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

# Statistical Quality Assurance Report – Rounded Population Estimates

**Contents**

## 1. Introduction

This quality information and methods report accompanies the first release of results from Scotland's Census 2022.

The information in this report will help you to:

- understand the strengths and limitations of the data
- understand quality considerations of the data
- understand the methods we used to create the data
- decide suitable uses for the data

This report provides information on:

- how census data was processed to produce the statistics
- the methods used
- the quality of the data

Key facts and figures about how the data was collected is available on our website. Information on the design of the census is also on our website.

Census statistics represent the total population rather than just those who completed the questionnaire. Statistical modelling has been used to produce total population estimates across the United Kingdom since the 2001 censuses.

The quality of census data is important. We made innovative changes to our approach in response to the return rate during the collection phase of Scotland's Census.

We conducted a Census Coverage Survey (CCS) immediately after the 2022 Census collection as we always intended. The results of this survey were used in our modelling. The key change we made was to supplement the CCS with administrative data to reduce uncertainty in our estimates of the total population.

### 1.1 Census Quality Summary

- The Office for Statistics Regulation has **designated census statistics as National Statistics** meaning they meet the requirements of the Code of Practice for Official Statistics.

- An International Steering Group (ISG) was established in response to the challenges faced during the collection phase and NRS has been working with these experts as we've been adapting our methods. The ISG made supportive statements at the end of collection, at the end of method development and also now that the methods have been implemented. Professor James Brown, chair said:

**'The International Steering Group commends NRS for its systematic and robust application of the census estimation methodology that has led to today's published estimates.'**

- In May 2023, the Registrar General wrote to the SG Chief Statistician to commission a review of the coherence assessment for first outputs. This included reviewing the quality assurance and validation processes ahead of first outputs. In August, the Chief Statistician confirmed by letter:

**"I am therefore satisfied… that the outputs will provide the robust population statistics to support the wider statistical system and that analysts and the general public can have confidence in the outputs, and that they can be published on 14 September 2023."**

- Before the census we published our Statistical Quality Assurance Strategy, which included targets for quality.
- A usual resident population response rate of 89.8% was achieved, lower than our target of 94% at national level, with 31 out of 32 councils above our target of 85% for each council area. Only Glasgow City was below 85% at 82.1%.
- The confidence interval targets for Scotland's 2022 Census estimates were +/- 0.4% at national level and +/- 3% at council area level for both people and households. At national level the population confidence interval achieved was +/- 0.51%, with 31 out of 32 council areas being within +/- 3%. East Lothian was slightly outside the interval at 3.07%.
- This means we have a population estimate for Scotland of 5,436,600 and a confidence interval of +/- 0.51% (27,900 people). That is the true population count is expected to be within 27,900 of the published estimate. In other words we can be confident that if we were to repeat the survey selection 100 times, 95 of the population estimates would be between 5,408,700 and 5,464,500.
- These population confidence intervals are similar to 2011. At the national level the confidence interval in 2011 was +/- 0.44% (+/- 23,000 people), with all council areas being within +/- 3%.
- We have carried out extensive quality assurance work to produce high quality census estimates, including sharing early estimates with council areas for quality assurance purposes, as allowed under the Code of Practice for Official Statistics.

### 1.2 Statistical Quality Assurance Strategy

In May 2021 we published our Statistical Quality Assurance Strategy, which provided an overview of how we will measure the quality of the census. The document outlines why statistical quality assurance is important for the 2022 census given the wide use of the data and the important decisions it informs. It also set out the following quality targets, which we have reported against in Table 1.

Table 1: Results against the quality targets.

| How will we achieve high quality results? | How will we measure success? (Key Performance Indicators (KPIs) and acceptance levels) | Result |
|---|---|---|
| We will maximise our overall person response rate | Person response rate of at least 94%. | Person response rate of 89.8% |
| We will ensure a minimum level of response within every local authority in Scotland | Person response rate in every council area of at least 85%. | 31 of 32 council areas with a response rate of at least 85%. Only Glasgow City was lower at 82.1%. More information on response rates is included in Section 4.3. |
| We will maximise the accuracy of our national population estimates | Variability: national estimates will achieve 95% Confidence Intervals (CI) +/- 0.4%; Bias[1]: < 0.5% | National confidence interval for persons of +/- 0.51% and bias of 0.15%. |
| We will maximise the accuracy of our local authority population estimates | Variability: Council Area estimates will achieve 95% CI +/- 3% | 31 of 32 local authorities achieved a 95% confidence interval within +/- 3%. Only East Lothian was slightly outside this at +/- 3.07%. More information on variability is included in Section 4.2. |

[1] Bias indicates a potential systematic difference between our estimate and the true population. Bias due to the sampling of the CCS has been measured by calculating the difference between the national estimate and the average estimate from the bootstrap resamples used in calculating variance. There are other sources of bias, such as dependence between the census and CCS or overcoverage within the census, which we have corrected for within our estimates. Availability of definitive comparator sources makes it very difficult to estimate any other residual bias in the estimates. We will publish further work on the quality of the statistics in later releases.

| | | |
|---|---|---|
| We will minimise the non-response to all mandatory questions | Achieve or exceed target non-response rates for all mandatory questions | Non-response rates for age and sex published in Section 2.3. More information on non-response for mandatory questions will be released alongside the release of data for each question. |
| We will assess all national and local authority level results for each main release by means of quality assurance panel | Undertake quality assurance panel and resolve any issues present | We went a step further than we intended by providing all councils with the opportunity to highlight inconsistencies before publication. More information is included in Section 6.4. |
| We will publish details of methods and full details of all our data quality indicators | Publish on our website | Information on methods is published on our website. More information will continue to be published following first release. |
| We will publish the results of an independent methodology review | Publish on our website | Published in the Peer review and governance section of our website. |
| We will maintain our National Statistics Accreditation | Accreditation maintained throughout | Accreditation maintained. More information is included in Section 1.3. |

## 1.3 National Statistics Status

The statistics being released have been assessed as National Statistics by the Office for Statistics Regulation (OSR). This is the result of a rigorous assessment of the quality, trustworthiness and value of these statistics and the compliance of National Records of Scotland (NRS) with the Code of Practice for Official Statistics. More information about the assessment can be found on the OSR website.

## 2. Assurance of Processes

### 2.1 Background

The census data has been through a number of statistical processes that form the overall census data journey. Ahead of the census, our methodology for data processing was approved by our External Methodology Assurance Panel. More information on this process and panel is available on our website.

Since the end of collection the International Steering Group (ISG) has provided clear advice as we adapted our methods to account for the lower than expected return rate. More information about ISG is on our website.

We have performed quality assurance at each stage within the census data journey. Standard checks were repeated at each stage of processing to check that no data had been removed unintentionally during processing and no unexpected changes had been made. In addition to this, bespoke checks were performed for each process. These checks were performed on census processes and also their Census Coverage Survey equivalents where applicable. More information on what we planned for quality assuring census statistical data processing is available on our website.

### 2.2 Data Processing

#### 2.2.1 Capture and Coding

In order to process the data it needs to be captured from the questionnaires and then transformed into a digital dataset suitable for further processing. This step is referred to as capture and coding. The capture process includes both paper and online questionnaires, and the coding of responses into values in a dataset happens either automatically, or manually if the automatic process is unable to assign a valid code.

