

Scotland's Census 2022

Statistical Quality Assurance Strategy

Version 1.0, May 2021

Contents

| | |
|--|-----------|
| 1. Introduction | 4 |
| 1.1 Scope of this document | 4 |
| 2. What is Statistical Quality Assurance, Control and Management? | 4 |
| 2.1 What is Statistical Quality? | 4 |
| 2.1.1 Code of Practice for Statistics | 5 |
| 2.1.2 European Statistical System's Dimensions of Quality | 6 |
| 2.2 What is Statistical Quality Assurance? | 7 |
| 2.3 What is Statistical Quality Control? | 7 |
| 2.4 What is Statistical Quality Management? | 8 |
| 3. Statistical Quality Assurance Strategy | 8 |
| 3.1 Scotland's Census 2022 Objectives | 8 |
| 3.2 How Will We Produce High Quality Results? | 9 |
| 3.3 Statistical Quality Assurance Strategy and Principles | 10 |
| 4. Statistical Quality Assurance Approach | 11 |
| 4.1 Assurance of Processes | 11 |
| Our approach to the assurance of the results of each process is to: | 11 |
| 4.2 Validation of population estimates | 12 |
| 4.3 Administrative data | 13 |
| 4.4 Timing | 13 |
| 5. Statistical Quality Assurance End-to-End Journey | 13 |
| 5.1 Question Development | 14 |
| 5.2 Questionnaire Design | 15 |
| 5.3 Census Address Register (CAR) | 17 |
| 5.4 Enumeration Strategy | 17 |
| 5.5 Online and Paper Collection | 19 |
| 5.6 Field Force Strategy | 20 |
| 5.7 Census Coverage Survey | 20 |
| 5.8 Census Quality Survey | 21 |
| 5.9 Load and Validation | 21 |
| 5.10 Coding | 22 |
| 5.11 Remove False Persons and Resolve Multiple Responses | 22 |
| 5.12 Name Re-ordering | 23 |
| 5.13 Filter Rules | 24 |

| | | |
|-----------|---|-----------|
| 5.14 | Edit and Imputation..... | 24 |
| 5.15 | Census to Census Linking and Census Coverage Survey to Census Linking 25 | |
| 5.16 | Estimation | 26 |
| 5.17 | Adjustment | 27 |
| 5.18 | Validation of Population Estimates and Topic-based Quality Assurance..... | 27 |
| 5.19 | Derived Variables Created for Outputs | 29 |
| 5.20 | Output Areas | 29 |
| 5.21 | Statistical Disclosure Control | 30 |
| 5.21.1 | Record Swapping..... | 30 |
| 5.21.2 | Cell Key Perturbation..... | 30 |
| 5.22 | Census Outputs | 31 |
| 5.22.1 | Pre-Built Tables | 31 |
| 5.22.2 | Geographies and Totals | 32 |
| 5.22.3 | Additional Outputs | 32 |
| 5.22.4 | Origin-Destination | 32 |
| 5.23 | Management Information | 33 |
| 6. | Stakeholder engagement | 34 |

1. Introduction

1.1 Scope of this document

The success of Scotland's Census 2022 will be judged primarily on its ability to deliver high quality population and household statistics in accordance with agreed timescales. The required quality of these statistics will be highly dependent on the programme's ability to produce high quality systems, business processes, products and services that in turn support the collection and processing of Census data and the production and dissemination of statistical outputs.

This strategy gives an overview of how we will assess and measure the level of quality being achieved throughout the collection and processing of Census data and the production and dissemination of statistical outputs.

With that in mind, this document is designed to:

- outline what statistical quality assurance, quality control and quality management are and why they are important for Scotland's Census 2022 ([section 2](#));
- describe our proposed strategy for Statistical Quality Assurance for Census 2022 ([section 3](#));
- provide an overview of the methods we will use to quality assure the Census 2022 processes and results ([section 4](#));
- describe and explain the Statistical Quality Assurance end-to-end journey built into Scotland's Census 2022 ([section 5](#)); and
- explain how we have engaged with stakeholders and how we plan to continue doing so ([section 6](#)).

This work will be led by Scotland's Census 2022 Statistical Quality Assurance team.

2. What is Statistical Quality Assurance, Control and Management?

2.1 What is Statistical Quality?

Quality means that statistics fit their intended uses, are based on appropriate data and methods, and are not materially misleading.

Quality requires skilled professional judgement about collecting, preparing, analysing and publishing statistics and data in ways that meet the needs of people who want to use the statistics.

2.1.1 Code of Practice for Statistics

Scotland's Census 2011 was assessed and designated a National Statistic by the [UK Statistics Authority](#). National Statistics are produced in line with the [Code of Practice for Statistics](#) which denotes good statistics as those that command public confidence by demonstrating trustworthiness and providing high quality outputs that enhance public value.

Scotland's Census 2022 intends to deliver on trustworthiness, quality and value. The [Office for Statistics Regulation \(OSR\)](#), the regulatory arm of the UK Statistics Authority, assesses compliance of our statistics against the three pillars of the Code of Practice for Statistics: trustworthiness, quality and value . We are undertaking a three-phase accreditation process to ensure we [retain our National Statistics designation](#) for Census 2022.

Quality in the [Code of Practice for Statistics](#) means that statistics fit their intended uses, are based on appropriate data and methods, and are not materially misleading. Quality requires skilled professional judgement about collecting, preparing, analysing and publishing statistics and data in ways that meet the needs of those who want to use the statistics.

Specifically, the Quality pillar states that 'Producers should demonstrate how they assure themselves that the statistics are robust and reliable'. The Quality pillar contains five principles, which producers of Official Statistics have to provide evidence of adherence:

1. Statistics should be produced to a level of quality that meets users' needs. The strengths and limitations of the statistics and data should be considered in relation to different uses, and clearly explained alongside the statistics.
2. Quality assurance should be proportionate to the nature of the quality issues and the importance of the statistics in serving the public good. Statistics producers should be transparent about the quality assurance approach taken throughout the preparation of the statistics. The risk and impact of quality issues on statistics and data should be minimised to an acceptable level for the intended uses.
3. The quality of the statistics and data, including their accuracy and reliability, coherence and comparability, and timeliness and punctuality, should be monitored and reported regularly. Statistics should be validated through

comparison with other relevant statistics and data sources. The extent and nature of any uncertainty in the estimates should be clearly explained.

