Extensive validation process by HESA make the data as complete as possible
- Captures hard to reach population age group that might not appear in other Admin Data sources (Young Adults).

Limitations of the source

- The data for an academic year is usually available from the following winter.
- 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.

Risk Matrix

This section contains a risk/profile matrix for the HESA dataset. The matrix reflects the levels of risk of data quality concerns and the public interest profile of the statistics. 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 the use of data for this project, the cell highlighted is appropriate:

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]

Justification for Matrix Score

The public interest profile has been set to “medium” for the following reasons:

- One of the objectives of this project is to support future recommendations for the census beyond 2022. Therefore, there is a strong interest in evaluating the viability of the estimates to maximise the use of available data sources to provide accurate and timely evidence to measure Scotland’s population.
- Currently administrative data based estimates are experimental statistics 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’.

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.
- 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)

Data Supplier:	Scottish Funding Council (SFC)
Supplier info:	The SFC is a Non Departmental Public Body of the Scottish Government. The SFC invests around £2 billion a year in Scotland's 19 universities and 24 colleges (within 13 college regions) for learning and teaching, skills development, research and innovation, staff, buildings and equipment.
Data type	Individual level data
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 • College code number • Mode of attendance
Time period covered	2019/20, 2020/21 and 2021/22 academic year
Use of Data:	Production of administrative data based estimates.

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 student's 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 years data.

Quality Assurance undertaken by data supplier

There are four Management Information System software suppliers in the college sector (ESS, Tribal, Civica, and One Advanced) and they annually update college Management Information Systems (MIS) to the latest Further Education Statistical (FES) guidance published by SFC[6]. 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 and other desk-based exercises to ensure the student records submitted by colleges are accurate and comparable across the sector. Aggregations of the FES data are then used to produce Accredited Official Statistics publication 'College Performance Indicators'.

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 to check that individual names are correct. However any errors will be usually be corrected by students throughout their time

[6] Guidance Notes for FES 2 2021/2022 can be found at: http://www.sfc.ac.uk/publications-statistics/statistics/statistics-colleges/college-data-collections/college-data-collections.aspx

studying at a college. However it is possible that names will differ from official names, i.e. 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 if appropriate.

Strengths of the source

- Could be useful data source for young adults who are not as likely to update other their 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 Accredited Official Statistics publication.

Limitations of the data source

- 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 is from pre-study, so may not match other datasets where a student may have provided a postcode for their term-time address.
- Also some postcodes particularly for school pupils at primary will be the school.

Risk/Profile Matrix

This section contains a risk/profile matrix for FES data. The matrix reflects the levels of risk of data quality concerns and the public interest profile of the statistics. 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 the use of data for this project, the cell highlighted is appropriate:

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 public interest profile has been set to “medium” for the following reasons:

- One of the objectives of this project is to support future recommendations for the census beyond 2022. Therefore, there is a strong interest in evaluating the viability of the estimates to maximise the use of available data sources to provide accurate and timely evidence to measure Scotland's population.
- Currently administrative data based estimates are experimental statistics 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'.

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

- 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 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.
- 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 likely to provide extra evidence of someone's existence rather than being the primary evidence that they are in Scotland, the quality risk remains low.

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	Individual level data
Data content:	Birth, death, marriage and civil partnership registration records at individual level. The variables included in the dataset are: **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. **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. **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.
Time period covered	Births and deaths: 27 March 2011 to 30 June 2022

	Marriages and Civil Partnerships: 1 July 2014 to 30 June 2022
Use of Data:	Production of administrative data based estimates.

Data Source Information

Vital events data records births, deaths, civil partnerships and marriages in Scotland. It is a legal requirement to register these events.

For births and deaths, there is a process to show that the event has occurred. For example for deaths a Medical Certificate of Cause of Death (MCCD) is usually provided and this is preserved by the registrar to prevent multiple registrations of the same death.

A marriage or civil partnership cannot be legally recognised without registration. The registration of these events is a step in the ceremony and this means they cannot be registered multiple times.

The data are input by the local registrar into the NRS Forward Electronic Register (FER). A standard set of questions are asked to the individual(s) registering the event, the system flags any potential errors and the registrar then reviews a printed copy of the registration with the individual(s). The record is then locked, however corrections can be made if an error is discovered in the future.

Data supply and communication

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.

Quality Assurance undertaken by data supplier

The data from the electronic NRS system was passed to the Vital Events team to check. These checks included:

- Numerical or administration registration errors in the statistical database compared to the electronic NRS FER system. These errors are rectified through a thorough investigation and usually include records that are missing from the statistical database or those records that should potentially be deleted.
- FER uses a coding system to identify a variety of different variables and the computer system will highlight and help correct any errors in these codes. Further quality checks are also carried out by Vital Events coding staff.

