Scotland's Census
Shaping our future
A' dealbhadh ar n-àm ri teachd

Scotland's Census 2022

Quality Assurance of

Administrative Datasets (QAADs)

Version 2 – October 2023

Contents

1. Introduction

Scotland's Census 2022 is the official count of every person and household in the country and their characteristics. The data collected are processed carefully to ensure they accurately reflect the makeup of Scotland's population. Prior to release of the results of the census to the public, statistical quality assurance was carried out to ensure the quality of census data. This included using administrative data sources to compare to the data from the census.

This document summarises the quality of each of the administrative data sources identified for use in the quality assurance of Census 2022 data. It provides details on the nature of each dataset, how the data supplier quality assures the data, how the data are supplied to the data supplier and the scope of the quality assurance that was carried out using these datasets.

The majority of the data sources referred to In this document were used during the Validation of Population Estimates (VoPE) process, which ensured that the national and sub-national census estimates are plausible when compared with data from survey, administrative, and demographic comparator sources. VoPE occurred at the end of the census data journey but prior to statistical disclosure control and outputs. More information on the VoPE methodology can be found here, including information on which data sources are used to validate which census questions.

Some of the datasets were also used in production of a linked administrative dataset. This was used to supplement the Census Coverage Survey (CCS) records in estimation, assist placement of records in adjustment, and for quality assurance of communal establishment population estimates during data processing. Inclusion of administrative data in these processes helped to reduce uncertainty in the estimates of the total population.

Our comparator data sources are a mixture of administrative and survey datasets as well as datasets that are themselves based on both administrative and survey data (i.e., the Mid-Year Population Estimates). We have made use of the Administrative Data Quality Assurance Toolkit to guide how we quality assure each dataset, even though not all our datasets are administrative datasets and our aim is not to produce new statistics using these data. Nevertheless, the toolkit and Quality Assurance of Administrative Datasets (QAAD) framework is a helpful way to assess each dataset and document its strength and limitations.

The toolkit provides a risk matrix that is usually used to assess the public interest profile of each administrative dataset (lower, medium, higher) and the level of risk of data quality concerns (lower, medium, higher). We have completed the risk matrix for each dataset but we have set the public interest profile as high reasoning that we are using each dataset to validate a component of the census and the census outputs have high public interest due to their comprehensive nature and crucial roles in national and local policy- and decision-making across a range of topics.

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.

2. QAADs

2.1 Mid-year Population Estimates

Data Supplier	National Records of Scotland (NRS)
Supplier information	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 Mid-year Population Estimates (MYE) publication produces statistical estimates of the population of Scotland, council areas and NHS board areas.
Data type (counts or unit records)	Counts
Data Content:	This publication covers a variety of different areas of the population of Scotland including the overall outlook of the population, the age structure and area population within Scotland. **Overall population** gives a high-level overview of the size of the population and reasons for its changes from the previous publication. **Age structure** focuses on the age composition of the population. **Area population** provides analysis on the changes in each council area in Scotland and which areas exhibit increases and decreases in the overall population and why this is happening through components such as natural change or net migration.
Time period covered:	The Population and Migration Statistics team produced a bespoke population estimate for census day (20th March 2022) available to us in autumn 2022.
Use of Data:	Validation of population estimates.

Source information

The mid-year population estimates refer to the population as at 30 June each year. The estimates are usually published 10 months after this date. These statistics are used in conjunction with other data for the allocation of government funds to Scotland and its Council and NHS Board areas. They are also a main component of many other statistics such as population projections and are used as the denominator for per capita statistics.

The MYE publication therefore uses a number of different organisations to gather the necessary data. These data consist of a mixture of administrative and survey data. The Office for National Statistics (ONS), Northern Ireland Statistics and Research Agency (NISRA), Labour Force Survey (LFS), the Home Office, Ministry of Defence: Defence Statistics, Civil Registration System, National Health Service Central Register (NHSCR) and the Justice Analytical Services of the Scottish Government are all agencies that contribute to the MYE including birth and death numbers, migration statistics and prisoner data.

In terms of accuracy the MYE are based on the 2011 Census and updated annually by 'ageing on' the population and applying information on births, deaths and migration. Births and deaths data are used from the civil registration system which is considered to be virtually complete. Migration is more difficult to estimate because there is no comprehensive registration system in the UK. Migration is usually derived from the National Health Service Central Register (NHSCR) and Community Health Index (CHI) for migration within the UK, and from the International Passenger Survey (IPS) and Home Office data for international migration. Due to the COVID-19 pandemic, the IPS was suspended in March 2020. ONS led research into alternative data sources including using statistical modelling to estimate UK-international migration over this period.[1] For the period March to June 2020, migration flows were estimated based on the Scottish proportion of the UK modelled migration. This, in turn, was based on the Scottish proportion of the IPS flows from July 2019 to March 2020. In addition, it is expected that during the COVID-19 pandemic people may have moved address without registering this change with their GP, however NRS did not adjust its established methods of estimating within Scotland and Scotland-UK migration.

Data supply and communication

The data are provided annually by a variety of different stakeholders. The population statistics team in NRS have forged close links with ONS, NISRA, Scottish Government and NHSCR. These ongoing relationships with different organisations mean that any quality assurance issues or changes can be resolved with minimal delay.

Quality Assurance undertaken by data supplier

When the Population and Migration Statistics team within NRS receive data for the various components, checks are carried out and comparisons made with previous years' data to gauge consistency and completeness of coverage. The data are then

[1] Using statistical modelling to estimate UK international migration - Office for National Statistics (ons.gov.uk)

processed electronically to produce the mid-year estimates. Quality assurance takes place at each stage of this process. This may include calculation of demographic rates and ratios to help check the plausibility of the data. NRS also have processes in place to check the suitability of the administrative sources used in producing population estimates.

The Population and Migration Statistics team also meet with the data suppliers and discuss any issues that might affect the data. These close working relationships ensure that issues and planned changes to data collection can be considered before the data are used to produce population estimates.

A page is available on the NRS website providing information on the quality assurance arrangements for administrative data used in population estimates, along with information on the suitability of each data source used in the production of population estimates.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.
- Checking the age distribution of the population.

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 population of an area includes all those usually resident there, regardless of nationality.
- Population estimates for Scotland are comparable with population estimates from the UK and most countries around the world.

Limitations of the source

- Short-term international migrants are excluded.
- Those aged 90+ are reported as a single age group rather than reported by single year of age.

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low", as there is a well-documented quality assurance process to ensure MYEs are achieving high data quality standards. The MYEs are a National Statistics publication, so they have already been judged of sufficient quality for that accreditation. The statistics are based on the best available sources, including high quality administrative data sources for births, deaths, migration and special populations.

2.2 Centenarians Population Estimates

Data Supplier	National Records of Scotland (NRS)
Supplier information	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 Centenarians publication produces statistical estimates of the population of Scotland aged 90 or above.
Data type (counts or unit records)	Counts
Data Content:	This publication covers a variety of different areas of the population of Scotland including the overall outlook of the 90 or over population, the sex structure and the causes of changes in centenarian populations. **Overall** gives a high-level overview of the size of the population and reasons for its changes such as growth and proportion of this part of the population **Sex structure** focuses on the sex composition of the population to assess the marked differences in male and female numbers among centenarians. **Changes in the population** provides analysis on observed changes in the population and the reasons behind these changes. For example, the decrease in births during the Great War has had an impact on the decline of centenarians in recent years, but life expectancy increases have contributed to an increased number of centenarians in the population. The improvement of male life expectancy relative to females in recent years means the proportion of males aged 90+ is increasing relative to females.
Time period covered:	The Population and Migration Statistics team produced a bespoke population estimate for census day (20th March 2022) available to us in spring 2023.
Use of Data:	Validation of population estimates.

Source information

The centenarians publication details the estimated number of people by sex aged 90 to 104, by single year of age, and the number of people aged 105 and over in Scotland. The Population and Migration Statistics team produced a bespoke census day estimate for use in validation with the estimated number of people by sex aged 90 to 120, by single year of age. Age relates to age at last birthday. These statistics give an important insight into the most rapidly growing age-group of Scotland's population and are used in the calculation of life expectancy statistics for Scotland.

This publication uses a number of different data sources to gather the relevant data needed to produce these outputs. The mid-year population estimates from National Records of Scotland, deaths from the civil registration system and Pension Age Client Group from the Department of Work and Pensions (DWP) are all used and collected on an ongoing basis throughout the year.

In terms of accuracy, the Centenarians data are based on the Kannisto-Thatcher (KT) method. This method has also been adopted by the Office for National Statistics to produce estimates for the very old in England and Wales and by the Northern Ireland Statistics and Research Agency to produce equivalent estimates for Northern Ireland. The KT method is considered robust at national level. It is used by the World Health Organisation to compute population estimates for ages 85+, it is the method used to construct the Human Mortality Database and it has outperformed other methods in numerous studies. The estimates each year are measured against DWP, Scottish National Health Service Central Register and the Community Health Index (CHI). These measurements are found to be broadly similar to the data these organisations hold.

Data supply and communication

This data are provided on an ongoing basis by a variety of different stakeholders. The Population and Migration Statistics team have links with different bodies such as the DWP so that any issues with quality assurance can be resolved with relative ease.

Quality Assurance undertaken by data supplier

The Population and Migration Statistics team check the data they receive from other stakeholders by carrying out checks against previous years' data. For example, the number of 90 year olds between 2004 to 2009 should reflect the lower birth rates within the war years while from 2010 onward, evidence of the large 1920 post war cohort is expected. These checks ensure consistency and completeness of the data.

The estimated very old population includes all who reside in Scotland, whatever their nationality, so members of the UK and non-UK armed forces are included but short-term international migrants are excluded.

