Scotland's Census 2022

Validation of Population Estimates

Methodology

November 2021

Contents

1.   Plain English Abstract

Scotland's Census is the official count of every person and household in the country. Data is also collected on characteristics of Scotland's people and homes. The data collected are processed carefully to ensure they accurately reflect the makeup of Scotland's population. Prior to release of the results of the census to the public, we will compare the Census data to other sources to ensure counts are plausible and any discrepancies are accounted for. This document provides detail on this process, known as Validation of Population Estimates.

2.   Abstract

Scotland's Census is the official count of every person and household in the country, conducted by the Register General for Scotland who heads a non-ministerial government department called National Records of Scotland (NRS). Scotland's next Census will be taken on Sunday 20 March 2022.

Scotland's Census 2022 will produce population data for Scotland. The population data will be available at national and sub-national levels of geography, including Council Areas and Data Zones. Census 2022 will also produce population statistics on a number of individual and household topics such as demographics, qualifications and employment, household occupancy and communal establishment characteristics. Full details of all Census 2022 topics and questions can be found at Scotland's Census website: www.scotlandscensus.gov.uk.

Statistical quality assurance will be carried out to improve the quality of census data. To ensure that the population totals are plausible, Validation of Population Estimates (VoPE) will be part of this process. VoPE will ensure that the national and sub-national estimates are plausible when compared with data from survey, administrative, and demographic comparator sources. VoPE occurs at the end of the census data journey but prior to statistical disclosure control and outputs.

This paper gives a high-level description of the proposed VoPE methodology.

3.   Background and Introduction

Scotland's Census 2022 will produce population data for Scotland at national and sub-national levels. These data will cover a range of topics for individuals and households. The Validation of Population Estimates (VoPE) process will compare census data with existing data, identifying any areas where census estimates differ significantly from prior data. At this point in the quality assurance process of the 2022 Census, data are expected to be high quality and, therefore, VoPE is a final validation check. Where discrepancies do arise these will be verified so that the difference is adequately explained and if not, investigated further. Census data is considered the gold standard for data on Scotland's population as the census will have near complete coverage of the population and census data will go through a vigorous processing and quality assurance journey prior to publication.

Figure 1. VoPE process within the census data journey
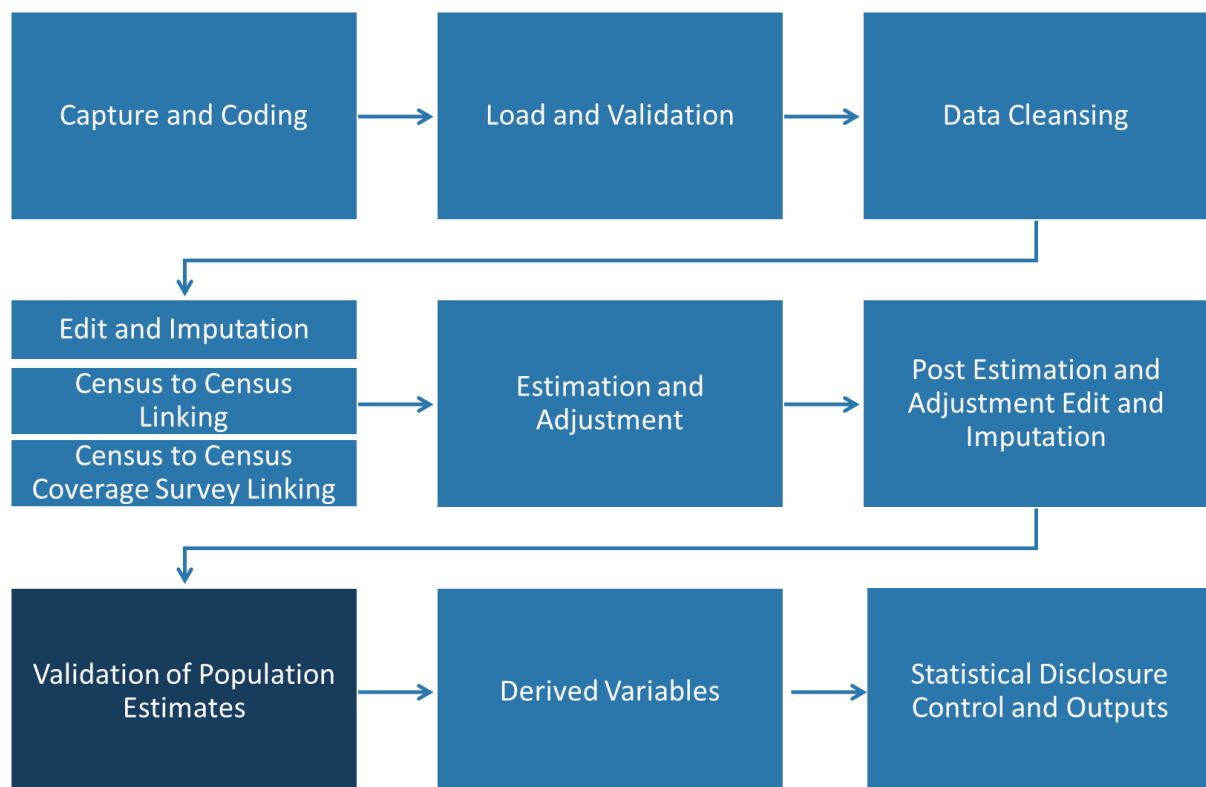
VoPE planning, preparation and execution process is proposed as follows:

Planning and Preparation

- Identify populations and topic estimates to be validated.

- For all these populations, determine whether validation is reasonably possible. Where not possible, explanation will be given.

- Identify suitable comparator data for each population.

- Present planned VoPE process to internal and external review groups for feedback.

- Consult with data owners as necessary, and identify the specific data that will be required for VoPE.

- Draw up the necessary governance documents, such as Data Protection Impact Assessments, etc.

- Create Quality Assurance of Administrative Data (QAADs) documents on datasets used to perform VoPE.

Execution

- Comparator data is shared with NRS and stored in a secure IT area.

- Census data is made available for VoPE checks.

- Validation checks are carried out for the specified populations, returning either verification that the census data is expected given the comparator data, or returning an indication that further work may be undertaken to verify that the change in population has a reasonable explanation. This explanation may be communicated to users if necessary.

- Where a reasonable explanation is not found, further investigation may be undertaken on that population.

Reasonable explanations are known factors that will influence the census populations, but have a justified reason for not appearing in the comparator data. For example, construction or demolition of large housing areas after the mid-year population estimates (used as a comparator data source) but before census could lead to discrepancies in local area populations.

A range of datasets will be used as comparator data sources. Conversations with data suppliers and information from any official statistics sources will help with the understanding of the quality of these datasets. Knowing the strengths and limitations

will help when making decisions on the VoPE checks. We do not expect the census and comparator data to match exactly as comparator data sources will not reflect the same point in time as the census and may have limitations due to being produced primarily for administrative purposes or be drawn from a much smaller sample of the population, but will be used as a guide for expected results from the census.

The characteristics of the census data to be checked and the comparator data will guide the type of check undertaken. There are three types of checks proposed, suitable for different circumstances.

1. **Range check:** Does the estimate lie within an acceptable range of the comparator data? This is suitable when the comparator data gives a good coverage of the same population as the census data. Taking into account the quality of the comparator data, suitable ranges can be constructed. For example, Pupil Census covers all pupils attending local authority and grant-aided schools in Scotland, this gives good coverage of children aged 6 to 15. This does not cover home schooling and independent schools. The comparison to the census would take this into account, as there should be a slight undercount in some council areas due to independent schools not included in the comparator data.

2. **Proportion check:** Are the distribution and/or proportions present in the census data similar to those in the comparator? This is suitable in cases where the comparator data are based on a sample of the population and therefore do not give full coverage at a national level. While the national population is not covered, for a characteristic measured by these datasets the proportion of responses can be taken as a reflection of the nation. For example, Scottish Survey Core Questions (SSCQ) asks a question on religion and publishes percentage results for the different response options with further breakdown by age band. The census data on religion can be validated by comparing the proportions of responses in the census data to those published in the SSCQ. Survey data is published with confidence intervals associated with their estimates and these will be used as well to bracket acceptable proportions.