Quality assurance checks completed during the coding process included:

- Manual checks of a sample of images from paper forms against captured and coded data.
- Manual sample checks on all coded returns, carried out for each census question.
- Analysis of the distribution of responses to all questions to spot any values having an unusually high/low number of responses.
- Comparison to other data sources for selected questions. Selected questions were: population by sex, age, household size, ethnic group, student population and accommodation type.
- Manual checks of specific codes, which are known to be likely to be given incorrect codes at automatic coding of paper responses. These are commonly responses to text questions.
- Sample checks of data that had been manually coded.
- Quality assurance of coding performed by manual coding operators by sending them responses that were auto-coded, then comparing the auto and manually coded values.

- Ad-hoc checks, for example a comparison of item non-response to that of the Office for National Statistics (ONS), Northern Ireland Statistics and Research Agency (NISRA) and Scotland's Census 2011.

Following the quality assurance of the capture and coding process, we re-coded some responses during our internal coding process. This fixed some issues identified during quality assurance, and assigned codes to some responses that were not coded automatically or manually.

### 2.2.2 Data Cleansing

After the data was coded, a number of steps were undertaken to 'clean' the data. The main steps were:

- Name Re-Ordering (NRO) — this applies only to paper questionnaires. It makes sure the names on a household form match the names on individual forms.
- Remove false persons (RFP) — RFP looks for possible false person records in the census data and removes them if they meet certain criteria.
- Resolve Multiple Responses (RMR) — This process finds and merges duplicate records. These could be records of households, people or communal establishments.
- Filter Rules — these apply only to paper questionnaires. It involves fixing problems with the answering path a person takes through the questionnaire.

More information on the methodology for these processes is available on the Scotland's Census website.

Quality assurance checks performed during data cleansing included:

- Clerical review during the NRO, RFP and RMR processes to manually decide whether to re-order, remove or resolve records respectively that were not easy to classify automatically.
- Quality assurance of each clerical review stage, double checking a sample of decisions made by each individual involved in the review.
- Checking aggregate statistics on the number/proportion of records changed during each process and comparing this to what was expected, for example what proportion of records were changed during the same process in 2011.
- Use of administrative data to help decide whether a record was a false person during RFP and whether a record was a multiple return during RMR.
- Checking a sample of records altered by each process (for example a sample of records that were re-ordered during NRO).
- Ensuring every individual had a matching household or communal establishment record and ensuring the address information was consistent between the records. Where a matching record was not found, a household or communal establishment record was created.

### 2.2.3 Edit and Imputation

Edit and Imputation (E&I) checks and corrects inconsistencies and fills in gaps in census data. We use editing and imputation where people have not answered individual questions or where there are inconsistencies between answers.

More information on the methodology for E&I is available on the Scotland's Census website.

Quality assurance checks performed during E&I included:

- Checking item non-response and imputation rates for each variable.
- Comparing the distribution of each variable before and after imputation, including checking for outliers in the distributions.
- Checking the number of donors used during imputation and the geographic distance between donors and imputed records.
- Applying edit rules a second time outside of the E&I process to independently check if they had been applied correctly.
- Separately checking that communal establishment records had been imputed successfully. For example, checking the age/sex distribution of individuals in student communal establishments before and after imputation.

## 2.3 Item Missingness for Age and Sex

In the first release of census information the population estimates have been published by age group and sex for each council area. Table 2 shows the item non-response and imputation rates for age and sex. The item non-response rate is the number of records where a question was not answered divided by the total number of records that were asked that question. The imputation rate shows the proportion of records that had either age or sex (or both) imputed. This can be slightly higher than the item non-response rate if a record failed an edit rule during E&I. Part of the reason for item non response was due to the design of the census, for example, how we used forms that had not been fully completed from our online questionnaire.

Table 2: Item non-response and imputation rates for age and sex.

| Variable | Item non-response rate (%) | Imputation rate (%) |
|---|---|---|
| Age | 1.3 | 1.4 |
| Sex | 1.8 | 1.8 |

### 2.3.1 Age

Age is calculated from the date of birth question asked on the census form. Where this information was missing, we matched the record to administrative data using other information that was present, for example name, sex and address. Where a match was found, we used the age from the administrative record to guide the imputation of age during the E&I process. The age that was imputed was still done so from a census donor record, however, the administrative age was one of the variables used by the algorithm to select a donor record. This made the imputed age be more likely to be similar to the administrative age, improving the quality of the imputation of this variable. No administrative data was added to the census dataset.

### 2.3.2 Sex

#### 2.3.2.1 Remove False Persons (RFP)

During the RFP process, the 2 of 7 rule finds and removes false records from the census. It considers a person to be false if they do not answer at least 2 of 7 questions:

Name in the individual form
Name in the relationship matrix
Name in the household member listing
Date of birth
Relationship to others in household
Sex
Marital status

One of the 2 questions must be either the name or date of birth.

If a record only had either a name or date of birth, it was matched to administrative data to provide additional evidence that this was an actual person. If a match was found, the record would be kept and would not be considered a false person. This process enabled us to be confident about using responses where limited information was provided. However, it did result in higher non-response rates for some variables, including sex compared with previous censuses. The method used in 2011 would have removed these records as false persons and they would not be counted in the calculation for item non-response. This was deemed to be an acceptable trade off due to more records from genuine respondents being kept during the RFP process.

The method of matching on date of birth also means that records that were successfully matched to administrative data were more likely to have a missing value for sex than age. This is because date of birth (which is used to derive age) was used in the matching – records with this variable missing were less likely to be matched to administrative data; these records that failed to match were removed as a false person.

#### 2.3.2.2 Edit and Imputation

During the E&I process, records that had sex missing were first imputed using the first name of the person, where this was present. To do this a reference dataset was created from the census dataset with first names, sex and the amount of times this combination of sex and first name appeared in the dataset. Thresholds were then applied to ensure that only names that appeared at least a certain amount of times and had a specific sex ratio or higher were included in the reference dataset. Then the person's first name was matched to the name in this reference dataset and the sex was pulled across.

For example, if the person's first name was Olivia and their sex was missing, the person's sex was imputed as female as a certain amount of census respondents with the first name Olivia were female and Olivia appeared enough times in the dataset. For names with a more even split between male and female or more unusual names that were therefore not on the reference dataset, the imputation of sex would be the same as the imputation of other variables, by finding a donor that is as close as possible to the record with the missing sex.

## 3. Coverage Estimation

Census statistics represent the total population rather than just those who completed the questionnaire. Statistical modelling has been used to produce total population estimates across the United Kingdom since the 2001 censuses.

In response to the lower than expected return rate for the census, we adapted our methodologies to estimate those who did not respond.

As intended we conducted a Census Coverage Survey (CCS) immediately after the 2022 Census collection. The results of this survey were used in our modelling. The change we made was to supplement the CCS with administrative data to improve the accuracy of our estimates of the total population.

The coverage estimation process combines the results of the Census Coverage Survey (CCS) with administrative data to create a combined survey and administrative frame, which is used to identify and adjust for the number of people and households not counted, those counted more than once, and those counted in the wrong place. Firstly, we estimated undercoverage and overcoverage in the collected data by:

- linking the combined survey and administrative frame records with those from Census 2022 using automated and clerical linking
- using the linked census and combined survey and administrative frame within a Dual System Estimation technique to estimate the number of people and households missed by the census
- searching the Census 2022 database for duplicates and using the combined survey and administrative frame to estimate the level of overcount in the census
- estimating populations for each local authority by age, sex, ethnic group, economic activity and other important characteristics, balancing over-estimates and under-estimates, using a combination of statistical regression and small area estimation techniques
- checking and correcting for biases by evaluating against comparator data

More information on the new methods used for coverage estimation is also available in our Methodology Enhancements to Secure High Quality Census Outputs and Population Estimates report on our website. More detailed methods papers on coverage estimation and the use of administrative data will be published in due course.