The team regularly compare their estimates with those from the DWP to increase confidence in the estimation methods used. For example, Scotland level estimates for 2013 had a correlation coefficient of greater than 0.99 for both males and females.

A report is available on the NRS website providing information on the quality assurance arrangements for data used in the centenarians publication.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.
- Checking the age distribution of the population.

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

- Death registration is compulsory and must be registered within 8 days of the death date.
- This is an administrative data source so it is not vulnerable to sampling error.
- Coverage and validation of multiple data sources.

Limitations of the source

- Date of birth on death certificates is not validated, leading to possible inaccuracies of age.
- Any error in the 90 and over 2011 census estimate is thereby carried forward to the inter-census mid-year population estimates and will be reflected in the single year of age 90+ estimates.
- The source for the 90+ population is the NRS Mid-year Population Estimates (MYEs), and is dependent on the quality of those estimates (see 2.1 above).

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low" as these data are designated as National Statistics which means the data meet the standards set by the Code of Practice for Statistics and the team working with these data has provided comprehensive information about what they do to make sure the data are good quality.

2.3 Small Area Population Estimates

Data Supplier	National Records of Scotland (NRS)
Supplier information	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 Small Area Population Estimates (SAPE) publication produces statistical estimates of the population of Scotland in small areas known as data zones by sex and by age.
Data type (counts or unit records)	Counts
Data Content:	The publication provides population counts by sex and by age (single year of age 0-89 and 90+) by data zone. Data zones nest inside the 32 Local Authorities/Council Areas that make up Scotland. They are designed to include 500 to 1000 people. There are 6976 2011 data zones (designated following Census 2011). In 2020 there was an average of 784 people in each data zone.
Time period covered:	The Population and Migration Statistics team produced a bespoke population estimate for census day (20th March 2022) available to us in late 2022.
Use of Data:	Validation of population estimates.

Source information

The Small Area Population Statistics (SAPE) are produced using the demographic cohort component method, the same method used to produce the mid-year population estimates (MYE; see 2.1 above for further details). The SAPE for mid-2011 onwards are based on the 2011 Census and 2011 Data Zone boundaries and relate to the usually resident population. In simple terms, this means that population estimates are estimates of people where they usually live.

SAPE are published after the MYE but based on the same input data as for the MYEs. They are also adjusted to match the population estimates from the MYE for council areas. Data zones are designed to nest into council areas although boundary changes may mean their borders may no longer align precisely. Data zone population estimates are aggregated to create population estimates for various geographies e.g. urban and rural populations or to feed into the calculation of settlement and locality population estimates on a 'best-fit' basis. A methodology report is available with further details of the method used to produce SAPE including an overview of the cohort component method.

Data supply and communication

The data to produce the MYE and SAPE are provided annually by a variety of different stakeholders. The Population and Migration Statistics team in NRS have forged close links with ONS, NISRA, Scottish Government and NHSCR, some of the main data suppliers for their publications. These ongoing relationships with different organisations mean that any quality assurance issues or changes can be resolved with minimal delay.

Quality Assurance undertaken by data supplier

When the Population and Migration Statistics team within NRS receive data for the various components, checks are carried out and comparisons made with previous years' data to gauge consistency and completeness of coverage. The data are then processed electronically to produce first the mid-year estimates and then the SAPE. Quality assurance takes place at each stage of this process. This may include calculation of demographic rates and ratios to help check the plausibility of the data. NRS also have processes in place to check the suitability of the administrative sources used in producing population estimates.

The Population and Migration Statistics team also meet with the data suppliers and discuss any issues that might affect the data. These close working relationships ensure that issues and planned changes to data collection can be considered before the data are used to produce population estimates.

A page is available on the NRS website providing information on the quality assurance arrangements for administrative data used in population estimates, along with information on the suitability of each data source used in the production of population estimates.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.
- Checking the age distribution of the population.

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 population of an area includes all those usually resident there, regardless of nationality.
- Population estimates for Scotland are comparable with population estimates from the UK and most countries around the world.

Limitations of the source

- Short-term international migrants are excluded.
- Those aged 90+ are reported as a single age group rather than reported by single year of age.
- The population base are the population estimates for Census 2011 which are now over 10 years old. There are no means of verifying the true population between censuses. As a result, any uncertainty in the population estimates will increase with time as we move further from the previous census.

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low", as there is a well-documented quality assurance process to ensure SAPE are achieving high data quality standards. The SAPE is a National Statistics publication, so they have already been judged of sufficient quality for that accreditation. The statistics are based on the best available sources, including high quality administrative data sources for births, deaths, migration and special populations.

2.4 Scottish Household Survey & Scottish Household Condition Survey

Data Supplier	Scottish Government
Supplier information	The Scottish Government has a range of responsibilities including: the economy, education, health, justice, rural affairs, housing, environment, equal opportunities, consumer advocacy and advice, transport and taxation. The Scottish Household Survey (SHS) is a survey of private residences in Scotland.
Data type (counts or unit records)	Counts
Data Content:	This publication covers a variety of different areas of the population of Scotland including the overall characteristics of households in Scotland, its neighbourhoods and communities. Topics covered in its publication include housing, neighbourhoods and communities, finance, internet, physical activity and sport, public services and institutions, environment, volunteering, culture and heritage and childcare.
Time period covered:	Survey results from 2021, published in summer 2023.
Use of Data:	Validation of population estimates.

Source information

The SHS is usually a face-to-face survey of a sample of people in private residences in Scotland. This survey covers all 32 local authorities in Scotland and it is voluntary to take part. This survey also includes evidence of the physical condition of households (Scottish Household Condition Survey, SHCS) and robust evidence on the composition, characteristics, attitudes and behaviour of private households and individuals. This survey helps meet central and local government needs for policy relevant data in this field.

This publication also has built in flexibility to respond to different data needs regarding geography and frequency for example to provide some data annually at local authority level, and some biennially at national level.

The publication's accuracy or robustness in terms of quality of the data is usually helped by using Computer Assisted Personal Interviewing (CAPI). This has a number of important advantages over 'pen and paper' interviewing techniques, including faster interviews, automatic edit checks and a quicker flow of information from the survey interviewer to the survey database.

All face-to-face interviewing, including the SHS, was suspended in 17th March 2020 to help prevent the spread of COVID-19. The 2020 and 2021 surveys were carried out using a telephone/video approach.

Response rates for telephone surveys are generally lower than for face-to-face interviews. Around 4500 households were interviewed for the 2020 SHS (one third of which were face-to-face before March 2020) compared to around 10500 in 2019. Due to this smaller sample size, the 2020 data is not broken down as in previous years, e.g., data is not provided for individual local authorities. The 2021 survey achieved 9,952 interviews, just below the target of 10,450, meaning data are available at local authority level. Due to the difference in the mode of interviewing for both 2020 and 2021 and an observed change in the demography of those responding, the SHS team state data from these years are not comparable to previous years.

Following suspension of face-to-face interviewing in March 2020, no further data collection for the Scottish House Condition Survey was made in 2020. In 2021, the physical survey is being carried out by an 'external+' approach including an external-only inspection, supplemented with additional information e.g. from Energy Performance Certificates and the householder speaking to surveyors over the telephone. The approach has been designed to provide reliable estimates of key statistics including on fuel poverty, energy efficiency and external repairs. No data was collected on internal aspects or housing standards. The full number of inspections (3000+) are expected to be carried out. SHCS statistics for local authorities require three years of survey data, and further consideration is being given to how to present local authority level analysis given the lack of data for 2020.

Data supply and communication

The data are usually gathered through interviews in the individual householder(s) home. This approach has been amended to a telephone/video approach for 2020 and 2021. The survey asks questions of a random sample of people in residences

across Scotland, ensuring all 32 local authorities are represented. In a usual year, around 10,000 households take part in the survey, due to design improvements the precision of these estimates is highly accurate.

Quality Assurance undertaken by data supplier

The data are subject to various checks which are important to maintain the high quality of the data. The main data file is subject to checks and editing involving range checks, simple logic checks and complex logic checks. The data then undergo two additional processes. Firstly, the calculation of derived variables such as the age and sex of the Highest Income Householder and secondly, the imputation of household income. The data are then forwarded onto the Scottish Government who run further checks on the data.

The SHCS forms are also subject to robust checks by being uploaded into the physical survey validation system. Surveyors are then shown a list of all the errors picked up by the validation program. Additionally, they are shown a list of all the entered data, with a description of the variable next to each piece of data, and with the data split into representations of each page of the form. The validation system shows the data and the failed edits as well as showing the photographs of the property. Corrections are then made and each form rechecked until it passes all edits.

A methodology report is available providing further information on the quality assurance arrangements for data used in the SHS publication.

The SHS team conducted a thorough analysis of the effects of the change of mode to telephone/video survey for the 2020 survey due to the COVID-19 pandemic. Any change in approach means that, in addition to any real change, estimates may be affected by a) changes to the profile of the responding sample (non-response bias) and/or b) changes to how questions are asked and answered (measurement error). The team found evidence of both effects – response rate dropped from 63% in 2019 to 20% in 2020 and the profile of those responders changed (e.g., an increase in owner occupiers and a decrease in responders from the most deprived areas). The team also found that interview mode affected how some people answered questions, e.g., a reduction in people answering neutrally (don't know/neither agree or disagree) likely due to the absence of showcards legitimising this option. After calibration weighting, for most measures where major changes would not be expected, the estimates were in line with those from 2019. Notable exceptions include tenure and length of time at the address.

The pandemic itself likely influenced how people answered certain questions e.g. relating to how lonely they felt or the confidence they had in their local health services. Although the 2020 and 2021 data may not be comparable to previous years, it does not mean they are not robust in themselves.

A report is available providing further information on the analyses the team did on the effect of the change in survey mode for the 2020 data collection.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.