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

3. **Ad-hoc comparison:** For some topics to be checked by VoPE, the available comparator data do not contain the level of detail required for the methods above. These cases are generally where the question is newly included in the census, or where methodological issues have been identified in previous data collections. These topics include language proficiency (English, Scots, Scottish Gaelic), British Sign Language use, trans status and history, and smaller ethnic groups. For these topics, an individual VoPE approach will be created using the data available. Taking into account the coverage and precision of the comparator data, a meaningful verification of the census data can be made.

Dependent on the census data to be checked and the comparator data sources, one or more of the above methods may be used. The desired outcome is to verify that the census data are as expected given the comparator data, identifying any areas that differ significantly.

If there are any topic areas or populations where there is a larger than expected difference between the census data and the comparator data, we will ensure that the relevant information is included in the report on data quality that will accompany the census outputs. Depending on the nature and scale of the difference, it may also be highlighted more prominently to users following further investigation.

3.1   Order that census populations and topics will be validated

Population estimates will be validated in order of priority. A prioritisation scheme was made to ensure headline figures can be validated prior to first outputs, scheduled for one year after census day (i.e., March 2023).

Priority One: Total population coverage

These checks will validate the total size of the population at national and subnational levels.

- Population size estimates at Scotland, Council Area and Data Zone level by age and by sex
- Household counts and household size estimates at Scotland, Council Area and Data Zone level

Priority Two: Population sub-groups

- School pupil population
- Ethnic group population breakdown
- Armed forces population
- Prison population
- Children's homes population
- Care homes population
- Student population
- Gypsy/Traveller population
- Populations within hostels and shelter for the homeless and homeless people sleeping rough

We will also validate topics included in the Hard to Count index[1], an index that ranks planning areas by combining the ranks of four variables, with low ranking corresponding to harder to count areas. The four variables are:

1) proportion of the population that is 16 to 29 years old (covered by population validation in priority one);

2) proportion of school pupils where English is an additional language;

3) proportion of dwellings (occupied or not) that are flats;

4) Scottish Index of Multiple Deprivation (SIMD).

For the latter variable, we will correlate population counts with SIMD rank for that area.

Priority Three: Protected characteristics

Checks on these topic areas provide additional validation on protected characteristics data and/or special category personal data. Age, sex and ethnic group will already be covered in priority groups one and two.

- Marital or civil partnership status
- Trans status or history
- Sexual orientation
- Long-term conditions
- Long-term health problem or disability
- Religion

---

[1] For more information about the Hard to Count index see PMP004: Developing a hard to count index | Scotland's Census (scotlandscensus.gov.uk)

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

Priority Four: Remaining individual and household topics

The remaining individual and household topic questions will be validated following the other priority topics, prioritised depending on the order of further releases of census data outputs.

Health
- General health

Household relationships
- Household and family relationships
- Address one year ago

Housing
- Tenure
- Landlord
- Number of bedrooms
- Type of central heating

Education
- Qualifications held

Identity and Language
- Country of birth
- Date of arrival in the UK
- English language proficiency
- Scottish Gaelic language skills
- Scots language skills
- British Sign Language use
- National identity

Labour market
- Ex-service status
- Unpaid care
- Occupation

Economic activity
- Economic activity: employment status or reason for unemployment
- Supervisory status
- Industry
- Ever worked
- Hours worked
- Looking for or available to work
- National Statistics Socio-economic Classification (NS-SeC)
-

Transport
- Cars or vans
- Method of transport to work or study

Topic-based checks will not be performed for the following groups either because they are derived variables, which means data have been created from the input variables that were validated in the earlier VoPE process, or due to lack of suitable comparator data. These include:

- Names of people usually living at address
- Visitors
- Self-contained accommodation
- Address or place of work or study
- Passports held

3.2   Method used in 2011

For details of how population estimates were validated for the census in 2011, see: [2011 Census in Scotland: Population and Household Estimates: Quality Assurance Process.](#) The process was similar to 2022 with a range of comparator data sources used to validate population estimates at the national, Council Area and Data Zone levels as well as specific topics where appropriate.

3.3   Methods used by ONS and NISRA

Office for National Statistics (ONS)[2] and the Northern Ireland Statistics and Research Agency (NISRA)[3] are responsible for planning and carrying out the census for England and Wales, and Northern Ireland respectively.