4. Scheduled revisions, or unscheduled corrections that result from errors, should be explained alongside the statistics, being clear on the scale, nature, cause and impact.
5. Systematic and periodic reviews on the strengths and limitations in the data and methods should be undertaken. Statistics producers should be open in addressing the issues identified and be transparent about their decisions on whether to act.

2.1.2 European Statistical System's Dimensions of Quality

Alongside the Code of Practice for Statistics, Scotland's Census 2022 also follows the principles set out by the [Quality Assurance Framework of the European Statistical System](#). This framework is in place for all producers of official statistics in the European Union (EU) to ensure a consistency of approach when developing, producing and disseminating European statistics.

The European Statistical System's Dimensions of Quality are:

| | |
|----------------------|---|
| RELEVANCE | the degree to which a statistical product meets needs for content and coverage; |
| ACCURACY | the closeness between an estimated value and the true value; |
| RELIABILITY | closeness of the initial element value to the subsequent estimated measure; |
| TIMELINESS | the time between the date of publication and the date to which the data refers; |
| PUNCTUALITY | the time between the actual and planned publication of a statistic; |
| ACCESSIBILITY | the ease with which users can access data; |
| CLARITY | the quality and sufficiency of metadata, illustrations and accompanying advice; |

| | |
|----------------------|--|
| COMPARABILITY | the degree to which data can be compared over time and domain; |
| COHERENCE | the degree to which data from different sources on the same topic, is similar. |

2.2 What is Statistical Quality Assurance?

The [Government Statistical Service Statistical Policy and Standards Committee \(GSS SPSC\)](#) define Quality Assurance (QA) as the anticipation and avoidance of problems. It is about having agreed systems that check and validate the work that we do so that our end product is robust and received well by the end user.

Statistical Quality Assurance

- covers all procedures focused on providing confidence that quality requirements will be fulfilled;
- requires processes and systems in place that are planned and tested, and which should self-correct or flag problems under exceptions;
- aims to prevent, reduce or limit the occurrence of errors in a statistical product and therefore, to get it right first time.

Examples of Statistical Quality Assurance include, but is not limited to:

- data entry systems programmed to pick up scanning errors;
- validation gates programmed into computer systems;
- trend analysis;
- regular checks with administrative data sources;
- peer reviews;
- expertise checks; and
- recreating outputs using different applications.

2.3 What is Statistical Quality Control?

The Statistical Policy and Standard Committee ([SPSC](#)) set up by the Government Statistical Service ([GSS](#)) define Quality Control (QC) as responding to observed problems. This means that, if something has gone wrong, we have to have an agreed approach to putting things right. Having agreed responses for quality control allows us to respond when things go wrong quickly and effectively.

Examples of Statistical Quality Control include, but are not limited to:

- querying errors with organisations that supplied the information (for example, administrative data);
- retrieving the raw data to enable comparisons before and after adjustments;
- amending errors in released documents, updating metadata and re-releasing statistical outputs;
- responding to validation gate errors by manually editing the returned value.

2.4 What is Statistical Quality Management?

The GSS SPSC define Statistical Quality Management as the encompassing approach to quality. It can also be defined as the application of the documented policies, processes and procedures relating to statistical quality assurance and statistical quality control.

These include:

- adhering to the principles as set out in the Code of Practice for Statistics;
- ensuring all staff are trained to produce quality outputs;
- quality initiatives that ensure we are compliant with the Code of Practice;
- promoting a culture of continuous improvement;
- monitoring and management of data quality;
- ensuring all risks are managed in accordance with the risk management strategy;
- auditing of statistical processes.

3. Statistical Quality Assurance Strategy

The Statistical Quality Assurance Strategy describes how we will measure the level of quality being achieved throughout the collection and processing of Census data and the production and dissemination of the results.

3.1 Scotland's Census 2022 Objectives

The key objectives for Scotland's Census 2022 are:

- to produce high quality results;
- to generate outputs that meet the needs of users;
- to maximise online response rates for the census;
- to produce timely outputs to maximise benefits;
- to protect, and be seen to protect, confidential information;

- to do so in a cost effective way; and
- to make recommendations for the approach to future censuses in Scotland.

The implementation of a structured but pragmatic approach to statistical quality assurance will contribute to Scotland's Census 2022 achieving these strategic objectives.

This Statistical Quality Assurance Strategy will also help to support Scotland's Census in meeting the strategic objectives of the UK Statistics Authority, as set out in [Statistics for the Public Good](#). It will also help to ensure that we meet our obligations under the [Code of Practice for Statistics](#).

3.2 How Will We Produce High Quality Results?

This document outlines the Statistical Quality Assurance strategy which focuses on the first of Scotland's Census objectives to produce high quality results.

Critical Success Factors (CSFs) describe what success will look like and are aligned to Scotland's Census 2022 objectives. While CSFs will evolve as the Statistical Quality Assurance Strategy develops, the overarching definitions of success for producing high quality results as at November 2019 are:

| How will we achieve high quality results? | How will we measure success? (Key Performance Indicators (KPIs) and acceptance levels) |
|---|---|
| We will maximise our overall person response rate | Person response rate ¹ of at least 94%. |
| We will ensure a minimum level of response within every local authority in Scotland | Person response rate ¹ in every council area of at least 85%. |
| We will maximise the accuracy of our national population estimates | Variability ¹ : national estimates will achieve 95% Confidence Intervals (CI) +/- 0.4%; Bias: < 0.5% |

| | |
|---|--|
| We will maximise the accuracy of our local authority population estimates | Variability ¹ : Council Area estimates will achieve 95% CI +/- 3% |
| We will minimise the non-response to all mandatory questions | Achieve or exceed target non-response rates for all mandatory questions |
| 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 will publish details of methods and full details of all our data quality indicators | Publish on our website |
| We will publish the results of an independent methodology review | Publish on our website |
| We will maintain our National Statistics Accreditation | Accreditation maintained throughout |

Notes

1. The acceptance level for this Key Performance Indicator is under review.

We expect our plans to evolve in light of further stakeholder engagement, as a result of our experience from the [2019 census rehearsal](#), and as a result of developments and availability of tools and alternative data sources.