### 3.1.1 International Steering Group (ISG)

At the direction of the ISG we performed a review of UK data sources to identify which would be most appropriate to meet the specific needs of Scotland's Census 2022.

Working with ISG, we identified several administrative data sources with the potential to help us in improving the statistics. We prioritised them as either essential or desirable for our ability to enhance the coverage estimation process (see tables 3 and 4). All sources deemed as essential were obtained. We also

obtained two desirable sources: council tax exemption data from all council areas and combined aggregate information from DWP and HMRC at council area level.

Table 3: Essential administrative data sources.

| Data source | Description |
|---|---|
| NHS Central Register | People who are or have been registered with a GP in Scotland, or whose birth was registered in Scotland |
| Health Activity | Patients who have interacted with selected NHS services in the previous 3 years |
| School Pupil Census | People enrolled in state funded schools in Scotland |
| Electoral Register | People registered to vote in Scotland |
| Higher Education Statistics | Higher education students studying or domiciled in Scotland |
| Vital events (births, deaths, marriages and civil partnerships) | Birth, death, marriage and civil partnership registrations |

Table 4: Desirable administrative data sources obtained.

| Data source | Description |
|---|---|
| Council tax | List of occupied or unoccupied properties in local authority (residential properties that have council tax exemptions or discounts because they are unoccupied) |
| DWP | Aggregated tables with people who have interacted with DWP or HMRC |
| HMRC | Aggregated tables with people who have interacted with DWP or HMRC |

More information on the ISG is available in Section 6.2.

## 3.2 Census Coverage Survey

It is necessary to impute whole missing persons or household records that were estimated to have been missed by Scotland's Census 2022. To support this adjustment of population and household estimates, the Census Coverage Survey (CCS) estimates the degree of under and over coverage.

The format of the 2022 CCS was broadly similar to the 2011 CCS. We conducted the CCS in the weeks following Census Day, sampling around 1.5% of postcodes in Scotland within council areas. We used a Hard-to-Count index to identify and include areas where people were less likely to complete the census form. The response rate was 57%, similar to the 61% achieved by ONS in their CCS. Participation in the CCS was voluntary. We then supplemented the CCS by adding carefully selected administrative data from CCS areas into the sample. The addition of the administrative data gave us greater confidence in the estimates given the lower than expected return rate.

## 3.3 Administrative Data

Administrative data is information created when people interact with public services, such as schools, the NHS, the courts or the benefits system, and collated by government.

These public bodies must keep records of these interactions for operational purposes: to enable them to carry out their day-to-day work, to monitor and improve their performance, and keep providing services in an effective way.

This information has been used alongside the Census Coverage Survey to help estimate how many people did not respond to the census. We used administrative data to enhance the CCS, no administrative data is used in the census dataset itself. We did this by linking together several administrative data sources and applying criteria to only include people who we are confident were usually resident in Scotland on Census day.

### 3.3.1 Data Governance and Confidentiality

We have put in place data sharing agreements with each of the organisations that have provided administrative data for the Census. The agreements set out the lawful basis for sharing administrative data with us and how we are allowed to use it. The data sharing agreements make sure we use data in the way people would expect based on the privacy information provided to them by each of the data providers.

No administrative data makes its way into the census dataset. People who did not respond to the census will still be treated as non-response in the census processing. The administrative data was used to supplement the CCS to help estimate how many people did not respond to the census.

NRS worked closely with partner organisations to secure a range of administrative data. All data obtained have Data Sharing Agreements and the use was agreed by the Public Benefit and Privacy Panels who ensured that all legal, privacy and ethical standards were met for the research. A Data Protection Impact Assessment[2] was also conducted.

## 3.4 Logistic Regression Model

We used an approach based on a mixed effects logistic regression to convert our collected returns to population estimates representative of 100% of the population. This is the same approach used by the Office for National Statistics (ONS) for its 2021 Census. The logistic regression approach uses a single national model to calculate record-level non-response weights, which are used to estimate the household and person populations. The model in effect predicts the likelihood of a person responding to the census accounting for age, sex, ethnicity, and other potential variables.

Our model selection process aimed to identify a model that explains coverage patterns using objective, data-driven methods.

[2] Data Protection Impact Assessment - Administrative Data for Estimation | Scotland's Census (scotlandscensus.gov.uk)

The key principles we followed were:

- Objectively explainable model structure using best practice techniques
- Include key demographic predictors
- Minimise prediction error
- Compare data groupings in parallel

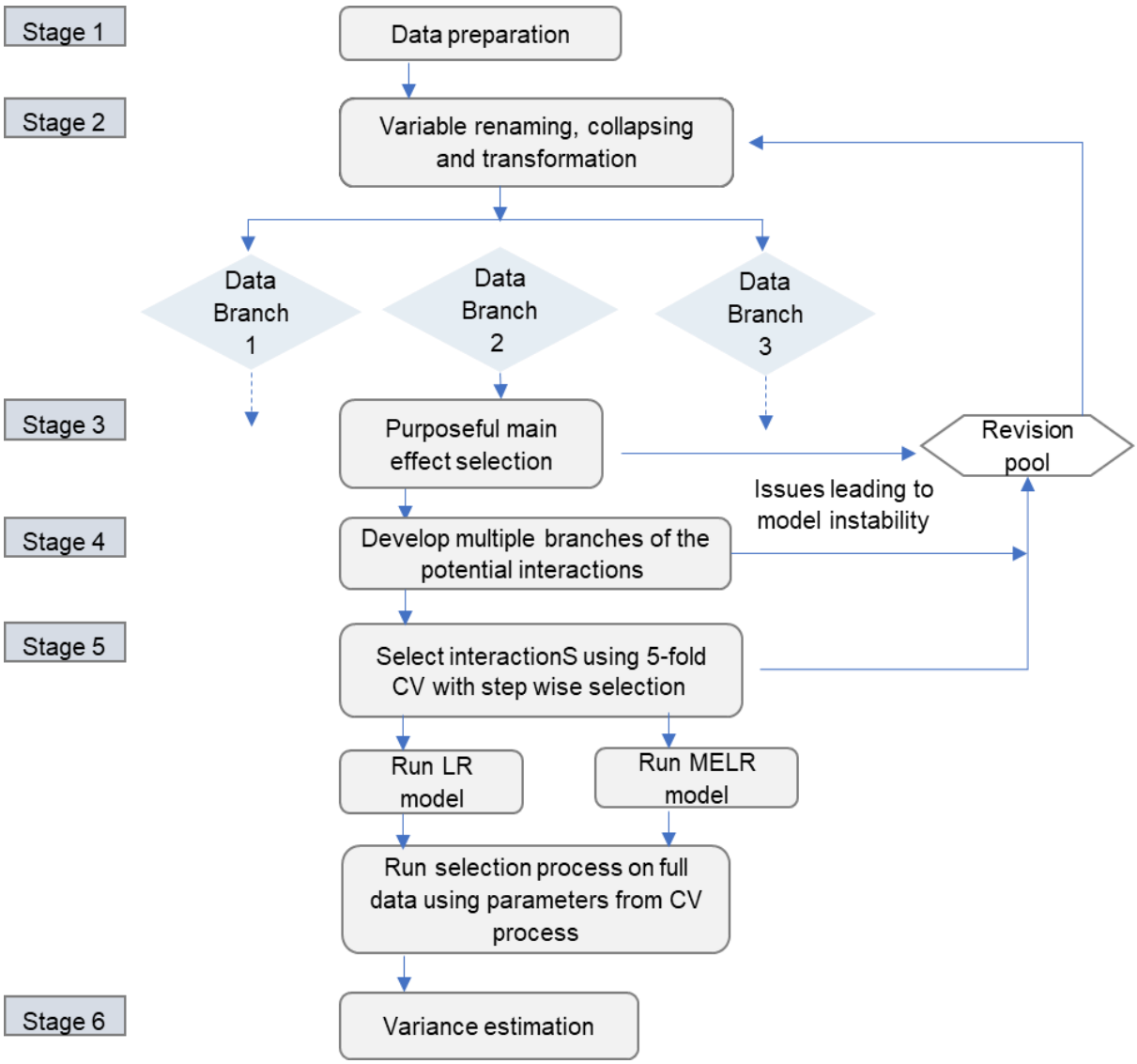
Alongside our key principles, we used the following techniques:

- Standard statistics, frequency tables and diagnostic tests (AIC, BIC, log likelihood) to compare models
- Stepwise selection to select interactions
- K-fold (fivefold) cross-validation to check prediction error

### 3.4.1 How the modelling was implemented

We followed six stages in our model selection process, as illustrated in Figure 1.