ONS has published its [strategy for quality assurance](#) online as has [NISRA](#). Both strategies contain many of the same strands as detailed below for Scotland's Census. Validation of population estimates at the national and Council Area levels by age and by sex are the main focus of all three quality assurance strategies. All three strategies will make use of a range of comparator data sources and will use quality assurance panels made up of area and/or topic experts to help understand any anomalies identified. Following validation of headline population estimates, all

---

[2] For more information see [Census - Office for National Statistics (ons.gov.uk)](#)
[3] For more information see [Census | Northern Ireland Statistics and Research Agency (nisra.gov.uk)](#)

census topics will be validated using additional comparator data sources and expert panels.

4.   Proposed Method for 2022

This section describes the proposed method for Validation of Population Estimates for Census 2022. Table 1 gives a list of all comparator data sources and the census topics that they will be used to validate.

Table 1. Comparator sources used in Validation of Population Estimates for Scotland's Census 2022

Comparator Data Sources	Census topics to be validated
Mid-year population estimates	Population counts
Centenarians (and people aged 90 and over)	Population counts
Small area population estimates	Population counts
Census 2011	Population counts
National Health Service Central Register (NHSCR)	Population counts
Further Education Statistics (FES)	Student status, student term-time address
Higher Education Statistics Agency (HESA)	Student status, student term-time address
SG Pupil Census	School pupil population, ethnicity, main home language, Gaelic language skills, English language proficiency, long-term health conditions
SCIS School Census	School pupil population
Ministry of Defence statistics	Armed forces population

Scottish Prison Service - population statistics	Prison population
Children's Social Work Statistics	Children's home population
Care Home Census	Care home population
Estimates of Households and Dwellings in Scotland	Number of households, household size, accommodation type
Small Area Statistics on Households and Dwellings	Number of households, household size
Vital Events data	Fertility and mortality rates (as a check for population count anomalies)
Communal Establishment Register (CER)	Student halls of residence
Scottish Survey Core Questions	Ethnicity, marital/civil partnership status, sexual orientation, religion, any conditions that limit day-to-day activities, general health, tenure, landlord, unpaid care, number of cars or vans per household
Scottish Household Survey	Address one year ago, number of bedrooms, supervisory status, hours worked
Scottish House Condition Survey	Central heating
Scottish Health Survey	Nature of long-term health conditions, general health

	for population under 18
Transport and Travel in Scotland statistics	Method of transport to work or study
Labour Force Survey	Main home language
Annual Population Survey	Qualifications held, country of birth, national identity, occupation, economic activity, ever worked, looking for or available for work, NS-SeC
ONS Business Register and Employment Survey	Industry
Annual Survey of Hours and Earnings	Hours worked

4.1   Priority One: Total population coverage

4.1.1   Validation of population estimates for geographic areas

The VoPE process for geographic areas will validate the census data produced for Scotland's population at the national and Council Area levels, broken down by age and by sex. The main comparator data source will be the Mid-Year Population Estimates produced annually by NRS. NRS have agreed to produce a bespoke population estimate for census day taking account of births, deaths and migration that have occurred since their last official estimate produced for June 2021. In addition, data produced by NRS on the number of centenarians and people aged over 90 will be used to validate the national population data for this age group. These data will not correspond to the bespoke census day estimates but to the last published data in June 2021.

In order to validate the national and Council Area population data by age and by sex, population counts from Census 2022 and MYE will be compared at Council Area levels for:

a) the whole area population

b) the area population by age

c) the area population by sex, and

d) the area population by age and by sex.

We will also validate population counts at the Data Zone level using the NRS Small Area Population Estimates, using estimates produced in June 2021. The reasoning for this approach is that discrepancies in small areas could be hidden at Council Area and at National level, therefore, small areas should be examined. The overall census population estimate for each Data Zone will be compared with the NRS Small Area Population Estimates, with further comparison by age and sex taking place for areas with relatively large differences between the two sources in total population.

In addition, the NRS centenarians dataset will be used to compare population counts for single year of age from 90 to 104 and then 105 and over by sex. These estimates are produced using a different method to the Mid-Year Population Estimates, but are consistent with those statistics for the total population aged 90 and over.