3.3 Statistical Quality Assurance Strategy and Principles

In achieving the CSFs above ([section 3.2](#)), the Statistical Quality Assurance strategy for the 2022 Census aims to:

- ensure that the Census results provide a reliable basis for decision-making;
- give data-users confidence that the Census results are fit for purpose;
- minimise the risk of errors in the Census estimates, however minor, which might undermine the overall credibility of the statistics;
- leave a legacy of methods, tools and skills for the quality assurance of post-2022 population statistics.

To achieve this, we will adopt the following strategic principles:

- Quantify accuracy wherever possible before, during and after the 2022 Census by establishing quantitative estimates of various aspects of error. This will allow us to:
 - prioritise quality assurance activities;
 - help in making evidence-based decisions on any interventions required;
 - assess the success of the census against the agreed success criteria; and
 - subsequently help users understand the reliability of the results.
- Ensure every quality assurance check has a purpose, such as informing a recommendation on an intervention or to help people use the data sensibly.
- Use flexible methods and tools – every census is different and we will need to react quickly to unexpected anomalies in the census data, which will be useful for the quality assurance of post-2022 population statistics.
- Be outward looking by sharing experiences and learning from other census-taking organisations and work in partnership with data users to understand what would give them confidence in the census data, and to provide that accordingly.
- Take explicit account of all [lessons learned from Scotland's Census 2011](#).
- Conduct the statistical quality assurance activities in parallel to the processing of data. By applying the quality checks as soon as the first data comes into NRS we can identify issues as soon as possible.

4. Statistical Quality Assurance Approach

4.1 Assurance of Processes

Scotland's Census collects information from every person and household in Scotland, and produces essential data, which is not available in any other data sources. The census data undergoes a number of statistical processes that form the overall census data journey. We will review quality assurance procedures performed internally for each process within the census data journey, and will provide a secondary level of quality assurance for every individual process.

Our approach to the assurance of the results of each process is to:

1. Understand each process and the nature of the errors which might be introduced:
 - Define metrics to understand the nature of errors, measures and acceptance criteria
 - Define interventions to resolve and sign-off for each process
2. Estimate the expected size and likelihood of the errors within each process
3. Prioritise the Quality Assurance work based on points 1 and 2:
 - Define dependencies, risks, timeframes, and resources for the Quality Assurance of each process;

The Scotland's Census 2022 Statistical Quality Assurance team will work closely with the owner of each process, drawing on their expertise to ensure that quality assurance checks are meaningful and can be easily produced, avoiding duplication of work, and ensuring methods of intervention are available where the results of a process are causing concern.

4.2 Validation of population estimates

By validating census population estimates, we will consider the likely accuracy of the census results for geographic areas, population groups and topics. We aim to ensure that the main demographic statistics are plausible for Scotland and its constituent Local Authorities, as well as other geographies and topic areas.

Specific quality assurance activities include:

- We will use a suite of tools and methods that will allow us to focus on geographic areas and population groups or topics where there are inconsistencies or need for further analysis.
- In 2022, we will aim to make best use of all sources available, including:
 - 2011 Census;
 - mid-year population estimates;
 - administrative data;
 - survey data;
 - other sources.
- The evidence to support our population estimates will be reviewed by internal and external Quality Assurance Panels who will advise on whether estimates are fit for purpose or require further work or adjustment.

4.3 Administrative data

NRS will support data linkage and the quality assurance of administrative datasets before use following the Census in 2022. As part of this, supporting documentation for the purposes of data sharing and linkage will be published such as:

- Quality Assurance of Administrative Data (QAAD) documents; and
- Data Protection Impact Assessments (DPIA).

This data will feed in to the assurance of processes and validation of population estimates.