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

- Strong and varied response rate across Scotland is key for quality of the data.
- The weighting strategy used by the survey minimises the extent of bias.

Limitations of the source

- Geographical limitations below local authority area.
- Limited in the amount of detail it can collect.
- The survey only produce estimates and these estimates are limited by a number of factors. The factors are sample coverage, sampling variability and the bias in the achieved sample.
- Analyses identified effects of the change of interviewing mode due to the ongoing COVID-19 pandemic both in terms of non-response bias and measurement error.

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concern has been set to 'Medium.' The Scottish Household Survey publication is usually a National Statistics publication which means the data meet the standards set by the Code of Practice for Statistics and the team working with these data have provided information about what they do to make sure the data are good quality. Due to the change in interview mode and reduction in sample size, the 2020 and 2021 publications are Experimental Statistics publications. However, the team has provided an extensive evaluation of likely effects of the change in mode meaning the 2020 and 2021 data are still of high quality.

2.5 Children's Social Work Statistics

Data Supplier	Scottish Government
Supplier information	The purpose of Scottish Government is to hold a range of responsibilities that include: the economy, education, health, justice, rural affairs, housing, environment, equal opportunities, consumer advocacy and advice, transport and taxation. The Children's Social Work Statistics area produces statistical estimates of children who were formally looked after, under child protection measures or in secure care.
Data type (counts or unit records)	Counts
Data Content:	This publication covers a variety of different areas of children's social work statistics in Scotland including looked after children, child protection and secure care accommodation. **Children looked after** gives a high-level overview of those defined as looked after by a local authority. These children may be looked after for a number of reasons including neglect, abuse, complex disabilities which require specialist care, or involvement in the youth justice system. **Child protection** focuses on the number of children in Scotland who need protecting from abuse or neglect. The risk of harm or neglect is considered at a Child Protection Case Conference. **Secure care accommodation** provides analysis on the number of young people who present high risk to themselves or others and can only be authorised following a decision through the Children's Hearing System or a Court.
Time period covered:	Data for 2021/2022 published in summer 2023.
Use of Data:	Validation of population estimates.

Source information

Children's social work statistics present information collected from local authorities and secure units on children and young people, generally younger than their mid-twenties, who were formally looked after, in receipt of aftercare support, under child protection measures, or in secure care at some point during the previous year before publication (for example August 1st 2019 to July 31st 2020).

There are a number of ways that a child may become looked after, on the child protection register or in secure care. Children may be referred to the Children's Reporter, become voluntarily looked after or come via the criminal justice system.

The majority of referrals to the Children's Reporter are on care and protection grounds, although a small proportion are on offence grounds. Young people are placed in secure care either as an outcome of the criminal justice system or through a referral to the Children's Reporter.

Data supply and communication

Data are provided on an annual basis by all local authority areas in Scotland as well as secure units in Scotland. The Children and Families statistics team have strong links with those involved in providing this data which ensures the quality of the data continues to be delivered at a high standard.

Quality Assurance undertaken by data supplier

The data for all three parts of this publication – looked after children, child protection and secure care – come from administrative data held by local authorities and secure units. As this information is used to monitor and manage these sectors it should be robust and accurate.

Automated validation checks are undertaken at the point the data are submitted. The Children and Families statistics team undertake a range of validation checks on administrative data as part of the quality assurance process of preparing this National Statistics publication. These procedures include: trend analysis, comparing against other available sources, and checking outliers with data providers. The data providers are then asked to confirm their data – for looked after children and child protection data this confirmation comes from local authorities, for secure care accommodation this confirmation comes from secure units. In cases where concerns about data quality outweigh the value of having an estimated figure publicly available, the data would not be published.

The data on looked after children sometimes experience delays between the data being updated on the local authority management system and the event affecting the child occurring. Scottish government are currently investigating using new data to update the figures for previous years in order to resolve this issue.

A report is available on the Scottish Government website providing information on the quality assurance arrangements for data used in children's social work statistics publication.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.
- Checking the age distribution of the population.

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

- Data come directly from local authorities and secure units, ensuring accuracy.
- The use of administrative or management information from other organisations enables Children's Social Work statistics team to reduce the burden it places on data providers in terms of the time and cost involved in data collection exercises.

Limitations of the source

- Delays sometimes occur between updating the data on system(s) and the event involving the child.
- Figures on individual secure units have been estimated due to issues with the provision of the data.

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low" because these data are designated as a National Statistic which means the data meet the standards set by the Code of Practice for Statistics and the team working with these data has provided an adequate summary about what they do to make sure the data produced are of high quality.

2.6 Care Home Census

Data Supplier	Public Health Scotland (PHS) Information Services Division (ISD)
Supplier information	PHS was established in 2020 as Scotland's new national organisation for public health. It incorporates Information Services Division, NHS Health Scotland and Health Protection Scotland. From 2003 to 2009, the Scottish Government published the annual Care Home Census. From 2010 onwards, it was published by ISD Scotland. ISD Scotland became part of Public Health Scotland (PHS) on 1 April 2020. ISD provided health information, health intelligence, statistical services and advice to support the NHS in progressing quality improvement in health and care.
Data type (counts or unit records)	Counts
Data Content:	This publication covers a variety of different areas in the care home sector in Scotland including care homes for adults, care homes for older people and care homes for those with disabilities. The **All adults** section of the publication covers all Main Client Groups and provides analysis on the number of adults in care homes by sector. For long stay residents, information is also provided on disability, age, sex and length of residency. **Care homes for older people** focuses on how many care home residents in Scotland stay in older people care homes and what sector provides the care for these residents/number of admissions and deaths/length of residency and the number of long stay residents with different health characteristics, for example, dementia. **Care homes for learning disabilities, mental health problems and physical and sensory impairment** provides analysis on the sector providing care to these residents with disabilities.
Time period covered:	March 2022 Care Home Census published summer 2023.
Use of Data:	Validation of population estimates. To assist communal establishment population estimates during data processing.

Source information

The Scottish Care Home Census for adults takes place on 31 March each year with data published later that year. No Census took place in March 2020 to reduce the administrative burden on care home managers due to the ongoing COVID-19 pandemic.

The Scottish Care Home Census includes all the care homes for adults (aged 18 and older) registered with the Care Inspectorate, as listed on their website. The Care Inspectorate was formed under the Public Services Reform (Scotland) Act 2010 and it is the independent regulator of social care and social work services across Scotland. All adult care homes are run by one of the following three sectors: Local authority/NHS, Private sector or Voluntary sector.

The Scottish Care Home Census collects data annually via the Care Inspectorate's eForms system. All adult care homes registered with the Care Inspectorate in Scotland are asked to submit their data, however, it is not mandatory for them to do so.

The report is accompanied by fifteen data tables which provides more in-depth information at Scotland and local authority level.

In early 2020, the Care Home Census had its National Statistics status suspended due to delays in publishing historic data caused by slow progress in finalising a data sharing agreement with the Scottish Government. The delays have been rectified and the publication currently has Official Statistics status until the Office for Statistics Regulation reassesses this status.

Data supply and communication

The data are provided by care homes across Scotland using a form which asks for aggregated data at care home level. A further part of the data collection then asks for information such as gender, date of birth and health characteristics on individual Long stay residents only.

Quality Assurance undertaken by data supplier

The Care Inspectorate's IT department have set up several automated checks in their eForms system. For example an alert will pop up on the screen when a weekly cost is entered that appears to be too low or too high. An alert will also occur when the total number of residents at the end of the census year does not equal the total number of residents at the start of the census year plus the number of admissions minus the number of deaths and discharges. The electronic census form cannot be submitted to PHS until this is correct.

PHS further validates the data by comparing the aggregate data to previous census years to identify potential anomalies. Where there are differences, PHS will then contact the care homes to ask them to resubmit the correct figures.

The Care Inspectorate compare the total number of residents and the total number of care homes collected from the Care Home Census with similar data collected from the Care Inspectorate's annual return.

PHS, the Scottish Government, and the Care Inspectorate also meet regularly to discuss and update guidance on data items where there are suspected data quality issues.

Data not submitted by care homes were estimated by PHS using data from the previous year, or average values from care homes with the same Main Client Group and Sector and which are of a similar size, or by scaling-up (or down) proportions.

A report is available on the PHS website providing more information on the data quality and completeness of this publication.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.
- Checking the age distribution of the population.

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

- Advanced automated checking eForm system
- Designated Official Statistics, complying with the Code of Practice for statistics

Limitations of the source

- Care homes are not required to complete the census: of the 1,051 care homes for adults open on 31 March 2022, there were 737 (70%) care homes that submitted data for at least part of the Census and 314 (30%) care homes that did not submit any data.

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low" as these data are designated as Official Statistics which means the data meet the standards set by the Code of Practice for Statistics and the team working with these data have provided information about what they do to make sure the data are good quality.

2.7 Estimates of Households and Dwellings in Scotland (including Small Area Statistics on Households and Dwellings)

Data Supplier	National Records of Scotland (NRS)
Supplier information	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 Estimates of Households and Dwellings in Scotland publication produces the latest statistics on occupied and vacant dwellings, second homes and trends in household types
Data type (counts or unit records)	Counts
Data Content:	This publication covers a variety of different housing topics in Scotland, such as national household and dwelling estimates, council area estimates and neighbourhood level estimates (Small Area Statistics). National household and dwelling estimates focus on any changes in counts of households and dwellings including occupancy and the length of time homes and dwellings are unoccupied. Average size of the household and its characteristics are also analysed in this section. Council area estimates focus on the decreases/increases in household dwellings in council areas, percentage increases in each council area and take into account the number of unoccupied households and dwellings in these areas. Neighbourhood level estimates provide analysis on data zone areas, which are small enough to represent local communities. This section takes on a similar structure to the others by again including vacant dwellings and dwellings that are exempt from council tax.
Time period covered:	September 2021 (Households and Dwellings in Scotland, 2021) The Household Estimates and Projections team also produced a bespoke household number estimate for census day (20th March 2022) available to us in summer 2023.
Use of Data:	Validation of population estimates. Production of the Alternative Household Estimate. Application of a household adjustment to the census estimates

Source information

The Estimates of Households and Dwellings in Scotland publication provides estimates for: Scotland, council areas, Strategic development plan areas, national parks and data zones.