Details of this process can be found following this link:

Quality of NRS Data on Vital Events

Quality Assurance undertaken by the Admin Data team within NRS

A number of data consistency and validation checks are performed, which included:

- Checking the proportion of missing values for requested 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 suggests the data needs to be amended/adjusted the issues are communicated with the data supplier so the data can be amended when appropriate.

Known Issues

Registration offices were closed to the registration of births between March and late June 2020. This might have postponed registrations but should have been reconciled so the impact on statistics will be minimal.

Strengths of the data source

- Legal requirement of these events to be registered means complete coverage across Scotland.
- Efficient and effective QA process throughout the collection and analysis of the data means errors are picked up on quickly and are usually minor.
- NRS uses these data in NRS official publications.
- The Vital Events dataset can show relationships between individuals which may not be found in other datasets.
- The dataset has a good coverage of address/geographical information which can be used to obtain UPRNs (Unique Property Reference Numbers).

Limitations of the data source

- Events that include Scottish nationals out with Scotland are not included in the data collection.
- Citizens of other countries events are included if the event happens in Scotland.
- There could be a time lag following the events and it being registered (E.g. births can be registered up to 21 days from the date of occurrence).

Risk Matrix

This section contains a risk/profile matrix for the Vital Events data. The matrix reflects the levels of risk of data quality concerns and the public interest profile of the statistics. 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 the use of data for this project, the cell highlighted is appropriate:

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 Public Interest profile has been set to value of “Medium” for the following reasons:

- One of the objectives of this project is to support future recommendations for the census beyond 2022. Therefore, there is a strong interest in evaluating the viability of the estimates to maximise the use of available data sources to provide accurate and timely evidence to measure Scotland’s population.

- Currently administrative data based estimates are experimental statistics 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’.

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 Accredited Official Statistics publications so has been judged to be of sufficient quality for those publications.

Electoral Register

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 Electoral register.
Data type	Individual level data
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. Also indicates where someone is an overseas voter)
Time Period Covered	December 2019, 2020, 2021 and 2022
Use of Data	Production of administrative data based estimates.

Data source information

The Electoral register 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[7], 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 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 Electoral Registration Officers (EROs) having a legal requirement[8] to invite anyone who is not registered to register to vote. Any non-

[7] 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

[8] Representation of the People Regulations (Scotland) 2001

responses to an Invitation to Register must be followed up with two reminders and a personal visit, although there are no personal visits to anyone under the age of 16. All EROs match their records to the Department of Work and Pensions (DWP) database and most also match to a local database e.g. Council Tax as well. If all electors match in a household, then that household just gets a single communication setting out who is registered to vote and asking them to advise the ERO of any changes. There is no follow up activity required in these cases. Where not all the electors match, or if the ERO is aware of a change at the property, then those households get a full communication and the ERO is required to carry out follow up activity to encourage a response.

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 by an ERO 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.

Data supply and communication

The data provided is done so annually plus ad hoc monthly requests 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.

Quality Assurance undertaken by data supplier

For the data covered by this report, the Register is published on 1st December with monthly updates to it January to November, adding new electors and to deal with address changes etc.. Forms were issued where the ERO believed there may be changes to the household. The information obtained through those requests 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. Most relevant 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 then 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 receiving notification from two different sources that the elector is no longer eligible.

Where an ERO believes that a person is no longer registered at a property but has not received the necessary documentary evidence they will carry out a review of the person's registration. This will involve writing to the person at the address at which they are registered advising that unless they confirm within 14 days of the review being issued that they are still resident at the property they will be removed from the register.

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.

Quality Assurance undertaken by the census teams within NRS

The admin data team receive the data in separate files according to electoral wards within each ER area.

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 suggested the data needed to be amended/adjusted then the potential issues were communicated with the data supplier so the data could be amended if appropriate.

Strengths of the data source

- A large proportion of the adult population in Scotland will be included in the data.

- 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 most areas.

Limitations of the data source

- The Registers are published 1st December while our population estimates are based on mid-year. This time difference may lead to a mismatch in the recorded location of individuals since their registered location could have changed.
- 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.
- 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.

Risk/Profile Matrix

This section contains a risk/profile matrix for Electoral Register data. The matrix reflects the levels of risk of data quality concerns and the public interest profile of the statistics. 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 the use of data for this project, the cell highlighted is appropriate:

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 **public interest profile** has been set to “medium” for the following reasons:

- One of the objectives of this project is to support future recommendations for the census beyond 2022. Therefore, there is a strong interest in evaluating the viability of the estimates to maximise the use of available data sources to provide accurate and timely evidence to measure Scotland’s population.
- Currently administrative data based estimates are experimental statistics 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’.

The **risk of quality concerns** has been set to “low” 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.

5. 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 upcoming publications on our website. If you would like us to tell you about future statistical publications.

Enquiries and suggestions

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

Lead Statistician: David Rowley

E-mail: statisticscustomerservices@nrscotland.gov.uk

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