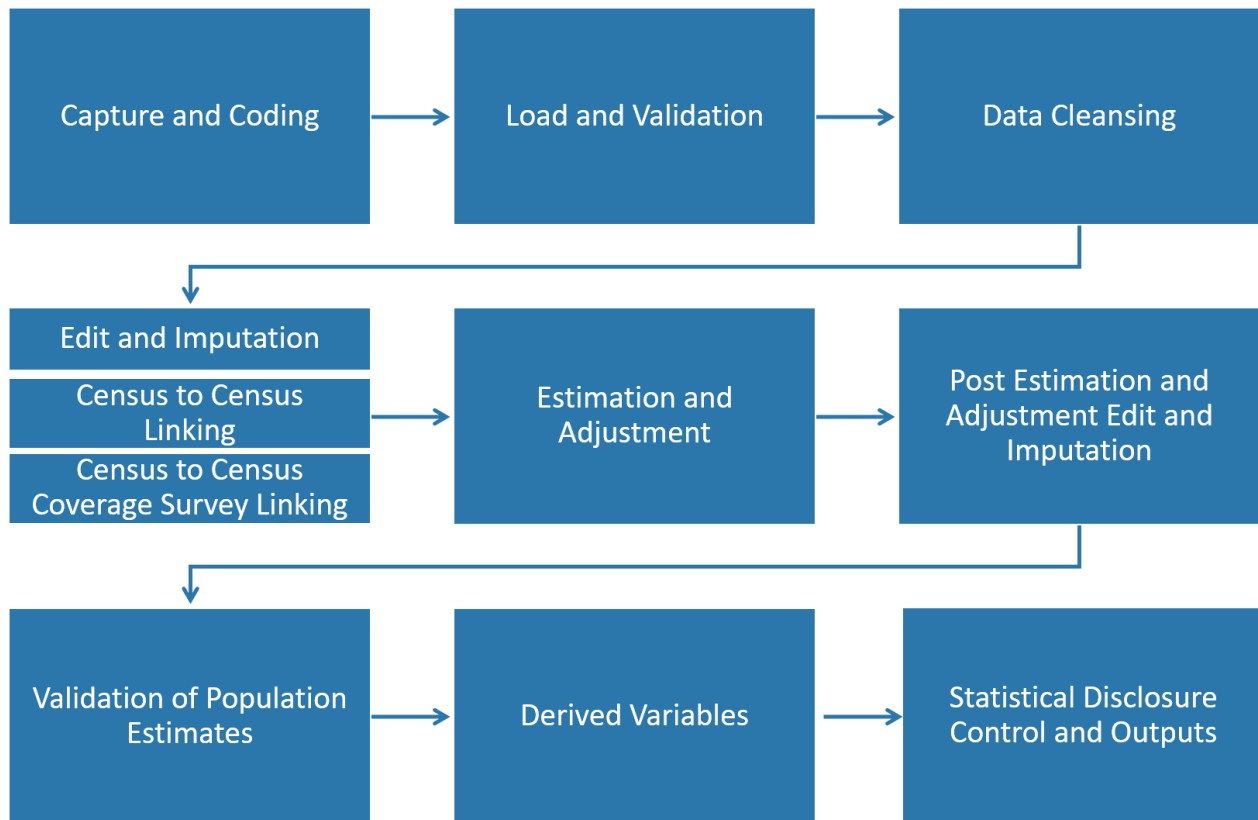
4.4 Timing

With the time constraint of publishing the first census outputs by the end of March 2023, we will need to conduct the Statistical Quality Assurance work in parallel with the processing of Census data and the production and dissemination of statistical outputs. With that in mind, the tools that we develop or procure for statistical quality assurance will be designed for flexibility, enabling analysts to interrogate data and adapt analyses as needed, rather than being constrained to a previously specified set of analyses. We will also seek to make best use of expertise elsewhere in NRS and the Scottish Government to maximise the trustworthiness, quality and value of our census outputs through expert-led and peer-reviewed statistical quality assurance.

5. Statistical Quality Assurance End-to-End Journey

Understanding the flow of census data through an end-to-end journey is vital for ensuring that an appropriate quality assurance plan, including the development of meaningful quality indicators, is in place and is being followed. The end-to-end journey will also help to bring together all of the statistical quality information for Scotland's Census 2022 to form an overall picture of data quality, a report of which will be published after the final census outputs have been released in 2024. [Census Outputs Strategy](#) can be found on our website. The quality assurance, control and management procedures detailed below will also be assessed as part of our evidence for [National Statistics accreditation](#). An overview of the steps in the census data journey is shown in [Figure 1](#).

Figure 1: Census data journey



We expect that the quality assurance activities proposed in this paper will evolve based on stakeholder feedback, evaluation of the [census rehearsal](#) in 2019/2020, and peer review of tools and methodologies. Our statistical data processing methodologies have been reviewed through the [External Methodology Assurance Panel](#).

5.1 Question Development

Census questions are thoroughly tested to ensure they produce high quality outputs that meet data user and respondent needs.

In deciding what subjects to include and what questions to ask, we consulted many stakeholders and organisations to take full account of circumstances in Scotland.

We also considered:

- how acceptable the questions are to the public;
- how to ask questions in a way that produces reliable answers; and
- whether other ways of collecting the information already exist.

The Scottish Parliament made the final decision on which questions to include in Scotland's Census 2022. More details on the legislative process for how Scotland's Census will be conducted, including the detail of the questions to be asked, see the [Legislation](#) page on the Scotland's Census website.

Specific quality assurance activities include:

- [Topic consultations](#) asked for views on the topics data users thought should be included in the next census. This included where the questions could be investigated for further development to improve the data quality of existing questions.
- We held a number of [topic events, meetings and focus groups](#) to gather and understand stakeholder needs and views.
- All [topic reports and supporting documentation](#) have been published on our website.
- We developed question evaluation criteria for [existing, alternative and new questions](#), and for [tick box response options](#).
- We undertook [Cognitive and Quantitative Question Testing](#) to build on an extensive programme of work on question development. Question testing purposively investigated impacts on data quality between digital and paper formats.
- The [Plans for Scotland's Census](#) document lays out proposals set out in 2018 for how the census will be conducted and the questions it will ask. The legislation reflects the final position.
- Analysis of any mode effects on the data. For the 2022 Census, people in Scotland will be encouraged to fill out their census questionnaire primarily online, and as such where a digital platform improves the data quality, these changes will be implemented. This means there will be modal differences in data quality between digital and paper responses. The digital platform for the online census questionnaire allows use of functionality not used in Scotland's Census previously. For example, as the use of predictive text, which allows the user to select from given list dependent on what they start to type. This additional digital functionality was evaluated following the [2019 census rehearsal](#).

5.2 Questionnaire Design

We expect most people to complete the census online in 2022. This builds on the success of the online household questionnaire used in 2011. The online questionnaire

needs to be easy to use as well as being suitable for a range of digital devices (for example, a computer, tablet or smartphone).

The paper and online questionnaires both capture the same information, but were designed as two separate methods of completing the census questionnaire. This way, we were able to make use of the advantages of the online census, maximising online take up and the gains in data quality.

Specific quality assurance activities include:

- The design of the online questionnaire takes account of good practice standards and guidance.
- Features of the online questionnaire include:
 - relevant on-screen information to help answer questions
 - use of functionality, such as predictive text, to reduce or eliminate altogether the amount of typing and subsequent coding required
 - more comprehensive built-in validation within and between questions
 - specific design of questions for a digital platform
- Research and assessment of the paper questionnaire used in 2011 census informed the design and development of the paper questionnaire for 2022.
- The design of the paper questionnaire takes into account good practice standards and guidance, as well as requirements to ensure quality data capture. A key challenge in designing the paper questionnaire is the limited space on each page. Features of the design of the paper questionnaire include:
 - use of fonts and font sizes that follow accessibility guidelines
 - optimal format and sizes for free-text write-in boxes to ensure quality data capture
 - use of optimal image size to maximise the use of available space while ensuring quality scanning for data capture
 - structural formatting guidelines throughout to ensure consistency and respondent understanding
 - incorporation of required barcodes to support the data collection process

A full evaluation of the questionnaire design was undertaken following the census rehearsal in 2019. A detailed [Census Rehearsal Evaluation Report](#) can be found on our website.

5.3 Census Address Register (CAR)

A high quality address list, or address register, is at the heart of the census design and operation. An address list which draws from the best national sources is essential for delivering the necessary census information to households and allows us to target follow-up those households which do not initially respond.