The COVID-19 pandemic has also had an impact on migration estimates used in the MYE, with the International Passenger Survey no longer being used and administrative data sources being used instead. The methodology for this is still under development by the Office for National Statistics.

In addition, we will also use the NHS Central Register (NHSCR) to validate the census data produced for Scotland's population at the national and Council Area levels. This analysis will take into account ages where there is known to be an undercount in the number of people on the NHSCR, e.g., among young males who are known to be less likely to register with a GP.

We will also look at sex ratios at the national and Council Area levels and at implied fertility and mortality rates. Anomalies identified in these analyses (e.g., single year spikes) may help identify underlying issues with the population data. The latter analyses will make use of data from the NRS Vital Events births and deaths datasets.

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

4.1.2   Household estimates

The comparator data source for census estimates of the total number of households and household size (i.e., number of usual residents) will be the NRS Estimates of Households and Dwellings for the national and Council Area levels. We will also validate number of households at the Data Zone level using the Small Area Statistics on Households and Dwellings.

4.2   Priority Two: Population sub-groups

The next populations and topics to be validated include those of particular public interest (e.g. ethnic group), populations living in communal establishments (e.g. prisons, care homes) and populations that are hard to enumerate (e.g. students).

4.2.1   School Pupil Population

The school pupil population refers to all children aged 6 to 15 and children aged 4, 5 or 16 and over if they are in school. Children aged under 4 are not counted by the census as school pupils.

The comparator data sources will be the Scottish Government (SG) Pupil Census (covering all local authority and grant aided primary, secondary and special schools), and the Scottish Council of Independent Schools (SCIS) School Census (covering independent schools).

The Pupil Census data accounts for the approximately 96% (SCIS 2019) of school pupils that attend local authority and grant aided schools. These data are collected at the individual student level and will be compared as counts at national and Council Area level by age and sex. The SCIS School Census covers the approximately 4% of students attending independent schools in Scotland. These data are published only as a national figure split into nursery, primary and secondary schools. For the purposes of school pupil validation, we will count only primary and secondary pupils from the SCIS School Census.

By using both data sources, the vast majority of school-age children in Scotland will be included in the comparison data for the 2022 Census estimate of school-aged children. Only the small number of children who are educated at home or in other settings not covered by the SG Pupil Census or SCIS School Census will not be included.

4.2.2 Ethnic Group

Census 2022 data on ethnic group will be compared with data from Scottish Survey Core Questions for the adult population. The Census 2022 data will be aggregated to match the categories available from Scottish Survey Core Questions, which are: White: Scottish, White: Other British, White: Polish, White: Other, Asian and All other ethnic groups. The comparison will be done at Council Area level overall and at national level by 10 year age band and sex.

For school-aged children, data from the SG Pupil Census will be used to compare with the Census 2022 data. This will be done by sex and detailed ethnic group at national level and by broad ethnic group at Council Area level.

There is currently no identified comparator data source for pre-school aged children.

4.2.3 Armed Forces Population

The Armed Forces population refers to all current members of the Royal Navy, the British Army, and the Royal Air Force, excluding reservists. This population is a special enumeration group in the census with armed forces base managers filling out communal establishment questionnaires for the personnel resident on their bases.

The Ministry of Defence (MoD) holds data on the serving Armed Forces population, this will be supplied for VoPE. At local level, there can be large changes in the resident armed forces population throughout the year; for example, if a unit is deployed overseas, the number of individuals present at that base can be much lower than usual. Therefore, we will consider this when comparing the Census 2022

data to the MoD data at local level, especially if there is a time lag between the two. We will validate population counts at armed forces base level.

4.2.4 Prison Population

The prison population refers to individuals living in prisons and young offenders institutions (ages 16-21) where persons are confined following criminal convictions or while on or awaiting trial. This does not include immigration removal centres and children's secure care. Convicted prisoners and prisoners on remand will be enumerated as part of the prison population if they have been resident for six months or more, otherwise they will be resident at their home address. If they have no home address, they will be included in the prison population.