This publication uses a range of data sources to gather the relevant data needed to produce the outputs in the publication. Council Tax billing systems, at council area level, from the Scottish Government's Council Tax Base (Ctaxbase) data collection are used as well as Council Tax billing systems from NRS data zone level collection on occupied and vacant dwellings and the Assessor's portal, which provides information regarding Council Tax, valuation for rating and electoral registration.

In terms of accuracy, data on occupied and vacant dwellings are obtained from Council Tax billing systems. Different types of properties are entitled to different Council Tax discounts and exemptions. There are inconsistencies between the ways in which some councils record these discounts and exemptions. There can also be differences between results from the councils' data collection from Council Tax billing systems and the data zone level data collection. It is possible that not all information held on Council Tax billing systems and the Assessors' Portal is up-to-date. For example, councils may not be notified immediately of a change in the circumstances of a household which affects eligibility for a Council Tax discount or exemption. It may also take time for changes as a result of new building or demolition to be recorded.

There are also some definitional differences between Council Tax and Census data, for example some communal establishments are counted as separate units for Council Tax purposes, but they are counted as a single communal establishment in Census.

Data supply and communication

The data are provided by council tax billing systems around September, with the assessor's portal data being communicated around January.

Quality Assurance undertaken by data supplier

To administer Council Tax, each council keeps a record of all dwellings, as well as information on those receiving Council Tax discounts or exemptions, such as vacant dwellings and second homes. This makes it possible to estimate the number of dwellings that are occupied, by subtracting the number of vacant dwellings and second homes from the total number of dwellings. An occupied dwelling is approximately equivalent to a household.

For each category, from all dwellings to occupied exemptions, a clear description is given and summary information provided about each category and about where the data and information has come from. 'Ctaxbase' is the main source of information for some of these categories and the figures from 'Ctaxbase' give the number of properties in each local authority area, including those with exemptions and discounts.

The Scottish Household Survey, Assessors' portal and the 2011 Census Reconciliation Report – Households are also used to compare and contrast the data and any changes in definitions or areas identified by this publication.

The publication is available on the NRS website providing information on the quality assurance arrangements for the data.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.

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

- Coverage and validation of multiple data sources.
- Council tax registration is compulsory, so the data sources used to produce this publication are as accurate as possible.

Limitations of the source

- Reclassification of vacant dwellings and second homes in 2013 has meant some council's billing systems are not always properly picked up in the Ctaxbase collection.
- There are differences between the number of occupied dwellings and number of households in the census, the reason being that an occupied dwelling might be shared by more than one household.

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low" as these data are designated as National Statistics, which means the data meet the standards set by the Code of Practice for Statistics. The publication has provided enough information about the quality assurance process to ensure the data are of high quality.

2.8 Scottish Health Survey

Data Supplier	Scottish Government
Supplier information	Scottish Government holds a range of responsibilities including: the economy, education, health, justice, rural affairs, housing, environment, equal opportunities, consumer advocacy and advice, transport and taxation. The Scottish Health Survey (SHeS) publication analyses the health and wellbeing of the population by carrying out interviews and data analysis on a variety of different health related issues in private households in Scotland
Data type (counts or unit records)	Counts
Data Content:	This publication covers a variety of different topics in Scotland including mental health and wellbeing, general health and smoking consumption among other areas of health in Scotland. The topics included in the 2021 survey are: general health, mental wellbeing, dental health, accidents, CPR training, discrimination and harassment, stress at work, loneliness, social capital, eating habits, physical activity, smoking, alcohol, diseases of the heart, lungs and chest and self-reported height and weight. A new topic measuring drug use has also been included for the first time and a module on COVID, including questions on long-COVID and vaccinations.
Time period covered:	2021 survey results published November 2022.
Use of Data:	Validation of population estimates

Source information

The SHeS reports on a variety of different health issues and areas amongst the population of Scotland. Scotland has one of the lowest life expectancies in Western Europe, so the work of this publication is ever increasingly important to national policy to understand the public health issues that arise in Scotland.

Fieldwork is usually an integral part of this survey, with households contacted in advance, should they wish to take part. Due to the COVID-19 pandemic, all face to face fieldwork was cancelled in March 2020. Instead, it was decided to conduct a telephone survey in August and September 2020 collecting data on some of the measures usually asked in the SHeS. Potential participants aged 16 and over were contacted by letter and asked to opt-in to taking part in an interview conducted over the phone. Interviewing was conducted using Computer Assisted Telephone Interviewing (CATI), where the questionnaire answers were input directly to a laptop.

The telephone survey methods differed from those used in the SHeS series previously and while the survey includes many questions and key indicators from the face-to-face surveys, the change in mode of administration, along with the different approach to sampling, is likely to have impacted the responses received and thus comparability with the previous SHeS data. The SHeS team worked hard to anticipate and control for any biases introduced by the new survey mode and to adjust their weighting scheme appropriately.

Due to the ongoing impact of the pandemic, the 2021 survey was also conducted by telephone and included an online self-complete section and an online dietary tool. Unlike the 2020 survey, all parts of the SHeS were asked to both adults and children increasing comparability with previous years. Following agreement from the Chief Medical Officer that doorstep contact could proceed, the last third of the sample was worked using a knock-to nudge approach with interviewers visiting sampled addresses to encourage response to the telephone survey.

The results of the 2021 health survey were published in November 2022. An increased incentive was provided in deprived areas to try to encourage response in these areas for which there were lower than usual levels of response in the August/September 2020 telephone survey. The inclusion of online self-completions for more sensitive questions (with paper versions for those who would prefer) should make the results for those questions more comparable to results from the usual face-to-face survey. Additional information and analysis on the effects of the change of mode can be found in a technical report published by the SHeS team.

Data supply and communication

The SHeS is conducted by ScotCen Social Research in collaboration with the Office for National statistics (ONS), the Social and Public Health Sciences Unit (MRC/CSO SPHSU) at the University of Glasgow, the Centre for Population Health Sciences at the University of Edinburgh and the Public Health Nutrition Research Group at the University of Aberdeen.

The data collected in this survey comes from a variety of different private households in every region of Scotland. The sample specification for the SHeS 2020

telephone survey was designed by the Scottish Government. A random sample of addresses (11,000 addresses) was selected from the Postcode Address File (PAF), using a stratified design.

Quality Assurance undertaken by data supplier

Experienced interviewers who had previously worked on the face-to-face SHeS were fully briefed in advance of fieldwork. The briefing covered the survey's content and procedures in detail and key changes from the face-to-face survey were highlighted. A full set of written instructions that covered survey procedures were also provided to interviewers before starting work.

A large number of quality control measures were built into the survey at the data collection stage and thereafter, to monitor the quality of interviewer performance. The computer program used by interviewers for example, has in-built soft checks (which can be suppressed) and hard checks (which cannot be suppressed) associated with particular interview questions. When uncommon or unlikely answers are entered, or answers outside a predetermined range, these checks were triggered and appear as a warning message on the interviewers' laptop. The interviewer is either encouraged to double-check the entered response (a soft-check) or asked to change it (a hard-check).

Finally, recalls are carried out at 10% of productive households. These recalls checked with the participants that interviewers had followed the correct survey procedures when conducting the interview.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.

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

- Uses respected research methods in different health areas such as the Warwick-Edinburgh mental wellbeing scale, which adds weight to the data quality.
- The number of quality assurance protocols for the interviewer ensures a high quality performance in the collection of the responses.
- Every region in Scotland is represented.

Limitations of the source

- An interview can cause biases. For example, the respondent's answers can be affected by his reaction to the interviewer's race, class, age or physical appearance, regardless of the type of questions being asked. Due to the change in mode from face-to-face to telephone interviews, additional biases are likely to be at play, particularly related to the opt-in selection process and lower response than usual in deprived areas.
- Small sample size, relative to the population.

Risk Matrix

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 value of "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concern of the SHeS has been set to "Low." The SHeS publication has National Statistics status adhering to the high professional standards set out in the Code of Practice for Official Statistics. The extensive checks and quality assurance performed by the SHeS team mean the data are of high quality.

2.9 Business Register and Employment Survey

Data Supplier	Office for National Statistics
Supplier information	The Office for National Statistics is the UK's largest independent producer of official statistics and the recognised national statistical institute of the UK. It is responsible for collecting and publishing statistics related to the economy, population and society at national, regional and local levels. It plays a leading role in national and international good practice in the production of official statistics.
Data type (counts or unit records)	Counts
Data Content:	The Business Register and Employment Survey (BRES) publication covers a variety of different areas in employment in Scotland and the UK such as numbers of those in employment, regional variations and results by industry. **Employment numbers** focuses on the estimated number of employees in the UK. **Regional variations** includes increases and decreases within different countries in the UK which includes Scotland. **Results by industry** analyses the increase or decrease in employment in different sectors such as manufacturing and financial and insurance activities.
Time period covered:	Survey results from September 2021, published in September 2022.
Use of Data:	Validation of population estimates

Source information

The BRES survey publishes employee and employment estimates at detailed geographical and industrial levels. It collects comprehensive employment information from businesses in England, Scotland and Wales, representing the majority of the Great Britain economy. Independently collected Northern Ireland data are then combined to produce estimates on a UK basis.