The address register will also underpin other processes that support the production of outputs following the census.

Specific quality assurance activities include:

- We are developing a high quality address register which we will continue to refine so it can be used in 2022. All published CAR versions are drafts and subject to constant change and ongoing improvement through a range of quality assurance activities.
- The CAR is based on administrative data. As such, a Quality Assurance of Administrative Data (QAAD) will be produced to help statistical assessors and other interested users review the quality assurance arrangements of the address register used to produce census statistics.

5.4 Enumeration Strategy

Enumeration is the process of collecting data from the people of Scotland during the census. This includes the operational aspects of census data collection: direct contact, encouragement of self-response (including digital self-response) and follow up. It also links with the public assistance strategy that aims to provide support to the people of Scotland to ensure their participation in the census.

For those living in private households, the census will target enumeration at household level, with individuals responding within those households. Any member of the household aged 16 or over can choose to use an individual questionnaire to provide answers in private. They can do this either online using an Internet Access Code or on paper.

Communal Establishments (CE) will also be part of Census 2022. A communal establishment is typically a managed residential accommodation where there is full-time

or part-time supervision of the accommodation. The person in charge of the communal establishment (referred to as the communal establishment manager) is required to complete a CE questionnaire. In addition each resident in a CE is required to complete a CE individual questionnaire.

International research suggests a postal first contact is the best and most cost-effective solution to ensure both coverage and address verification for households. We are therefore planning to post-out a letter with an Internet Access Code to every household in Scotland, which will allow them to complete their census securely online, or request a paper questionnaire. The letter will also be used as a tool to inform and encourage the public to complete the census.

Specific quality assurance activities include:

- All major household contact materials sent to respondents were trialled in the census rehearsal in 2019.
- The effectiveness of the contact materials informing people how to engage with the census and their obligations will be determined by the Census 2022 response rate and will help inform future censuses. Currently we are aiming for 94% national response rate.
 - We monitored the content of enquiries made to our contact centre during the census rehearsal and these have been used to inform the design of the census enumeration strategy for 2022.
- Publicity, marketing and communications will seek to raise awareness and maximise motivation to participate amongst all groups and communities.
 - Messaging will be tailored to different audiences using a range of platforms, including social media, marketing, advertising and media relations and will seek to educate and reassure whilst highlighting the benefits of the census, and relieving concerns around security of data.
- NRS will seek to maximise response amongst those groups who are considered to be at most risk of non-participation, by building relationships through direct engagement with their representative and support organisations, and local authorities.
 - Building on the detailed market research already undertaken, community engagement activities will seek to develop knowledge and intelligence at local levels to inform messaging and tactics. This will include local and regional prevalence of target populations and the communications

channels and networks they use. The communications team will subsequently develop appropriate communication strategies for those key audiences.

- Work to establish working stakeholder relationships to support our approach is already well underway and will grow and intensify moving forward towards 2022.
- An Equality Impact Assessment has been carried out to ensure that the census understands any impact it may have based on the protected characteristics under the Equalities Act.

5.5 Online and Paper Collection

Respondents will be able to complete the census questionnaire online, or can request a paper questionnaire for return by post.

Where we do not receive a response, as with previous censuses, we will use different strategies to contact addresses and encourage the residents to take part. The non-response follow-up operation will start four days after Census Day and will take the form of targeted reminder letters and visits by enumerators.

Specific quality assurance activities include:

- Understanding our respondents and how they wish to interact with the census collection exercise is key to achieving the maximum response rate and work is currently underway on this in the form of user, usability and accessibility testing.
- We will continue to undertake user testing to ensure that a robust design for our online collection is developed, which both improves data quality and reduces any unnecessary burden on respondents.
- Live webchat has been incorporated into our online service so respondents can seek support directly from our contact centre whilst completing their census online. Public assistance will also be available via telephone, post, face-to-face and social media channels. Gaelic and British Sign Language (BSL) language support will be put in place as well. In addition to this, relay service (text and video), telephone data capture, accessible internet pages that can be used by screen readers, large print questionnaires will be provided. We will remain open to further developments in all these areas and will continue to monitor developments in other countries and across other services.

A full evaluation of the online collection and paper capture procedures was undertaken following the Census rehearsal in 2019.

- On the back of the rehearsal, changes and improvements have been made in online questionnaire . For example, introduction of drop down menus and online validation to support and increase the quality of input.
- As part of quality assurance activities, Collect Service is also undergoing digital first assessment process.

5.6 Field Force Strategy

A large field force will seek to ensure every household and communal establishment is able to participate in the census. There will be targeted follow up of non-responding households.

All non-responding households will receive at least one follow up visit to maximise response and ensure the status of vacant or derelict addresses can be resolved.

Specific quality assurance activities include:

- We will refine and target the follow-up operation, deploying more resource to those areas with lower response rates whilst still ensuring all non-responding addresses are visited.
- Online collection return rates will also enable an up-to-date information about the location of populations that have not responded. This could be used to target communications or community engagement activities further.
- The prioritisation of resource in the field, communications and community engagement could be used as a key driver for ensuring a high response rate and quality of the data collected.

5.7 Census Coverage Survey

To account for people and households who may not have been counted by the census, statistical methodology is used to estimate and adjust the results. This methodology is

used to identify the number of people and households affected and to adjust the 2022 Census estimates accordingly. The Census Coverage Survey (CCS) is a vital part of this process.

The CCS for 2022 will take place six weeks after the census day. This independent voluntary doorstep survey will aim to survey around 53,000 households (114,000 individuals) from across Scotland collecting information about people and households. These data will then be matched to census records and used to estimate the size of the census under- or overcount.

Specific quality assurance activities include:

- The CCS is vital to the census and the data collected follows the same data journey and quality assurance checks as the main census data.

5.8 Census Quality Survey

Historically the Census Quality Survey (CQS) was a voluntary survey carried out across Scotland eight weeks after the census day. It aimed to capture the differences between responses in Census compared to the CQS.