Figure 1: Flow chart of estimation process.



The six stages include:

Stage 1: Data Preparation: Filter and select the data, including selecting administrative data and combining it with the CCS

Stage 2: Collapsing & Initial Data Analysis: We analysed the variables, distributions and linkage rates for different levels, and created data branches with different variable groupings.

Stage 3: Purposeful Main Effects Selection: In this stage we fitted a basic model with just main effects to produce diagnostics. This allowed us to see which variables were contributing to the model and overall model fit to compare data branches. We also checked for any instability caused by our data groupings

Stage 4: Initial Interaction Analysis: After examining main effects we analysed each possible interaction separately to assess their contributions to existing main effects. We also checked for model instability at this stage, such as quasi-complete separation.

Stage 5: Cross-validation: We used stepwise selection to select interactions, while using k-fold cross validation check prediction error and protect against potential overfitting

Stage 6: Variance Estimation: We used bootstrapping to create confidence intervals around our estimates. In this stage we also checked for excessively large model weights.

More information on the modelling will be published in more detailed methods papers in due course.

### 3.5 Estimating Communal Establishments

We estimated the number of individuals living in communal establishments (CEs) using:

- administrative and survey data
- census communal establishment managers form
- CCS for small communal establishments

The coverage estimation modelling approach explained previously is suitable for people in households, however a different approach for CE estimation was undertaken. Our CE estimation strategy involved comparisons between census responses and administrative data sources. By reconciling differences between these sources, we calculated census under-coverage by age and sex for each individual establishment.

The Communal Establishment Register (CER), a NRS list of all known CEs in Scotland, was used as the basis for the coverage estimation process given that not every establishment responded to the census. If additional CEs were identified that were not recorded on the CER, extensive validation was undertaken to ensure that they were legitimate establishments.

Conversely, analysis was undertaken to establish if there were establishments listed on the CER where there was limited or no information from other sources. Each establishment type was considered separately, with key CE types receiving further investigation.

Across all CE types, the main data sources included the census manager forms and administrative data (further information on this is in Section 3.3). Both sources contained aggregate counts of usual residents by age–sex group; and the CER record contained a total usual resident count.

In addition to these three sources, there were further additional data sources specific to certain CE types, as in Table 5.

Table 5: Data sources used in estimation of specific communal establishment types.

| Establishment Type | Data Source |
|---|---|
| Care home with/out nursing | Public Health Scotland Care Home Census |
| Defence Establishment | Ministry of Defence |
| Prison and young offenders institutions | Scottish Prison Service |
| Halls of Residence/student accommodation | Higher Education Statistics Agency |
| Schools | Scottish Government Independent School's Census |
| Mental Health hospital (including inpatient units) | Public Health Scotland Mental Health Inpatient Publication |

We worked with ISG on the approach to prioritising sources.

### 3.6 Post-Modelling Adjustments

We have made adjustments to the data after modelling. These adjustments are similar to the adjustments made by the Office for National Statistics (ONS) for their 2021 census for households, babies and school ages.

Four adjustments were made after modelling:

- Household adjustment
- School aged children adjustment
- Babies adjustment
- Sex ratio adjustment for ages 15 to 18

The household and school age changes were informed by the quality assurance process we carried out with council areas (see section 5.4). We took the evidence provided by councils and looked for patterns across Scotland rather than making changes on an ad hoc basis for individual council areas.

#### 3.6.1 Household Adjustment

Statistical bias is caused by systematic issues, leading to unrepresentative estimates. An example is non-response bias: where a particular population group does not respond. We aim to produce unbiased estimates. This is the principle behind our design and why we both focused our collection strategy to ensure we got

responses from all areas and used CCS and administrative data to correct for non-response across specific demographic groups or regions.

Our use of the combined CCS and administrative data frame in estimation is underpinned by an assumption of statistical independence between it and the census frame. This means that the causes for non-response across the two frames should not be related, or simply put, the people who do not respond to the census should not be the same as the people who do not respond to the CCS or appear in the administrative data. We take steps in our design to ensure that these sources are independent. Despite this, however, we know that this assumption will not perfectly hold, and this will cause bias in our estimates. To deal with this bias, we perform checks to detect and correct for it. We created the Alternative Household Estimate (AHE) to help assess the level of bias and we also had available high quality household estimates based on council tax, which NRS publish on an annual basis. This meant that we had a choice in which source to use when correcting for bias. Assessing both sources, we decided to use the 2021 household estimates which were the latest available official statistics.

To perform these adjustments we compared our census household estimates in each council area to the comparator, the housing statistics estimates. Where the comparator falls outside the confidence intervals in our estimates this suggests bias is statistically significant. We then use the difference between the upper bound of our census estimate confidence interval and the comparator to correct for the bias in the household estimates. This correction is then fed through to the census person estimates, as any bias affecting households will also affect the people in them.

Adjustments were made in the council areas detailed in Table 6.

Table 6: Council areas that received a bias adjustment and the size of that adjustment for households and persons (%).

| Council area | Household adjustment (%) | Person adjustment (%) |
|---|---|---|
| East Renfrewshire | 5.1% | 3.2% |
| Inverclyde | 4.9% | 3.5% |
| Renfrewshire | 4.2% | 2.9% |
| East Dunbartonshire | 3.4% | 2.3% |
| Argyll and Bute | 1.3% | 0.9% |
| Angus | 1.1% | 0.7% |
| Orkney Islands | 0.8% | 0.5% |
| South Lanarkshire | 0.8% | 0.5% |
| West Dunbartonshire | 0.6% | 0.4% |
| Glasgow City | 0.5% | 0.3% |
| Na h-Eileanan Siar | 0.5% | 0.3% |
| Moray | 0.4% | 0.2% |
| West Lothian | 0.1% | 0.1% |
| **Scotland** | **0.5%** | **0.4%** |

### 3.6.2 School-Aged Children Adjustment

An adjustment was also made for children aged between 6 and 15 to calibrate the census estimate to data from the school pupil census where the census estimate for a single year of age and sex was lower than the school pupil census in each council area. This was only applied to ages 6 to 15 as the school pupil census does not cover a significant portion of the population at age 5 due to children deferring entry to primary school and at ages 16 and over due to pupils leaving school. The scale of this adjustment in each council area is shown in Table 7.