The BRES publication provides data on the number of employees in the UK in the public or private sector working on full time or part time basis. It is regarded as the definitive source of official government employee statistics by industry. It publishes counts by Standard Industrial Classification (SIC 2007).

Employment in the publication is calculated by adding the number of working owners to the number of employees employed by a business, where working owners include sole traders, sole proprietors and partners who receive drawings and/or a share of the profits, but are not paid via pay-as-you-earn (PAYE).

The BRES data and estimates are widely used, both within and outside government, and are a vital source of business employee information. The main users of the output include: Eurostat, the Scottish and Welsh Government, Department of Business, Energy & Industrial Strategy (BEIS) as well as being used in other official statistics such as Workforce Jobs and the Annual Business Survey (ABS) statistics.

Data supply and communication

BRES uses the Inter-Departmental Business Register (IDBR) as the survey sampling frame to identify businesses within the scope of the survey. It uses a stratified random sample where the strata are defined by SIC 2007, country and employment size of a business to select which businesses are required to complete the survey. The survey is completed on a paper questionnaire, editable PDF or, for businesses with a large number of sites, through a Secure Electronic File Transfer spreadsheet through a secure online platform. Further work is being done to expand the online option to make it available for more businesses. Businesses are given one month from the reference date to return their completed survey.

Quality Assurance undertaken by data supplier

BRES is based on a sample survey of businesses to estimate the total number of employees, which gives rise to sampling error. The actual sampling error for any estimate is unknown but a standard error can estimated from the sample. This provides a means of assessing the accuracy of the estimate when an unbiased or approximately unbiased estimator is used. The lower the standard error, the more assured that the estimate is close to the true value. The coefficient of variation (CV) can be calculated as the standard error divided by the estimate and it is used to compare the relative accuracy across surveys or variables. The CV is one indicator of the quality of the estimate; the smaller the CV the higher the precision.

Various procedures are in place to ensure that errors are minimised. Year-on-year comparisons are made at respondent, local unit and aggregate level. Disparities are investigated by those working on the publication to ensure consistent annual returns.

Comparison checks are made against other surveys to ensure consistent values across industries from different surveys.

As BRES is used both to update the IDBR and to produce estimates, there is a risk of feedback bias. To reduce this bias to a minimal level, the register employee count used in the analysis is modelled using survey data and the modelled values are used in an auxiliary variable in calibration. Another indicator of accuracy is reliability, which can be measured by assessing the difference between the first published estimate and the final revised figure.

BRES adheres to a revisions policy whereby current survey estimates together with a revision of the previous year's survey estimates are published. Late returns or information received in the course of the following year's survey may lead to changes to the estimates after the provisional publication. Such changes are incorporated into the figures when the revised estimates are published in the following year.

More information about quality assurance undertaken for the BRES can be found in the Quality and Methodology Information report.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.

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 response rate from businesses is high. Even though the 2020 survey had lower than average response rates, these were still above 70%.
- Good quality estimates for detailed breakdowns by geography and industry due to large sampling size.
- BRES provides both employee and employment data and is particularly recommended for analysis of employee data.

Limitations of the source

- Since BRES is based on a sample of businesses, it can be affected by sampling variability. For example, quality of estimates might deteriorate for smaller geographies.
- Caution is needed if the CV variation is greater or equal to 20%.
- Businesses are registered where they are based. Due to the COVID-19 pandemic many people are working from home rather than on site and this

may lead to mismatches between counts of employees by industry at lower geographies (e.g., local authorities) as when asked where they mainly work from they will answer from home rather than give their work address. If their home is not in the same local authority as their work, the counts will be mismatched. The magnitude of this issue is likely to vary between local authorities.

Risk Matrix

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 value of "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low" despite the survey being open to sampling variability. This is a National Statistics publication and the quality assurance and validation processes carried out for the publication is assured and accurate throughout the publication process from sources to final outputs.

2.10 Annual Population Survey

For the purposes of validating Scotland's Census 2022, we used Annual Population Survey data published by the Scottish Government Labour Market Statistics team who produce an annual publication People, Places & Regions using the Scottish data from the APS.

Data Supplier	Office for National Statistics
Supplier information	The Office for National Statistics is the UK's largest independent producer of official statistics and the recognised national statistical institute of the UK. It is responsible for collecting and publishing statistics related to the economy, population and society at national, regional and local levels. It plays a leading role in national and international good practice in the production of official statistics.
Data type (counts or unit records)	Counts
Data Content:	The Annual Population Survey is a UK-wide survey with the largest sample of any household survey and allows generation of statistics at the local level. It is an important source of information on the labour market, capturing levels of employment and unemployment as well as covering topics such as ethnicity and education. Its data comes from the Labour Force Survey with sample boosts in England and each devolved nation.
Time period covered:	September 2021 to September 2022.
Use of Data:	Validation of population estimates

Source information

The Annual Population Survey (APS) is a continuous household survey (sample size 320,000), covering the UK. It covers topics including employment, housing, ethnicity, religion, health and education. Its purpose is to provide information on social and socio-economic variables at local levels. It enables monitoring of a variety of topics between censuses and provides local area level information for labour market estimates.

Data supply and communication

The APS is not a stand-alone survey; it uses data combined from 2 waves of the main Labour Force Survey, collected on a local sample boost. These boosts are sponsored by the Department for Work and Pensions, Department for Business, Innovation and Skills, the Welsh Government and the Scottish Government.

The Scottish Government funds the boost to the LFS sample in Scotland, taking the sample size from approximately 4,000 households to 13,000 households for the latest time periods.

Quality Assurance undertaken by data supplier

As the Annual Population Survey (APS) is a sample survey, it provides estimates of population characteristics rather than exact measures. In principle, many random samples could be drawn and each would give different results, due to the fact that each sample would be made up of different people, who would give different answers to the questions asked. The spread of these results is the sampling variability. Confidence intervals are used to present the sampling variability. For example, with a 95% confidence interval, it is expected that in 95% of survey samples, the resulting confidence interval will contain the true value that would be obtained by surveying the whole population.

The publication has extensive checks on the response rate from households. Full response indicates a household in which each household member has answered all applicable questions. Partial response indicates a household in which questions were not completed because someone refused to be interviewed, refused part way through the questionnaire, or refused to let someone else answer on his or her behalf. However, at least one question block must have been completed. Outright refusal indicates a household that refuses to respond to the survey and the interviewer feels that there is no chance of an interview at the current or in any future wave.

Circumstantial refusal indicates a household where the respondent refuses to respond because of a temporary circumstance (for example, going on holiday, too busy during the field period). A circumstantial refusal enables an interviewer to call back at the next wave.

A report is available on the methodology and quality assurance practices of the APS.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.
- Checking the age distribution of the population.

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

- Large sample size with Scotland-specific boosted sampling
- Covers broad range of topics relating to employment and the labour market
- Established methods applied to reduce sampling bias

Limitations of the source

- The survey only produce estimates and these estimates are limited by a number of factors. The factors are sample coverage, sampling variability and the bias in the achieved sample
- The sample design provides no guarantee of adequate coverage of any industry, as the survey is not industrially stratified and workers under 16 years of age are not covered. The coverage also omits communal establishments apart from NHS housing and students in halls of residence. Members of the armed forces are only included if they live in private accommodation.
- Impacts of the COVID-19 pandemic on the labour market and labour market statistics can change rapidly, it will be important to be aware of policies and practices in place in the time period when the data were collected versus the snapshot of the population captured on census day.

Risk Matrix

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 value of "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low" despite the survey being open to some limitations which have been highlighted. The quality assurance and validation processes carried out for these data are assured and accurate throughout the publication process from sources to final outputs.

2.11 Transport and Travel Statistics Scotland

Data Supplier	Transport Scotland
Supplier information	Transport Scotland is the national transport agency for Scotland. Its statistics branch produces estimates of how the population of Scotland travels.
Data type (counts or unit records)	Counts
Data Content:	This publication reports the transport and travel data gathered by the Scottish Household Survey. **Personal travel** gives a high-level overview of who and why people travel where they do for work, school etc. **Public transport travel** focuses on the satisfaction of public transport services, as well as accessibility and changing modes of travel. **Walking and cycling** provides analysis on the numbers of people walking or cycling as part of their travel experiences and also charts bicycle accessibility. For Census validation, we require data related to travel to place of work or study.
Time period covered:	2021 survey results, published in 2023.
Use of Data:	Validation of population estimates

Source information

This publication provides the results of the Transport and travel related questions asked in the Scottish Household Survey (SHS, see 2.4 above), including information from the travel diary, and uses data from other sources to provide some context around transport and travel in Scotland.

The Travel Diary is a component of the survey which involves respondents recounting details of all the journeys they made the previous day. A journey can consist of one or more stages. A new stage is defined when there is a change in the form of transport or when there is a change of vehicle requiring a separate ticket.

Data supply and communication

The majority of the data is gathered through the SHS, which takes a random sample of households in Scotland, usually around 10,000. Other sources include the Department of Transport, Office of Rail Regulation and Department of Business, Energy and Industrial Strategy.

Quality Assurance undertaken by data supplier

The data are subject to various checks which are important to maintain the high quality of the data. The main data file is subject to checks and editing involving range checks, simple logic checks and complex logic checks. The data then undergo two additional processes. Firstly, the calculation of derived variables such as the age and sex of the Highest Income Householder and secondly, the imputation of household income.

The physical survey forms are also subject to robust checks by being uploaded into the physical survey validation system. Surveyors are then shown a list of all the errors picked up by the validation program. Additionally, they are shown a list of all the entered data, with a description of the variable next to each bit of data, and with the data split into representations of each page of the form. The validation system shows the data and the failed edits as well as showing the photographs of the property. Corrections are then made and each form rechecked until it passes all edits.