However, for Scotland's Census 2022 we have changed our approach and we will not carry out a CQS in the traditional way. Instead we will use alternative data quality measures which offer robust and extensive assurance that Scotland's Census 2022 produces high quality data and is of value to the people of Scotland. We have developed additional quality assurance processes that significantly improves our approach to ensure Census data is high quality. The new approach we have adopted relies on the robustness, quality and value of existing statistical quality assurance processes.

5.9 Load and Validation

Census data comes in to NRS at regular intervals from when the Internet Access Codes (IACs) are sent out until six weeks after census day.

Specific quality assurance activities include:

- Ensuring the data received is in the correct format and adheres to pre-determined specifications and processing rules.

- These processes have been evaluated in the 2019 rehearsal and captured in rehearsal evaluation.

5.10 Coding

Coding assigns each response from a census question to a particular code that can be processed to produce census outputs.

Returned paper questionnaires need to be scanned, so that the paper responses can be captured as digital data first to allow coding. Paper and online questionnaires will be automatically coded where possible. However, values from some paper and online responses will need coded manually. Both automatic and manual coding processes risk introducing error.

Specific quality assurance activities include:

- Checking a sample of returns to understand the rate of error in capture and coding.
- Assessing the completeness of the coding lists (e.g. how many text answers can be automatically assigned to an existing code).
- A percentage of automatically coded records will be sent to manual coding for quality assurance, and checks for discrepancies in manual coding will occur.
- Manual coding will be quality assured, potentially by having the same entry coded by multiple staff to assess accuracy.
- Checks for systematic errors (e.g. 'major' in job title being interpreted as military officer when not appropriate).
- Distribution of responses to each question and variable will be monitored throughout the coding period and compared to expected response rates based on comparator data sources, including the mid-year population estimates and the 2011 census.

5.11 Remove False Persons and Resolve Multiple Responses

These processes seek to identify and resolve census returns which relate to a person already correctly covered on another return from the same or very close location – Remove Multiple Responses (RMR). Or to remove responses which do not contain sufficient information to be treated as a response – Remove False Persons (RFP).

Specific quality assurance activities for RFP include:

- Analysis of the number of missing variables for accepted and rejected records as records containing information about a genuine persons. Examination of accepted records with little data or rejected records with missing fields.
- Checks on the operation of the process by examining records that should have passed and failed the rule and ensuring that they were treated correctly.

Specific quality assurance activities for RMR include:

- Clerical review of identified cases of duplicate responses.
- Analysis of the effects of merging records, including summary statistics and distributions comparisons.

We are exploring the use of administrative data to help determine if a record is a false person or a multiple record that can be removed from the data set. The methodologies for using administrative data for quality assurance purposes will be published in due course.

5.12 Name Re-ordering

For a census return, respondents are asked to add people to the household questionnaire in the same order that people will complete their individual questions. However, sometimes this does not happen. This is mostly relevant to the paper questionnaire responses, as the online questionnaire has built-in functionality to automatically include the individual names throughout the questionnaire.

This process is used to identify and correct where names do not appear in the same order in the household and individual questions. This improves the data quality for further data processing steps.

Specific quality assurance activities include:

- Clerical review of identified cases of inconsistent ordering, to check they have been identified correctly.
- Analysis of the effect of changes to ordering (to ensure that changes made do not negatively impact data quality elsewhere).

5.13 Filter Rules

Filter Rules is a process which resolves issues in the answering path (routing) of a paper questionnaire. For example, not everyone is required to answer every question on the census. However, a number of people do so, either because they wish to answer a question, or because they miss following guidance appropriately. The extra answers may cause issues with later statistical processing. Filter Rules process identifies and where possible resolves these issues to reinforce the questionnaire routing. Additionally, the online questionnaire has automatic routing built in. For example, age routing, which will route respondents to the questions they need to answer according to their age.

Filter Rules may also address inconsistent information that we find in groups of answers, which can also cause issues with later statistical processing. These are corrected taking into account other answers in the questionnaire.

Specific quality assurance activities include:

- Assessing the performance of the overall process using the overall counts of the records identified for filter rules application
- Analysis of a random sample of all records to determine if the rules appear to preserve the original characteristics of records and whether the filters make sensible changes.

5.14 Edit and Imputation

The Edit and Imputation process is designed to produce a dataset that is internally consistent, and has no relevant missing values for any returned questionnaire.

Item imputation is the process of imputing for a missing, invalid or inconsistent value on a census return, and is conducted in the census using, primarily, nearest-neighbour hot-

deck donor-imputation methodology. Quality issues around item imputation include predictive accuracy (how accurate is the method in imputing the correct value for an individual); distributional accuracy (how accurate is the method in reflecting the true multivariate distributions of variables) and 'spikes' – where the same donor is selected for multiple cases of missing values when that donor has rare characteristics.

Specific quality assurance activities include:

- Internal and external assurance panels to quality assure methodology.
- Distribution checks of the imputed variables before and after the imputation process.
- Sample check of the imputed variables to ensure the process does not produce any implausible record combinations.

5.15 Census to Census Linking and Census Coverage Survey to Census Linking

Sometimes people are recorded as usually resident at multiple addresses, or at a wrong address. This can lead to an overcount of population. Census to Census Linking helps us to identify and estimate this overcount. This is then used in the Estimation process to correct the population estimates.

To account for people and households who may not have been counted by the census, which can lead to an undercount of population, census to Census Coverage Survey (CCS) linking is applied. This identifies the persons and households that are captured only in the census, only in the CCS, or in both. The quality of the linking is important to produce accurate estimates.

Specific quality assurance activities include:

- Link census data to itself and other sources to compare distributions and assess quality issues arising from people being counted at different addresses, or at a wrong address. This especially is relevant for questions on address one year ago and visitors.

- Linking census returns to the Census Coverage Survey (CCS) – this will only cover a subset of the Census questions, however will inform the Estimation and Adjustment process.

5.16 Estimation

Estimation involves estimating the level of undercount in Scotland's Census. Dual System Estimation (DSE) is the method used to estimate the total population in a sample area by considering the number of people recorded by:

A: the census

B: the Census Coverage Survey (CCS), and

C: both the census and Census Coverage Survey.