Table 7: Size of the adjustment made to children aged 6 to 15 years old by council area.

| Council Area | Percentage adjustment applied |
|---|---|
| Aberdeen City | 0.0% |
| Aberdeenshire | 0.0% |
| Angus | 0.6% |
| Argyll and Bute | 0.2% |
| City of Edinburgh | 0.0% |
| Clackmannanshire | 1.4% |
| Dumfries and Galloway | 2.4% |
| Dundee City | 0.4% |
| East Ayrshire | 1.5% |
| East Dunbartonshire | 0.1% |
| East Lothian | 0.0% |
| East Renfrewshire | 0.1% |
| Falkirk | 0.9% |
| Fife | 1.1% |
| Glasgow City | 1.8% |
| Highland | 0.1% |
| Inverclyde | 0.8% |
| Midlothian | 0.2% |
| Moray | 0.1% |
| Na h-Eileanan Siar | 2.8% |
| North Ayrshire | 2.1% |
| North Lanarkshire | 1.2% |
| Orkney Islands | 0.5% |
| Perth and Kinross | 0.0% |
| Renfrewshire | 0.2% |
| Scottish Borders | 0.0% |
| Shetland Islands | 0.3% |
| South Ayrshire | 0.9% |
| South Lanarkshire | 1.3% |
| Stirling | 0.0% |
| West Dunbartonshire | 2.7% |
| West Lothian | 0.9% |
| **Scotland** | **0.8%** |

### 3.6.3 Babies Adjustment

People tend to miss babies from census forms, and we have evidence for this when comparing against information on the number of births from the registration system. Therefore an adjustment was made for 0 year olds to adjust to the mid-year estimate created especially for census day comparisons. The mid-year estimate was used rather than use birth registration directly as it takes account of migration. To ensure a smooth transition to older ages for 1 year olds we adjusted by three quarters of the distance between the census estimate and the mid-year estimate as at census day and for two year olds we adjusted by half of the distance between the census estimate and the mid-year estimate at census day. The Office for National Statistics and the Northern Ireland Statistics and Research Agency made similar adjustments for babies in the England and Wales and Northern Ireland 2021 censuses respectively.

Table 8: Size of the adjustment made to children aged 0 to 2 years old by age.

| Age | Percentage adjustment applied |
|---|---|
| 0 | 6.6% |
| 1 | 4.4% |
| 2 | 2.1% |

### 3.6.4 Sex-Ratio Adjustment for Ages 15 to 18

As part of our quality assurance we noticed inconsistences for a small number of ages in the sex ratio compared with other sources. There was no evidence of a wider issue and coverage patterns looked as expected. Therefore we made a minor change to adjust the sex ratio for ages 15, 16, 17 and 18 to be in line with the sex ratio based on the data directly collected in the census before any estimation took place (the census count). The census count sex ratio was also more in line with the sex ratio for these ages in other comparator sources. The adjustment was made by slightly reducing the number of males and slightly increasing the number of females for these ages. There were different options for how to make this adjustment, however because no wider issue was identified, both the numbers of males and females were adjusted , that is, there was no evidence males of these ages were missed meaning only the number of males should be adjusted.

Table 9: Size of the adjustment made to males and females aged 15 to 18 years old by age and sex.

| Age | Percentage adjustment applied to males | Percentage adjustment applied to females |
|---|---|---|
| 15 | -0.4% | 0.4% |
| 16 | -0.8% | 0.9% |
| 17 | -1.0% | 1.1% |
| 18 | -0.8% | 0.9% |

## 4. Population and Household Estimates — Working with Census Statistics, Confidence Intervals and Response Rates

### 4.1 Working with Census Statistics

Census statistics represent the total population rather than just those who completed the questionnaire. Since the 2001 censuses, statistical modelling has been used across the United Kingdom to produce total population estimates from census responses.

As with all estimates, there is a level of uncertainty. Users should consider uncertainty when working with census estimates. For example:

- Where population estimates show change since previous censuses smaller than +/-1%, this should be interpreted as minimal change rather than as an increase or decrease.
- Where differences between council areas are smaller than +/-1%, the areas should be interpreted as having a similar population size.

### 4.2 Confidence Intervals

Our targets for uncertainty published in our Statistical Quality Assurance Strategy referred to confidence intervals, which are a standard statistical measure produced as part of our modelling. A confidence interval is a measure of the sampling error in our estimates. As our population estimates rely on the CCS, a sample survey, this introduces a degree of uncertainty due to random variations in sample selection. The confidence interval estimates are the range within which the true population would fall 95% of the time if the survey was repeated.

The confidence interval targets for Scotland's 2022 Census estimates were +/- 0.4% at national level and +/- 3% at council area level for both people and households. At national level the population confidence interval achieved was +/- 0.51%, with 31 out of 32 council areas being within +/- 3%. East Lothian was slightly outside the interval at 3.07%

This means we have a population estimate for Scotland of 5,436,600 and a confidence interval of +/- 0.51% (27,900 people). That is the true population count is expected to be within 27,900 of the published estimate. In other words we can be confident that if we were to repeat the survey selection 100 times, 95 of the population estimates would be between 5,408,700 and 5,464,500.

These population confidence intervals are similar to 2011. At the national level the confidence interval in 2011 was +/- 0.44% (+/- 23,000 people), with all council areas being within +/- 3%.

These confidence intervals for population and households[3] are available in Table 10.

[3] A household can be: one person living alone, or a group of people (not necessarily related) living at the same address who share cooking facilities and share a living room or sitting room or dining area.

Table 10: Confidence intervals by council area for households and usual residents.

| Council Area | Person 95% confidence interval | Household 95% confidence interval |
|---|---|---|
| Aberdeen City | 1.99% | 1.94% |
| Aberdeenshire | 1.65% | 1.20% |
| Angus | 1.26% | 1.43% |
| Argyll and Bute | 2.59% | 1.16% |
| City of Edinburgh | 1.11% | 1.28% |
| Clackmannanshire | 1.88% | 1.41% |
| Dumfries and Galloway | 1.37% | 1.33% |
| Dundee City | 2.40% | 2.46% |
| East Ayrshire | 2.00% | 1.64% |
| East Dunbartonshire | 1.07% | 0.80% |
| East Lothian | 3.07% | 1.41% |
| East Renfrewshire | 1.08% | 0.89% |
| Falkirk | 1.78% | 1.25% |
| Fife | 1.96% | 1.36% |
| Glasgow City | 1.51% | 1.62% |
| Highland | 2.15% | 1.51% |
| Inverclyde | 1.59% | 1.37% |
| Midlothian | 1.23% | 1.05% |
| Moray | 1.46% | 1.05% |
| Na h-Eileanan Siar | 2.72% | 1.52% |
| North Ayrshire | 2.52% | 1.25% |
| North Lanarkshire | 2.38% | 1.95% |
| Orkney Islands | 1.14% | 1.20% |
| Perth and Kinross | 2.31% | 1.21% |
| Renfrewshire | 1.79% | 1.28% |
| Scottish Borders | 1.89% | 1.70% |
| Shetland Islands | 1.34% | 1.23% |
| South Ayrshire | 1.51% | 1.07% |
| South Lanarkshire | 1.87% | 1.34% |
| Stirling | 1.00% | 1.99% |
| West Dunbartonshire | 2.63% | 1.91% |
| West Lothian | 1.62% | 1.37% |
| **Scotland** | **0.51%** | **0.47%** |

## 4.3 Difference Between Return Rates and Response Rates

Return rates are our operational measure during the collection phase. They are calculated by dividing the count of unprocessed household census form responses received by the count we expected to receive, as per our collection address frame. We used these to ensure we gave sufficient attention to each area to achieve our aim of consistent coverage and quality across areas. Return rates are only calculated for households.