A technical report is available providing further information on the quality assurance arrangements for data used in this publication which is part of the SHS publication.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.

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

- Strong and varied response rate across Scotland is key for the quality of the data.
- The weighting strategy used by the survey minimises the extent of bias.

Limitations of the source

- Geographical limitations below local authority data
- The survey only produce estimates and these estimates are limited by a number of factors. The factors are sample coverage; sampling variability, the number of cases that analysis is based on and the bias in the achieved sample
- During 2020 and 2021 the survey was carried out by phone due to the Covid-19 pandemic.
- Statistics on travel to place of work or study will be different from previous years due to working or studying from home being a much more common practice since the Covid-19 pandemic began.

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Medium" as the 2021 TATIS publication was classified as Experimental Statistics rather than National Statistics due to the impact of Covid-19. To note, previous years' publications were designated as National Statistics, which means the data meet the quality standards set by the Code of Practice for Statistics.

2.12 Vital Events

Data Supplier	National Records of Scotland (Vital Events)
Supplier information	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	Unit records
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:	ABPE use: Five years until census day (20th March 2022) Estimation: From July 2020 until census day (20th March 2022)
Use of Data:	Production of an Administrative Data Based Population Estimate (ABPE) used for validation of population estimates for Scotland's Census 2022 (five years of data). Production of a linked administrative dataset used to: supplement the Census Coverage Survey (CCS) records in estimation, assist placement of records in adjustment, and for quality assurance of communal establishment

	population estimates during data processing (data from July 2020 to March 2022). Bias adjustment of local authority population estimates

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. 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.

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;
https://www.nrscotland.gov.uk/files//statistics/vital-events/checking-quality-nrs-statistical-data-on-ve.pdf

Quality Assurance undertaken by the census teams 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.

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 Admin 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 outwith 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

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 "High" because these data are being used in the validation of population estimates and production of a linked administrative dataset used to supplement CCS records in estimation. These census outputs will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low", because Vital Events datasets have achieved an enhanced/comprehensive assurance standards. The legal requirement to register these vital events, combined with the robust processes set up for collection and quality assurance of this data means that the risk level is low.

2.13 Pupil Census

Data Supplier	Scottish Government - Education Analytical Services
Supplier information	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 (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:	2021/2022 school session.
Use of Data:	Validation of population estimates. Production of an Administrative Data Based Population Estimate (ABPE) used for validation of population estimates for Scotland's Census 2022. Production of a linked administrative dataset used to: supplement CCS records in estimation, assist placement of records in adjustment, and for quality assurance of communal establishment population estimates during data processing. Bias adjustment of local authority population estimates

Source 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 a National 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 the census teams 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

- 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.

Limitations of the 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 Matrix

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 "High" because these data are being used in the validation of population estimates and production of a linked administrative dataset used to supplement CCS records in estimation. These census outputs will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concern has been set to "Low", due to these data already having obtained National Statistics accreditation, with good communication between suppliers and evidence of robust quality assurance in place.

2.14 Prisons

Data Supplier	Scottish Government: Justice Analytical Services (JAS)
Supplier information	JAS provides data on prisons through administrative data that captures characteristics of the prison population in Scotland. This QAAD refers to the data extract to be extracted on 20th March 2022 to cover the Scottish prison population as on census day (20th March 2022). The Prisons dataset is usually used to create 'Scottish prison population', an annual publication characterising imprisoned people in Scotland.
Data type (counts or unit records)	Counts
Data Content:	The publication provides analysis on the make-up of Scottish prisons (age, sex, ethnicity) as well as information on sentence length and the changing average size of the prison population.
Time period covered:	Data extract at Census day (20th March 2022)
Use of Data:	To assist communal establishment population estimates during data processing. Validation of population estimates.

Source information

Justice Analytical Services provided a snapshot of the Scottish prison population on census day (20th March 2022). JAS produce statistics on the prison population using administrative data from the Scottish Prison Service (SPS) management information system, PR2. These data are collected by prison officers and staff for operational intelligence and logistical support.

The prison population is not static: individuals move in and out of the population every day. The SPS management information system is live, updating on a daily basis to capture movements (e.g. to/from remand) within the prison population. An extract of data from this system reports on the population at the moment in time the extract is drawn.

The annual Scottish Prison Population publication is an Experimental Statistics publication that was first published with data for the prison population of 2019-2020. It was the first publication since 2015 when server updates within the Scottish Government meant it was no longer possible to generate datasets by combining and linking daily extracts from the live Scottish Prison Service operational database.

The publication presents new data on the Scottish prison population up to 2019-20. It explores population levels and composition and how these have changed over time, with a particular focus on the last 10 years. Data are provided on annual average population levels, the changing nature of the movements of the prison population and the changing characteristics of that population. Additional analyses of where the prison population are drawn from are also included.

Data supply and communication

Data are extracted from PR2 systems by a JAS analyst and transfer using an encrypted thumb drive to the workspace. Although care is taken when processing the data, the detail collected is subject to the inaccuracies inherent in any large scale recording system. While the figures are the most accurate available at the time of receipt, they may be subject to minor revisions due to updating and on-going quality assurance.

Quality Assurance undertaken by data supplier

Data on the prison population is derived from a single source which is also used to manage the prison population on a daily basis. This means that the data held on the PR2 system is routinely quality assured to rigorous standards and any quality issues are identified early on in the process to ensure accuracy and consistency as far as possible.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking that variables are in the expected formats and values
- Checking all the data are there

- Comparing the data with similar published data taking into account seasonally of prisons, change of age calculation and guidance of who should be recorded as usually resident in a prison for census enumeration purposes.
- Checking the age distribution of the population.

Strengths of the source

- Data are extracted from a 'live' management system so are expected to be up to date and robustly reflect the prison population on census day
- Date of birth information for each prison should mean that age distribution of the prison population will be robust

Limitations of the source

- Some risk that characteristics of the prison population (e.g. ethnicity) will not be perfectly recorded and/or will be missing for minority ethnic groups
- Some risk that prisoners will not be classified in the Prisons dataset in a comparable way to the census classification of 'usual resident.'
- Possible issue with male/female population counts for prisoners identified as 'other.'

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to "Low" given the requirement for prisoners to be registered on the PR2 system for management of prison populations on a daily basis, the routine quality assurance of the database and the strong links between the Scottish Prison Service and JAS.

2.15 Scottish Survey Core Questions

Data Supplier	Scottish Government
Supplier information	The Scottish Government has a range of responsibilities including: the economy, education, health, justice, rural affairs, housing, environment, equal opportunities, consumer advocacy and advice, transport and taxation. The Scottish Survey Core Questions (SSCQ) gathers survey responses from harmonised questions in the Scottish Crime and Justice Survey (SCJS), the Scottish Health Survey(SHeS) and the Scottish Household Survey (SHS) into one output.
Data type (counts or unit records)	Counts
Data Content:	Scottish Surveys Core Questions is the result of harmonised questions asked across three major Scottish Government household surveys. The pooled sampling method results in an annual sample size of approximately 20,000. It is possible to request data to be aggregated over several years to obtain a larger sample size. This may be helpful when validating small population groups. SSCQ produces detailed information on the compositions, characteristics and attitudes of Scottish households and adults across a number of topic areas including equality characteristics, housing, employment, and perceptions of health and crime. Data are produced and available at national and subnational geographies. It should be noted that there is likely a trade-off between the level of geography and the detail of the data available. The relevant SSCQ data for our purposes are: • Household and family relationships • Household tenure • Household landlord • Household cars or vans • Sexual orientation • Marital or civil partnership status • Country of birth • General health • Long term health problem or disability • Ethnic group • Religion • Unpaid care • Qualifications held
Time period covered::	Survey results from 2019 and a multi-year data extract covering 2016-2019.

Use of Data:	Validation of population estimates

Source information

SSCQ is an annual Official Statistics publication for Scotland. SSCQ reports on indicators such as: self-assessed general health, limiting long-term conditions, smoking and unpaid caring (released alongside the Scottish Health Survey), perception of crime in the local area (released alongside the Scottish Crime and Justice Survey) and satisfaction with local government services (released alongside the Scottish Household Survey). Detailed population characteristics are also available from the SSCQ data.

The pooling of Core Questions results in an annual sample of around 20,000 respondents, providing unprecedented precision of estimates at national level. This sample size enables the detailed and reliable analysis of national indicators by protected characteristics such as ethnic group, religion, country of birth, sexual orientation, age and gender. Further variables are education level, economic activity, tenure, car access and household type. Multi-level analysis is available on request. SSCQ also enables a more detailed analysis of sub-national geographies than source surveys allow. Annually, SSCQ reports for Local Authorities, Health Boards and Police Divisions.

Due to the different sampling nature of each survey, which is necessary to meet their primary aims, the number of respondents varies between different SSCQ questions, however the sample size is usually 20,000.

It should be noted that due to COVID-19, the SSCQ constituent surveys were paused and then moved to phone-based interviews, rather than face to face. The sample size was significantly lower during this period (e.g., Scottish Household Survey ~3,000 respondents vs the usual ~10,000). For more information on the impacts of COVID-19 pandemic on the constituent surveys see 2.4 and 2.8 above.

Publication of SSCQ for 2021 has been delayed so we are making use of data published for 2019.

Data supply and communication

SSCQ gathers data from identical indicator questions in three other Scottish Government publications, the Scottish Crime and Justice Survey, the Scottish Health Survey and the Scottish Household Survey. Links between the teams responsible for each publication and the data they provide are strong which ensures the quality of the overall data is high.

The required data was extracted from the SSCQ dataset by either SSCQ statisticians or VoPE statisticians under the instruction of SSCQ. This is yet to be determined. The data was securely transferred to NRS Census secure area.