The dual system estimate (the estimate of the total population including people missed by both the census and the CCS) is calculated by $A*B/C$ (A times B, divided by C). Dual System Estimation is also known as capture-recapture.

The process to estimate the number of households and their population consists of three stages: the application of Dual System Estimation (DSE) to estimate the level of undercount in CCS areas; the extension of this to non-sampled areas using ratio estimation; and the use of small area modelling to derive Council Area totals. A separate, simpler methodology is used to estimate the population of communal establishments.

While the estimates are primarily for undercount within the Census, there are cases where people are counted more than once or in the wrong place which give rise to overestimation. Dependence between responding to the census and CCS results in underestimation of the population. The level of overcount and dependence must be identified and measured, to allow a correction to the estimates.

Specific quality assurance activities include:

- The DSEs being assessed for any bias at household level using an alternative household estimate from the census process.
- The sample being assessed for balance, which would affect the ratio estimator, using the placeholder data from the census process.

- Using administrative data to assess undercoverage in communal establishments.

5.17 Adjustment

Whilst Estimation (section 4.16) produces overall population and household estimates, Adjustment creates new records for the missed population (including records for people in communal establishments). It creates records for those people missed in order to get a complete census dataset. Adjustment process predicts the likelihood of the type of persons or households being missed in census, and creates skeleton household and person records to represent the 'missed' records. These then are passed onto Post Estimation Edit and Imputation process to fill in the blanks for the remaining variables using donor imputation methodology. Estimation and Adjustment combined gives a census dataset adjusted for the whole population.

Specific quality assurance activities include:

- To ensure the population estimates are plausible after the estimation and adjustment processes, all of the population estimates will be quality assured using:
 - demographic analysis;
 - survey data;
 - qualitative information;
 - administrative data; and
 - local information.

See Validation of Population Estimates (section 4.18) and Demographic and Topic-based Quality Assurance (section 4.19) for more information about the comparator data sources.

5.18 Validation of Population Estimates and Topic-based Quality Assurance

This strand of work ensures that the main demographic statistics are plausible for Scotland and Council Areas, as well as other geographies and statistical outputs. Topic-based analysis is to examine the census results relating to each topic covered in the census (for example, housing or health) to check for any evidence of systematic errors in the results for that or related topics, and to understand, as far as possible, the reasons for any differences between the census results and other sources of information on that topic.

We aim to consider the likely accuracy of the census data outputs and will focus on validating the estimates for:

- National and Council Area population counts by age and sex; and
- Validate population estimates for all topics asked in the census with a focus on variables that are new to 2022;
- Use comparator data sources and topic experts to quality assure topics such as demography, housing, labour market and health to assess the provisional census results in the context of other evidence and trends.

Specific quality assurance activities include:

- Using a suite of tools and methods that will allow us to focus on geographic areas and population groups or topics where there are inconsistencies or need for further analysis.
- In 2022, we will aim to make best use of all sources available, including:
 - 2011 Census;
 - mid-year population estimates;
 - administrative; and
 - survey.
- The evidence to support our population and topic-based estimates will be reviewed by internal and external assurance panels who will advise on whether estimates are fit for purpose or require further work or adjustment.

To help assess the quality of our population estimates we can also use demographic indicators based on 2022 Census and compare with indicators produced using other sources, including mid-year population estimates. These analyses will include:

- Sex ratios – the ratio of the number of men per 100 women in the census population estimates checked by single year of age to identify any implausible values.
- Fertility rates – calculated based on census population estimates of women aged 15-44. Unusually high or low fertility rates could be compared to rates calculated

from mid-year population estimates and other sources, with further analysis of variation within Council Areas.

- Mortality rates – age-standardised mortality rates calculated using census population estimates and comparing to equivalent rates calculated using mid-year population estimates and further analysis carried out on age-specific mortality rates based on census population estimates.
- Comparison of census-based estimates with official migration estimates.

5.19 Derived Variables Created for Outputs

A primary variable is a variable that has been created directly from the source material: either from a single question on the census or Census Coverage Survey questionnaire⁶, or from a single piece of metadata such as capture mode.

A derived variable is created using a mix of other variables. Derived variables are created throughout the data processing journey and used in processing activities, whilst other derived variables are created for the production of census outputs.

Specific quality assurance activities include:

- Each derived variable will be tested using methods including:
 - analysis of the distribution of values derived and comparison with other sources
 - tabulations
 - distribution checks
 - sample checks of derived variable outputs;
- The logic for creating combinations and groupings for each derived variable will be tested prior to receiving the census data as part of methodology development.

5.20 Output Areas

Output Areas are a geography used in census outputs. Output Areas are made from postcode and census data and used as building blocks for other census geographies. They must contain at least 20 households and at least 50 people in total. Some Output Areas will change between censuses due to postcode boundary changes or population

and household changes. Output Areas nest exactly into council areas and use split postcodes to do this.

Specific quality assurance activities include:

- Whilst provisional 2022 Output Area boundaries will be derived before the census data is collected, the final boundaries will be derived by the NRS Geography team using the Census dataset following coverage adjustment.
- We aim to quality check that the final Output Areas are created with the correct numbers of households and people, and check the shape of each Output Area boundary.

5.21 Statistical Disclosure Control

Statistical Disclosure Control (SDC) is all methods applied to census outputs to protect the privacy of personal information. It includes making small changes to data, controlling access to data, and controlling the level of detail that is available to census data users.

There are two main methods of SDC for Scotland's Census 2022: record swapping and key cell perturbation.

5.21.1 Record Swapping

Record Swapping involves swapping the geographical locations of records. For example, household 1 in area A is swapped with household 2 in area B. In any published data the information from household 1 will be in area B.

Specific quality assurance activities include:

- We can estimate the expected scale of the error introduced by record swapping by looking at the evidence from 2011 in conjunction with any change in the parameters and rules being applied in this process.
- The approach in 2022 is expected to be a combination of the above and a simple comparison of the adjusted figures with the unadjusted figures.

5.21.2 Cell Key Perturbation

Cell Key Perturbation introduces small adjustments to cells within output tables.