Response rates are our output measure. They are calculated by dividing the count of sufficiently complete responses by the number we estimated should have responded. We use these to produce a measure of observed characteristics we collected, which we can compare with our estimate of the true population.

Response rates can be calculated for both persons and households.

Scotland's Census 2022 usual resident person response rate was 89.8%, this means we estimated 10.2% of the population.

We produce return and response rates for several reasons. They:

- help users to understand the coverage and quality of census data
- provide confidence we have maximised response to the census
- enable us to see what proportion of each census measure is the result of observed data
- assure us there are no gaps in the population estimates for specific groups

This last point is particularly important. Although we had a lower than expected return rate to the census, our collection strategy ensured that we did not have gaps in certain areas.

We are also able to estimate the person and household response rate for each council area, which is the number of households and residents who provided a census return divided by the estimated number of households and usual residents in the council area. 31 out of 32 councils had a person response rate of 85% or more and all councils were above 82%. These response rates are shown in Table 11.

Table 11: Person and household response rates by council area

| Council Area | Person response rate | Household response rate |
|---|---|---|
| Aberdeen City | 88.2% | 89.0% |
| Aberdeenshire | 92.6% | 92.6% |
| Angus | 92.6% | 91.4% |
| Argyll and Bute | 91.5% | 91.8% |
| City of Edinburgh | 90.4% | 91.0% |
| Clackmannanshire | 91.3% | 90.5% |
| Dumfries and Galloway | 92.7% | 91.1% |
| Dundee City | 86.2% | 87.3% |
| East Ayrshire | 92.0% | 91.9% |
| East Dunbartonshire | 91.8% | 91.7% |
| East Lothian | 90.6% | 92.2% |
| East Renfrewshire | 90.3% | 89.4% |
| Falkirk | 92.6% | 92.6% |
| Fife | 90.8% | 91.2% |
| Glasgow City | 82.1% | 83.9% |
| Highland | 92.2% | 91.2% |
| Inverclyde | 85.3% | 85.0% |
| Midlothian | 91.6% | 93.3% |
| Moray | 92.7% | 92.5% |
| Na h-Eileanan Siar | 92.9% | 91.8% |
| North Ayrshire | 90.9% | 90.1% |
| North Lanarkshire | 88.7% | 89.1% |
| Orkney Islands | 94.0% | 91.8% |
| Perth and Kinross | 91.8% | 91.8% |
| Renfrewshire | 87.6% | 86.6% |
| Scottish Borders | 91.5% | 91.4% |
| Shetland Islands | 92.7% | 91.4% |
| South Ayrshire | 91.4% | 92.1% |
| South Lanarkshire | 91.7% | 91.7% |
| Stirling | 90.5% | 90.9% |
| West Dunbartonshire | 88.5% | 87.5% |
| West Lothian | 92.7% | 91.5% |
| **Scotland** | **89.8%** | **89.9%** |

Response rates by age group and sex can been seen in Table 12. Generally speaking response rates are lower for men than women and lower for younger adults, very similar to patterns seen in previous censuses.

Table 12: Person response rates by age group and sex, Scotland

| Age group | All | Females | Males |
|---|---|---|---|
| 0-4 | 85.8% | 86.0% | 85.7% |
| 5-9 | 87.2% | 87.3% | 87.2% |
| 10-14 | 88.9% | 89.0% | 88.8% |
| 15-19 | 83.4% | 83.8% | 83.0% |
| 20-24 | 80.2% | 81.2% | 79.1% |
| 25-29 | 85.1% | 86.2% | 83.9% |
| 30-34 | 85.6% | 86.7% | 84.4% |
| 35-39 | 87.3% | 88.4% | 86.1% |
| 40-44 | 89.1% | 90.2% | 88.0% |
| 45-49 | 90.3% | 91.3% | 89.3% |
| 50-54 | 92.4% | 92.9% | 91.8% |
| 55-59 | 94.1% | 94.6% | 93.6% |
| 60-64 | 95.7% | 96.1% | 95.3% |
| 65-69 | 96.6% | 96.6% | 96.7% |
| 70-74 | 96.4% | 96.3% | 96.6% |
| 75-79 | 95.9% | 95.6% | 96.3% |
| 80-84 | 94.0% | 93.8% | 94.3% |
| 85-89 | 92.0% | 92.0% | 92.0% |
| 90+ | 86.8% | 88.8% | 83.0% |
| **Scotland** | **89.8%** | **90.4%** | **89.1%** |

## 5. Validation of Population Estimates

We have undertaken an extensive process to compare the Census data with other sources to ensure the population and household estimates are plausible. This is known as Validation of Population Estimates.

### 5.1 Population

The Census estimates of the usual resident population of Scotland were compared with a number of sources, including:

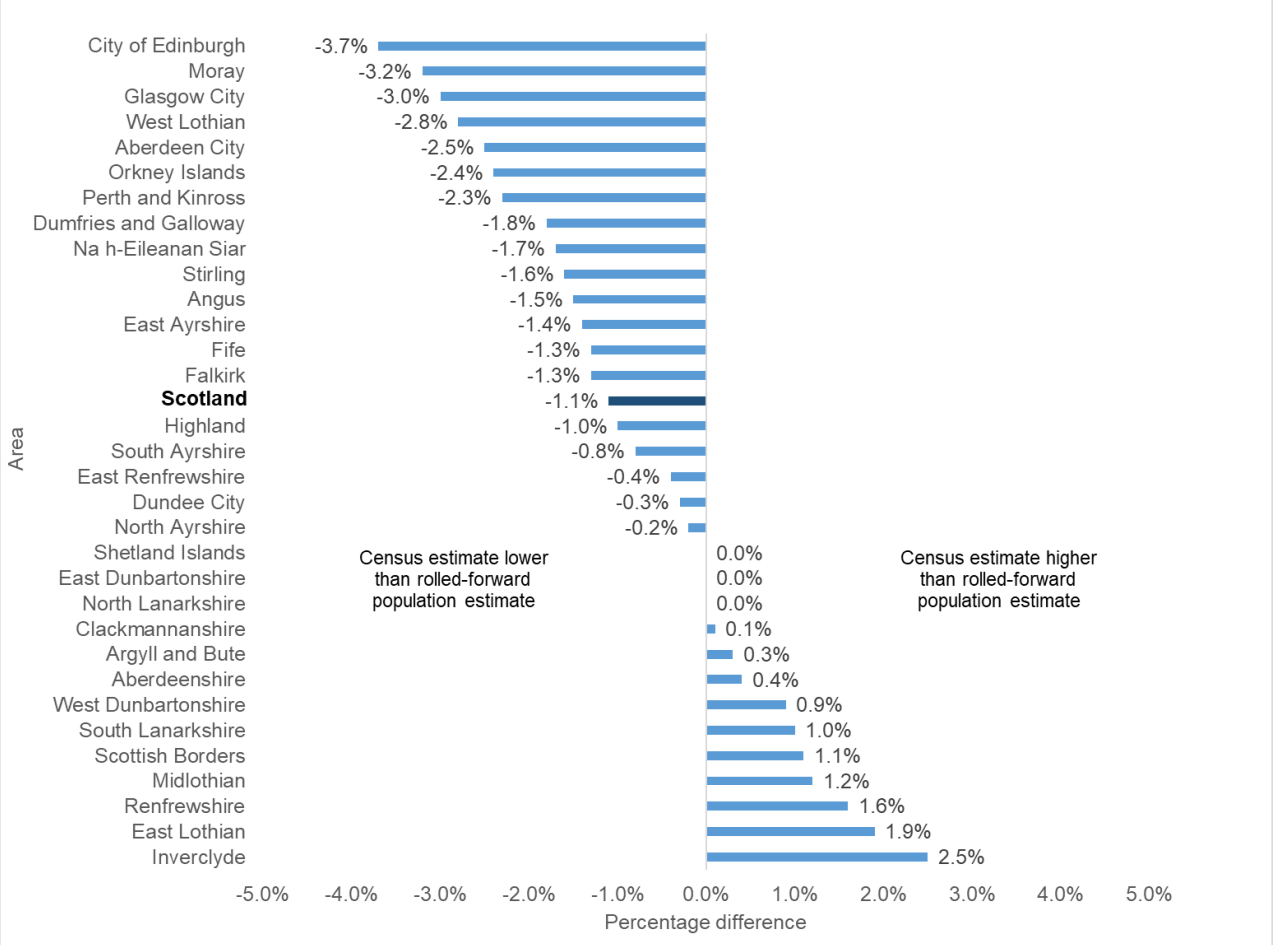
- Mid-year population estimates
- Administrative data based population estimates
- the National Health Service Central Register
- Department for Work and Pensions RAPID data
- Census 2011
- School pupil census
- Private school census
- Higher Education Statistics Agency student data