Quality Assurance undertaken by data supplier

SSCQ is an Official Statistics publication. Official and National Statistics are produced to high professional standards set out in the Code of Practice for Official Statistics. Both undergo regular quality assurance reviews to ensure that they meet customer needs and are produced free from any political interference.

Quality assurance undertaken by SSCQ includes:

- Estimates from the pooled SSCQ and its constituent surveys SCJS, SHeS and SHS are and assessed to determine agreement. This is done for certain characteristics.
- Information on statistical techniques and weighting is published here: Scottish Surveys Core Questions 2019 - gov.scot (www.gov.scot)

The SCJS, SHS and SHeS surveys also all conduct QA on the data collected.

Quality Assurance undertaken by the census teams within NRS

Once the data were received, a number of data consistency and validation checks were performed, including:

- Checking the dimensions of the dataset received
- Checking that all requested variables are present and in the expected formats and values
- Comparing the data received with published data and investigating when there appear to be significant changes.

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

- Ability to provide more detailed statistics for smaller geographies than the source surveys.
- The publication responses have a relatively large sample size of circa 20,000
- Only source of robust comparator data for some census questions, e.g., adult ethnicity.

Limitations of the source

- Relies completely on the other three publications' data.
- Although sample size is greater than in each source survey, data on small groups (e.g. minority ethnic groups) can still be unreliable or unusable due to small sample sizes and/or disclosure control.
- Surveys were disrupted and sample sizes reduced due to COVID-19, the impacts of which are not completely clear yet. When 2021 data become available, we will investigate using them; otherwise we will make use of the 2019 dataset.

Risk Matrix

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 "High" because these data are being used to validate population estimates for Scotland's Census 2022, estimates that will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to 'Medium' as the 2019 and 2016-2019 are likely to have lost some comparability to our snapshot data from census day 2022. However, the publication is designated as an Official Statistic which means the data meet the standards set by the Code of Practice for Statistics. Quality assurance of the data from the three main source surveys is comprehensive.

2.16 National Health Service Central Register (NHSCR)

Data Supplier	National Records of Scotland (NRS)
Supplier information	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:	Data extract at Census day (20th March 2022)
Use of Data:	Validation of population estimates. Production of an Administrative Data Based Population Estimate (ABPE) used for validation of population estimates for Scotland's Census 2022. Production of a linked administrative dataset used to: supplement CCS records in estimation, assist placement of records in adjustment, and for quality assurance of communal establishment population estimates during data processing. To assist the census cleansing processes: Remove False Persons, Resolve Multiple Responses, Edit and Imputation, and Overcount Correction.

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 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 DSA allows for additional ad hoc cuts to be requested, as required here to coincide with Census day in 2022.

The data are sent to the Census Administrative 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[2] 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 census teams 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.

Strengths of the 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 source

- It does not pick-up people who leave the UK (unless they informed their GP) leading to some inflation in the register.

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

- 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[3].
- 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.

[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

Risk Matrix

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 "High" because these data are being used in the validation of population estimates, production of a linked administrative dataset used to supplement CCS records in estimation and improvement of the quality and accuracy of Scotland's Census outputs. These census outputs will influence policy decisions in many areas including health, housing, education and employment. employment.

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 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.
- 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

2.17 Higher Education Statistics Agency Student Record

Data Supplier	Higher Education Statistics Agency (HESA), part of Jisc.
Supplier information	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: • First name • Middle name • Last name • Previous surname • Gender • 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 2021/22 academic year
Use of Data:	Validation of population estimates. Production of an Administrative Data Based Population Estimate (ABPE) used for validation of population estimates for Scotland's Census 2022. Production of a linked administrative dataset used to: supplement CCS records in estimation, assist placement of records in adjustment, and for quality assurance of communal establishment population estimates during data processing. Quality assurance of communal establishment population estimates during data processing.

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 National Statistics publication 'Higher Education Student Statistics: UK', links to the latest published data are here:

Higher Education Student Statistics: UK, 2021/22 | HESA
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 institutions 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 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 institutions (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 2021/2022 this information is found at:
HESA Collections | HESAStudent 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.

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

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 census teams 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.
- 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)

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

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

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 value of "High" because these are being used in the validation of population estimates, production of a linked administrative dataset used to supplement CCS records in estimation and improvement of the quality and accuracy of Scotland's Census outputs that will influence policy decisions in many areas including health, housing, education and employment.

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.

2.18 Electoral Register

Data Supplier	Electoral Registration Officers in Scotland
Supplier information	The Electoral Registration Officer (ERO) is an official appointed by the local authority to prepare and maintain the Register of Electors.
Data type (counts or unit records)	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) • Polling District • 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:	The full electoral register from December 2021 was used as the time period closest to census day.
Use of Data:	Production of the census address list. Production of the Alternative Household Estimate. Production of an Administrative Data Based Population Estimate (ABPE) used for validation of population estimates for Scotland's Census 2022. Production of a linked administrative dataset used to: supplement CCS records in estimation, assist placement of records in adjustment, and for quality assurance of communal establishment population estimates during data processing. Validation of population estimates.

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[6], 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 (between 1 July and 1 December). 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[7] 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, 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 Census Administrative 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 2021. Forms were issued to each household, requesting details of eligible residents. The

[6] 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
[7] Representation of the People Regulations (Scotland) 2001

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 identity – to verify someone's identity 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 Census Administrative Data team receive the data in separate files according to electoral wards within each ER area.

Once the Census Administrative 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 data source

- A large proportion of the adult population in Scotland will be included in the data. The Electoral Commission estimate of completeness and accuracy in December 2022 was: 81% for completeness and 88% for accuracy.
- 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 were published at 1 December while census day is 20 March 2022. This time difference may lead to a mismatch in where some individuals are registered.
- 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 Matrix

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 "High" because these data were used in the creation of a census address list, validation of population estimates, production of a linked administrative dataset used to supplement CCS records in estimation and improvement of the quality and accuracy of Scotland's Census outputs. These census outputs will influence policy decisions in many areas including health, housing, education and employment.

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.

2.19 Health Activity

Data Supplier	Public Health Scotland (PHS)
Supplier information	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 provide the data for Primary health care separately from Secondary health care. Each wave of data comprises a confidential data file, a payload data file and a Key file to facilitate linking.
Data type	Unit records
Data content:	Two sets of data were received from PHS, one with interactions including COVID 19 vaccinations and one where the interaction dates did not include vaccinations. The dataset including COVID 19 vaccination data was used in creating the health spine for CCS estimation. The other dataset, without COVID 19 vaccination data, was used to make Admin based population estimate. The following variables are included in the confidential files at an individual record level for both 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 Additionally, the payload files for both primary care and secondary care have the following data variables: • Marital Status • Ethnic Group • Date of last interaction (Day, Month and Year). • Transfers Out Status. These are the patients who are no longer resident within Scotland on the final date of the extract. • Random Identifier which has a one-to-one mapping with CHI numbers and is consistent over time.

Time Period Covered:	Data extract at 20 March 2022, with 'Last Interaction' variable indicating an individual's interaction with selected NHS services over the previous 3 years
Use of Data:	Production of an Administrative Data Based Population Estimate (ABPE) used for validation of population estimates for Scotland's Census 2022 (dataset not including COVID 19 vaccination interactions). Production of a linked administrative dataset used to: supplement CCS records in estimation, assist placement of records in adjustment, and for quality assurance of communal establishment population estimates during data processing (dataset including COVID 19 vaccination interactions).

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 selected NHS services, providing an up-to-date population register that can help confirm population estimates for Census.

Data supply and communication

Under the terms of a data sharing agreement, this data was provided to NRS as an ad-hoc transfer, specific to census day.
The data is transferred from PHS to NRS using a secure file transfer system.

Quality Assurance undertaken by data supplier

PHS perform internal quality assurance processes before sharing data. Data are subjected to checks by the NHS board where the data were recorded and by PHS once the data are received. Examples of validation carried out include: 'postcode exists', 'consultant worked in the location or specialty at the time of admission', 'age and sex at admission are consistent with diagnosis'. Any data errors (missing or invalid information) or queries (information that appears implausible) are sent back to the NHS board for further investigation. General data management includes checks on *completeness* and *timeliness*, with *dataset specific* checks as set out in their publication on Data Quality Assurance.

- *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 targets have generally been achieved with a national return of 97% to 100%. The majority of Health Boards sit around the averages

although there are some outliers, for example only 89% completion rate in Dumfries and Galloway (SMROO New) and 54% in Orkney (SMR02 Deliveries). For whatever reason, Orkney stands out as having a poor return rate for maternity deliveries (SMR02), as low as 50% in the 2022 calendar year.

- *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.

- 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

  SMR06 Cancer Registrations

  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 – Dental data are collected through GP17 forms submitted by dentists to claim payments for dental services provided. PHS only provide treatment contacts in the more recent data extracts; prior to 2019, where no treatment claims had been made for a patient, registration data was provided as a proxy for contact.
    A noticeable drop in contacts will be evidenced as a result of COVID-19 measures, where dental treatment was restricted to emergency care only. Public Dental Service (PDS) 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 of 98%. Outputs and metadata for contractor activity is published on 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 census teams within NRS

Once the Census Administrative Data team receive the data, a number of data consistency and validation checks are performed on Health Activity datasets. 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, 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 population estimates.

Strengths of the source

- Health Activity dataset is a comprehensive source of record level data that covers the vast majority of the population in Scotland.
- 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.
- The Secondary care data set has good coverage for ethnic group, with 94% completeness.