Specific quality assurance activities include:

- We can confidently estimate the expected scale of error introduced by the post-tabulation adjustment from first principles – that is, derived mathematically from the properties of the perturbation algorithm.
- As with Record Swapping, the quality assurance approach in 2022 is expected to be a combination of the above and a simple comparison of the adjusted figures for a particular set of tables with the unadjusted figures.
- We will check that cell key perturbation is applied consistently. We can do this by creating the same table multiple times to check that perturbation is applied the same each time. Also we will ensure that the perturbation is consistent across tables (for example, collapsing some categories, or changing age breakdowns should not affect the perturbed tables).

5.22 Census Outputs

5.22.1 Pre-Built Tables

Pre-built tables are built by census staff so users can view and use key census data without having to build their own data table.

Specific quality assurance activities include:

- Checking that cell counts produced using the flexible table builder will be compared (for all or selected areas) with counts produced using an independent system. The primary purpose of this is not to check that the tabulation tool can aggregate data correctly, but to help identify any errors or ambiguities in the coding of tables (for example, in presenting non-standard classifications).
- Comparing the tabular totals of pre-built tables with each other where populations are the same. For example, checking the tabular totals for the national population, population aged 3+, population aged 16-74, and so on. There will be a comparison sheet created with key population and household numbers to do this.

This will be applied to all outputs releases.

5.22.2 Geographies and Totals

The main building bricks for census geographical areas are Output Areas (OAs). All higher geographies or area types are built from OAs. Any area for which a Census output is produced is the aggregation of OAs that approximate best to the area. OAs will aggregate exactly to a Council Area, but not necessarily to any other type of area.

Specific quality assurance activities include:

- There will be several area and geography checks implemented during the creation of Census Outputs which might include:
 - checking all geographies are correctly nested
 - checking there are no rogue records in the wrong Council Areas, but correct Output Area
 - that totals for each geography sum into each other. For example, if you sum up all Council Areas, does this add to the Scotland total.

5.22.3 Additional Outputs

Alongside the pre-built tables, we will produce a number of other outputs and supporting information to accompany the key census outputs for 2022. These include metadata, analytical reports, and the content that will be published on our outputs website.

Specific quality assurance activities include:

- Internal peer-review of supporting materials and website content before publication.

5.22.4 Origin-Destination

Origin destination statistics are census data that deal with movement or flow of people; either as migration (from their address one year prior to the census) or travel to work or study (from their current address to their workplace address or place of study). These flows can be cross-tabulated by other variables of interest (for example, method of travel). Much of the origin destination data from the 2011 Census was [published at the UK level](#), providing flows for usual residents of Scotland, England, Wales and Northern Ireland.

Specific quality assurance activities include:

- Introducing age routing for zero to one year olds in the online questionnaire to increase the quality of data collected.
- Question design changes from 2011 for address of place of work or study questions to mitigate respondent misunderstanding of the question.
- Online questionnaire will have a postcode look-up function to help respondents select the correct address.
- Checks for implausible distances and methods of transport.
- Checks that the work or study addresses provided by respondents on the census match the registered addresses for their places of work or study.

5.23 Management Information

Access to detailed, high quality and timely management information is critical to ensure measures of key performances can be monitored and appropriate operational decisions can be made during the live census operation.

As part of the Data Collection Management Information (MI) project, we are aiming to develop a framework to collect, analyse and disseminate the MI data necessary within NRS so that managers and decision makers have access to the most accurate, timely and relevant information throughout the collection period to make the live operation a success.

Daily MI analysis and internal briefing was a vital feature of the [2019 census rehearsal](#). However, it was mostly done manually during the rehearsal. It was noted that this approach would not scale well to provide the volume of information needed for the census in 2022. Hence, a recommendation was made to evaluate how best to present management information that meets the requirements of the intended audiences.

Lessons learned from the rehearsal guided the development of the core Strategic Objective for the Management Information Project, to create a Reproducible Analytical Pipeline (RAP) for MI reporting for live census.

There will be a MI system put in place which will be able to pull management data from across the various levels of the operation and collate it in a useful, meaningful and concise way, meeting both the pre-defined and ad hoc needs of the broadest range of internal audiences possible. The management data will be aggregated counts and will not include personal data. The system will be designed to measure and monitor Key Performance Indicators related to response and data collection, both online and paper, as well as operations in the field.

Some areas where most important MI will be generated and used to assess operational performance and direct change are listed below.

- Enumeration:
 - Return Rates and Variability at various levels of geography.
 - Projected overall Return Rates based on current operational situation.
- Public Assistance:
 - Daily Contact Centre volumes (phone calls, email messages, webchat).
- Logistics and Print:
 - Daily paper questionnaire requests
- Census Coverage Survey:
 - Number of interviews and address listing tasks actioned each day.
 - Communal Establishment visits.
- Field Force:
 - Number and proportion of Prioritised Address List (PAL) visit tasks completed each day.

There will be many different internal audiences for management information during the live operation and beyond. This will necessitate the production of different reports and dashboards, both on a regular and ad-hoc basis, based on the needs of its users. Maximising the user experience of MI Reports and Dashboards will be fundamental to this work.

6. Stakeholder engagement

The Statistical Quality Assurance (SQA) team will seek the views of stakeholders in the development of our proposals for the SQA of 2022 Census data. Input from stakeholders will ensure that the quality assurance carried out will reflect good practice and include, where possible, the checks that other stakeholders would plan to conduct. This will also ensure that a built-in part of the quality assurance is to make available the evidence required by stakeholders to understand the reliability of the census results.

In particular, ways in which we have and will seek stakeholder views include:

- Developed plans in conjunction with other Census teams to ensure that the plans are feasible given the data and infrastructure likely to be available.
- Sought comment on early proposals from the Census Internal Peer Review Group, and the Census Design Authority (CDA).

- Sought comment on the proposed approach from the external assurance panels, which will include methodological and demographic specialists from inside and outside NRS.
- Participated as appropriate in stakeholder activities led by the wider census programme.
- Collaborating with a number of Local Authorities to understand their requirements and issues around the quality of census data.
- Inviting comments from all stakeholders on this published strategy for quality assurance.

Proposals developed in the light of the above, along with experience from the [2019 census rehearsal](#) and other developments in NRS, will be formally signed off at the appropriate level within Scotland's Census 2022 Programme.