This validation was done at national and council area level. Differences between the census and comparator sources were explored to aid the understanding of differences.

One of the main comparator sources used was the NRS estimates of the population as at 30 June for each year. These mid-year population estimates are based on the latest census and births are added on and deaths taken off and adjustments are made for migration. There are also some smaller adjustments for certain populations such as armed forces and prisoners. Information for the number of births and deaths is considered to be highly accurate, as it is based on the compulsory registration of births and deaths. Measuring migration is more challenging and over the period since the last census, estimates of migration have been made based on administrative and survey data. For census quality assurance, the 2021 mid-year population estimates were rolled forward to census day on 20 March 2022.

At Scotland level the rolled forward population estimates were higher than the 2022 census estimates by 62,000, a difference of 1.1%. This varied by council area with some being higher and some lower as shown in Figure 2.

Figure 2: Percentage difference between rolled-forward population estimate and Census 2022 population estimate, Scotland and Council Areas.

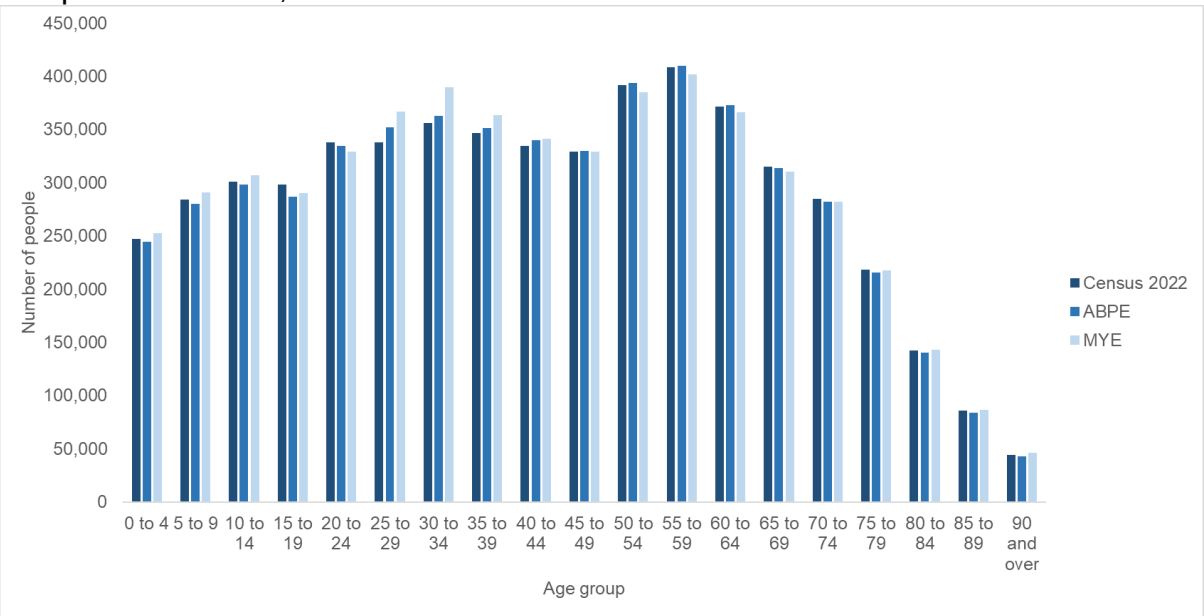



Administrative data Based Population Estimates were also used as a key source to compare with the census estimates. These population estimates are created using only a range of administrative data. This research is part of the evidence base to support future developments in Scotland's demographic statistics. For comparison with Census 2022, a Census day version of the Administrative data Based Population Estimates was created. More information on the Administrative data Based Population Estimates are available on the NRS website.

An Excel spreadsheet is available on our website which compares the census population estimates for council areas by age group and sex with the rolled-forward population estimate and the Administrative data Based Population estimates.

Figure 3 shows that for most ages the patterns of the census estimates are very similar to the mid-year estimates and Administrative data Based Population Estimates, especially for older people where there tends to be fewer migratory moves. For younger people especially in cities where the challenge of measuring migration is greater there is a difference between the census and the mid-year estimates but this is why we do a census to rebase the mid-year estimates. Interestingly the age pattern of the census for these ages groups is closer to the Administrative data Based Population Estimates. These differences will be the subject of further research.

Figure 3: Estimated population of Scotland by 5 year age band, Census 2022 and comparator sources, 20 March 2022.




A more detailed reconciliation of the census estimates with the mid-year estimates, analysing the differences between them, will be published next year.

## 5.2 Households

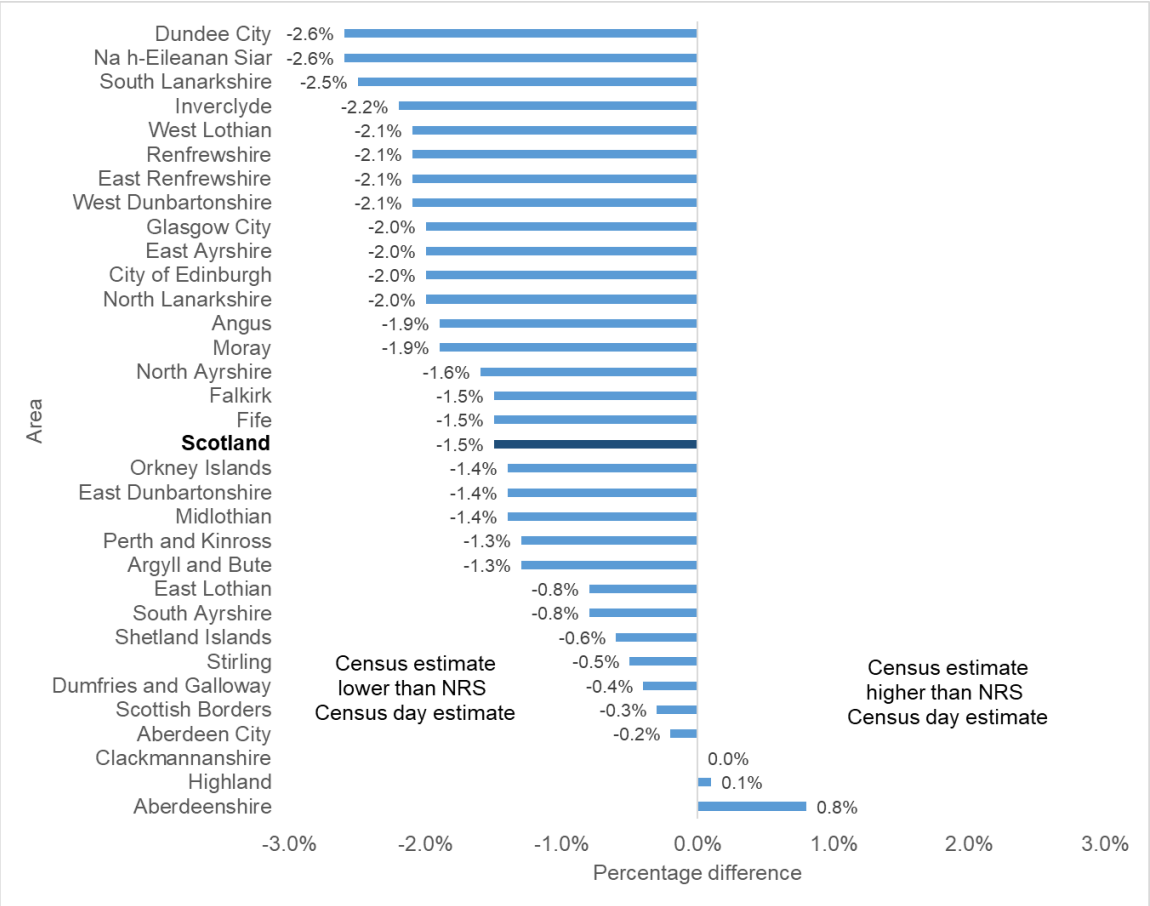
The Census household estimates were compared a number of sources, including:

- NRS rolled forward household estimates
- Alternative household estimates
- Census 2011

The NRS annual household estimates were one of the main comparator sources used to validate the census household estimates. The 2021 and 2022 household estimates were interpolated to produce a census day estimate as at 20 March 2022. The actual data the estimates are based on is at September, so some differences between this source and census was caused by differences in timing. There are also some known definitional differences between the two sources, for example some accommodation buildings for students are classed as individual households in the household estimates, but are classed as communal establishments by the census and are therefore estimated separately. More information on the NRS annual household estimates is available on the NRS website.

At Scotland level, the Census household estimate is 1.5% lower than the NRS rolled forward household estimate. 29 of 32 local authorities also have a lower Census estimate for households than the rolled forward household estimates, as shown in Figure 4.

Figure 4: Percentage difference between Census 2022 household estimates and NRS rolled forward household estimates, Scotland and Council Areas.




## 5.3 Sex question

Scotland's Census 2022 asked respondents 'what is your sex?' or 'what is this person's sex?' if they were responding on behalf of someone else within the household.

The following guidance was available to respondents, who had to click a link to see it in the online questionnaire. No specific guidance was provided alongside paper questionnaires.

> 'How do I answer this question?
> If you are transgender the answer you give can be different from what is on your birth certificate. You do not need a Gender Recognition Certificate (GRC).
>
> If you are non-binary or you are not sure how to answer, you could use the sex registered on your official documents, such as your passport.
>
> A voluntary question about trans status or history will follow if you are aged 16 or over. You can respond as non-binary in that question.'

The full guidance for the sex question is available on the Scotland's Census website.

The number of respondents who accessed this guidance was 0.13%, a very small proportion of the total number who responded to this question. Therefore, it is likely that the number of respondents requiring clarification on the definition was small.

The definition used for sex is consistent with our approach in the 2011 Census and we are confident that the data collected on sex is of high quality and is comparable to other sources, including the Scotland's 2011 Census and the 2021 Censuses in Northern Ireland and England and Wales.

## 6. External Assurance

### 6.1 National Statistics Accreditation

As mentioned at the start of this paper the statistics being released have been assessed as National Statistics by the Office for Statistics Regulation (OSR). This is the result of a rigorous assessment of the quality, trustworthiness and value of these statistics and the compliance of National Records of Scotland (NRS) with the Code of Practice for Official Statistics. More information about the assessment can be found on the OSR website.

### 6.2 International Steering Group

The International Steering Group was formed in May 2022. The group's role in Scotland's Census 2022 is to:

- provide assurance that the census programme was ready to move on from the collect phase
- provide direction and support to National Records of Scotland as we implement our statistical design after the collection phase
- propose amendments or changes for us to consider, including accessing additional administrative data

The group has made a statement regarding their confidence in the methods used by NRS to produce the census population and household estimates:

'The International Steering Group commends NRS for its systematic and robust application of the census estimation methodology that has led to today's published estimates.'

More information on the International Steering Group is available on our website.

### 6.3 Scottish Government Chief Statistician Coherence Group

In May 2023 Registrar General Dr Janet Egdell wrote to Alistair McAlpine, Chief Statistician for Scotland, setting out the assurance steps that NRS was taking ahead of publication of first outputs from Scotland's 2022 Census.

This included assessment by Office for Statistical Regulation (OSR) to publish as National Statistics, a statement from the International Steering Group on the statistical methods used and a review of the work that NRS will undertake on the coherence of census outputs with other statistics and data.

Janet asked Alistair in his capacity as Chief Statistician for Scotland to lead the review on coherence prior to publication. The Chief Statistician formed a group to undertake this work. The Chief Statistician for Scotland then wrote to the Registrar General to confirm they would form a group to assess the coherence of Census statistics and this letter is published on our website.

In August, the Chief Statistician confirmed:

> **"I am therefore satisfied…that the outputs will provide the robust population statistics to support the wider statistical system and that analysts and the general public can have confidence in the outputs, and that they can be published on 14 September 2023."**

More information on the Scottish Government Chief Statistician Coherence Assessment group is available on our website.

### 6.4 Quality assurance with Council Areas

NRS engaged with council areas, involving them in the quality assurance of census estimates. This involved sharing early population and household estimates for their areas with representatives from all councils who wanted to take part, providing them with the opportunity to provide feedback if there was evidence that these estimates differed from local data sources.

30 of 32 council areas were involved in this process and 28 provided feedback on the census estimates. Meetings were also held with 10 council areas to discuss the early census estimates for their areas. All council areas were also offered the opportunity to meet.

The themes from this process are set out in Table 13 alongside the action NRS took. This was a very useful process in quality assuring our data and we would like to thank council areas for their time and input.

Table 13 : Council area feedback themes

| Theme | Council comments | NRS response |
|---|---|---|
| Population lower than comparators for school age children | General comments about lower population than expectations, with specific reference to administrative data on school age children | An adjustment was made to school age children as described in Section 3.6.2. |
| Occupied households lower than Council Tax data or address list | Specific concerns raised by some local authorities | A household adjustment was made for several council areas for occupied households using information from an independent household estimate. See Section 3.6.1 for more details. |
| Armed Forces bases | Some feedback from areas with large armed forces bases | Checked MOD data, evidence from ONS Census, CE manager form, HMRC/DWP and Health data. Concluded that there is a definitional difference between how armed forces are counted in the census and mid-year estimates. Evidence of people working at these bases being counted at their family homes in other parts of the country in the census. No adjustment made to census estimates. |
| Small area differences (councils were sent very early indicative total estimates for 2011 census data zone population and households) | Referencing differences in household numbers at small area level | Most differences are relatively small with many large differences explained by definitional differences in how communal establishments are counted in the census compared with key sources like council tax, for example in the census some students' accommodation could be counted as one communal establishment whereas in council tax systems, it could be classed as individual flats. NRS will look further at these comments as part of the next step of census processing where we complete the unit level census database so that it is consistent with the estimates at council area level. |