Limitations of the source

- 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[8].
- Due to the number of datasets being used to create this dataset there may be a small percentage that are not linked correctly.
- The Secondary dataset is generally more complete than the Primary dataset for both identifiable and payload variables.
- Indication of instances where 1 January has been used as a default value for DoB, potentially inflating births on that date.
- A small number (around 50) of records have clearly incorrect date of birth. All these records have the exact same date of birth in the 16th century, suggesting that the issue only affects these particular records. It should be noted that none of these records are used to supplement the CCS.

[8] 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

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 Census Administrative 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 "High" because these data were used in the production of a linked administrative dataset used to supplement CCS records in estimation and improvement of the quality and accuracy of Scotland's Census outputs. These census outputs will influence policy decisions in many areas including health, housing, education and employment. Employment.

The Risk of quality concerns has been set to "Low" 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.
- Although the Health Activity dataset is dependent on multiple sources, there are various internal quality assurance measures in place.

2.20 Further Education Statistics (FES)

Data Supplier	Scottish Funding Council (SFC)
Supplier information	The SFC is a Non Departmental Public Body of the Scottish Government. The SFC invests around £1.8 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:	2021/22 academic year
Use of Data:	Production of an Administrative Data Based Population Estimate (ABPE) used for validation of population estimates for Scotland's Census 2022. Validation of population estimates.

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 (SDS) programmes as long as the student has attended at least once. 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 Census Administrative 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[9]. 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 National Statistics publication 'College Performance Indicators', the link below is for the 2021/22 version of this publication:

College Performance Indicators 2021-22 (sfc.ac.uk)

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

[9] Guidance Notes for FES 2 2021/2022 can be found at: https://www.sfc.ac.uk/publications-statistics/guidance/2021/SFCGD032021.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 census teams within NRS

Once the Census Administrative 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 National Statistics publication.
- Contains all the variables used when linking to other datasets.

Limitations of the data source

- There is a lag in being able to receive the data.

- 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 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 Census Administrative 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 "High" because these are being used in the validation of population estimates, that will influence policy decisions in many areas including health, housing, education and employment.

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.

2.21 Mental health inpatient activity

Data Supplier	Public Health Scotland
Supplier information	PHS was established in 2020 as Scotland's new national organisation for public health. It incorporates Information Services Division, NHS Health Scotland and Health Protection Scotland. The Mental Health Inpatient Activity dataset presents information on inpatient and day case care where a mental health diagnosis was recorded in psychiatric and/or non-psychiatric (general/acute) specialties in Scottish inpatient facilities. The dataset contains a summary of this information, drawn from hospital administrative systems.
Data type (counts or unit records)	Counts
Data Content:	• The number of admissions, discharges, stays, patients and hospital residents by NHS Board of treatment, hospital and treatment specialty, • Numbers and rates of discharges, patients and hospital residents by NHS Board of treatment, diagnosis grouping and treatment specialty, • Rates of discharges and patients by council area of residence and treatment specialty, • The number of discharges at Scotland level by discharge type, • The number of discharges at Health Board of Treatment level by discharge type group, • Numbers and rates of discharges and patients by NHS Board of residence, sex, age band and treatment specialty, • Numbers and rates of discharges, patients and hospital residents by NHS Board of residence, SIMD quintile and treatment specialty, • The Relative Index of Inequality for discharges, patients and hospital residents by treatment specialty, Scotland-level only, • The number of patients treated outwith their NHS Board of residence by NHS Board of residence and treatment specialty, • Percentage readmissions within 28 and 133 days by NHS Board of treatment for psychiatric specialties G1 and G4 only, • The number of stays by NHS Board of treatment, length of stay category and treatment specialty.
Time period covered:	1st April 2021 to 31st March 2022

Use of Data:	Validation of population estimates. Quality assurance of communal establishment population estimates during data processing.

Source information

The dataset contains a summary of mental health inpatient information for Scotland in both psychiatric hospitals and non-psychiatric hospitals, drawn from hospital administrative systems. A Scottish Morbidity Record 01 (SMR01) should be submitted for every episode of inpatient care in a general/acute hospital in NHS Scotland. A Scottish Morbidity Record 04 (SMR04) should be submitted for every episode of inpatient care in a mental health specialty in a psychiatric hospital or unit, or in a facility treating people with learning disabilities, in NHS Scotland. In addition, if the NHS contracts out psychiatric or learning disability care to a private care home or hospital, an SMR04 record should be generated for each episode.

When the SMR04 scheme was originally developed, inpatient stays in mental health specialties were often quite lengthy – perhaps several years – and this is still sometimes the case. To enable up-to-date monitoring of activity, the SMR04 record was developed with two elements: one completed and submitted on admission; and one completed and submitted on discharge. This was to ensure that it was not necessary to wait until discharge for information on the admission to become available (but if the stay is short, the admission and discharge information can be submitted together).
We use this dataset to obtain a snapshot of the count of individuals residing in a mental health establishment on census day.
A hospital resident is defined in this dataset as an individual who is an inpatient in a psychiatric facility at midnight on 31 March, before the beginning of the next financial year (1 April). Although the term 'hospital resident' is used for simplicity, some people will be looked after in care homes (under contract to the local NHS Board). This measure is not calculated for SMR01 information as, typically, patients do not experience long term stays in acute hospitals and, therefore, would not be thought of as 'residents'.

Data supply and communication

The data were downloaded from the PHS website.

Quality Assurance undertaken by data supplier

The data are sourced from hospital administrative systems across Scotland and extracted from the Scottish Morbidity Record (SMR) 01 and 04 datasets held by PHS. The Scottish Government target for SMR submission to PHS is 6 weeks following a hospital admission or discharge, or transfer/death/clinic attendance. For example, all SMR records with a March 2019 date of discharge would be expected to be submitted to PHS by 12th May 2019.

The SMR01 and SMR04 data are processed and monitored by the Data Support & Monitoring Team. The PHS Data Quality Assurance (DQA) team is responsible for ensuring that SMR datasets are accurate, consistent and comparable across time

and between sources. In order to assess the quality of recording, the DQA compares samples of records against source evidence in hospital systems and patient case notes. PHS have a 90% standard for accuracy and recording of data.

The Mental Health team who produce these data also carry out thorough quality assurance (QA) checks on the data after extraction from the databases, in accordance with the PHS checking guidance. For example, they compare figures for the same years between the current and previous publications in order to identify any large changes in the data sources. Additionally, they look at trends within the current publication in order to identify any unusual patterns. For changes or patterns in the data which cannot be explained by the known completeness estimates, the team contact the relevant board to highlight the issue and request an explanation.

Quality Assurance undertaken by the census teams within NRS

Once the Statistical Quality Assurance 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

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

- Designated National Statistics, complying with the Code of Practice for statistics
- Extensive quality assurance checks built into producing the dataset.

Limitations of the source

- In Scotland's Census 2022, individuals are considered usual residents in communal establishments such as mental health hospitals if they have been residing there for 6 months or more. This PHS dataset does not contain length of stay and we are considering the resident count at 31st March 2022 to reflect census day (20th March 2022) resident count with no adjustment for length of stay. Counts in this dataset are therefore likely higher than in the census dataset as all stays, however short, will be included.
- There may be lags in updating records so the resident counts in the publication may not reflect resident counts on census day.

Risk Matrix

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 “high” because these data are being used in the validation of population estimates and improvement of the quality and accuracy of Scotland's Census outputs. These census outputs will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concerns has been set to “low”, as there is a well-documented quality assurance process to ensure these data are of high data quality. The Mental Health Inpatient Activity is a National Statistics publication, so they have already been judged of sufficient quality for that accreditation.

2.22 Census of independent schools

Data Supplier	Scottish Government - Education Scotland
Supplier information	Education Scotland is a Scottish Government executive agency charged with supporting quality and improvement in Scottish education and thereby securing the delivery of better learning experiences and outcomes for Scottish learners of all ages. The Registrar of Independent Schools is an official of the Scottish Government, based in Education Scotland, responsible for maintaining a public register of independent schools which lists all registered independent schools in Scotland. The census of independent schools is a register of pupils attending non-publicly funded schools in Scotland. The information published is therefore a reflection of the information provided by school staff.
Data type (counts or unit records)	Counts
Data Content:	The census of independent schools covers all boarding and non-boarding non-publicly funded schools in Scotland and is collected at the level of the school by sex, boarding/not boarding and year of birth of the pupils.
Time period covered:	2021/2022 school session.
Use of Data:	Validation of population estimates. Quality assurance of communal establishment population estimates during data processing.

Source information

The census of independent schools is collated by the Scottish Government registrar of independent schools each year in September and reflects the number of pupils by year of birth, by sex, by boarding status and by school.

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 was provided to NRS by Education Scotland on an adhoc basis to facilitate census quality assurance.

Quality Assurance undertaken by data supplier

Comparisons are made with previous years' data to identify any potential issues and to gauge consistency or identify where additional scrutiny may be required.

Quality Assurance undertaken by the census teams within NRS

Once the Statistical Quality Assurance 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 age distribution of the population.

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

- These data are a comprehensive source covers the vast majority of the school aged population attending independent schools.

Limitations of the source

- The data are collected by school and no information is collected on the residence address of pupils attending each school. For example, it is not possible to tell what proportion of pupils attending schools in one local authority live in that local authority. This is in contrast to the SG Pupil Census for publicly-funded schools that does collect this information.
- Pupil counts are in year of birth rather than date of birth or age. We took the approach that if, e.g., 100 pupils had 2011 as their year of birth, on census day, 25% of these pupils would be 11 and 75% of these pupils would be 10. Because births are not uniformly distributed through the year, this simplification will introduce a small amount of error into the pupil age distribution.

Risk Matrix

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 "High" because these data are being used in the validation of population estimates and improvement of the quality and accuracy of Scotland's Census outputs. These census outputs will influence policy decisions in many areas including health, housing, education and employment.

The risk of quality concern has been as 'Medium' due to these data not being able to link pupils to their residence addresses.