The Scottish Prison Service produces prison population data with breakdowns by sex and age under 21 or 21 and over. These population data are produced from management information. The Scottish Government Justice Analytical Service receives these data and will share it with us for VoPE. This population is a special enumeration group in the census with prison managers filling out communal establishment questionnaires for their resident prisoners and prisoners given individual questionnaires for their personal details. The census figures on the population in a communal establishment type of prison will be compared with the Scottish Prison Service data by sex and age group.

4.2.5 Children's Homes

The population living in children's homes refers to all children living in a Children's Home, defined as a type of Communal Establishment providing accommodation and care for young people. This includes young people living in local authority homes, voluntary homes, residential schools, secure care units, and other residential accommodation such as private or independent residential placements for young people with complex needs.

Comparator data for this population will be provided by Children's Social Work Statistics, produced by Scottish Government, and compared to the census data on the population residing in a communal establishment type of children's home. We will validate at the national and Council Area levels by age and by sex.

4.2.6   Care Homes

The population living in care homes refers to all individuals resident in a care home with nursing, or a care home without nursing. These types of communal establishment provide accommodation to vulnerable adults. Care homes with nursing additionally provide nursing care, personal care, or personal support.

Comparator data for this population will be provided by the Care Home Census. A comparison will be made between the Census 2022 data using the communal establishment types of care home with and without nursing combined and the data from the Care Home Census at national and Council Area level.

4.2.7   Student Population

The student population refers to the following people:

- on an approved apprenticeship
- studying up to and including Advanced Higher standard or equivalent and their course includes more than 12 hours a week of study (i.e., school pupils)
- studying at a higher standard than Advanced Higher in the UK, and studying a course that:
    - lasts for at least one academic year, and
    - involves at least 24 weeks per year, and
    - involves at least 21 hours of term-time study, classes or work experience per week (which includes independent study

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

Students in their placement year of a course, sometimes called a sandwich year, and students abroad for less than 12 months in their international exchange year at an overseas university are also counted as being in full-time education.

To validate the size of the student population (i.e. all those individuals who indicated they are full-time students in the census) we will combine population estimates from the SG Pupil Census and the SCIS School Census with data from Further Education Statistics (FES, for college students and apprentices) and Higher Education Statistics (HESA, for university students). These will be used to compare with the census student population data at Council Area and Scotland levels. We are still developing the upper and lower bounds of expected difference between these sources and the Census 2022 data.

College and university students are known to be a difficult group to enumerate because they move regularly and between their home and term-time addresses. They are also often not well covered by administrative data sources such as the NHSCR if they do not register with a local GP whenever they move. We will also use the FES and HESA data separately to validate college and university student population counts.

Census 2022 data on student term-time address postcode will be validated using the Communal Establishment Register and data from FES and HESA to compare with information on where student halls of residence are located. These checks will not validate term-time addresses that are privately owned or rented properties.

4.2.8   Scotland's Gypsy/Traveller, Roma and Showperson Population

In Census 2022, groups with a mobile lifestyle and/or culture will be able to note their ethnic group as Gypsy/Traveller, Roma or Showman/Showwoman. Tick box options for Roma and Showman/Showwoman have been added for Census 2022. Testing has indicated that in Census 2011 most Roma and Showmen/Showwomen chose 'White' as their ethnic group then wrote in terms that could be linked to Roma or Showmen/Showwomen, although some chose White: Scottish or White:

Gypsy/Traveller. While our validation of ethnic group at the national and Council Area level will collapse these three groups into 'White: Other' due to the small size of their populations and level of detail available in comparator sources, we will specifically validate population size of these subgroups, if possible.

Analysis of Scotland's Census 2011 concluded that the population count of 4,212 (approx. 0.1% of the population) for the Gypsy/Traveller population was likely to be an undercount, though note that ONS recorded a similar proportion in the England and Wales in Census 2011. Groups that work with Gypsy/Traveller communities in Scotland estimated a population of 15,000-20,000 in 2012 (Amnesty international)[4] and 2013 (EHRC)[5]. This is the most recent published data on the size of the Gypsy/Traveller population in Scotland. We will validate the Census 2022 data of the total Gypsy/Traveller population in Scotland by making a direct comparison with 0.1% of the population as a lower bound and 20,000 as an approximate upper bound. No comparator data sources have been identified for Roma or Showmen/Showwomen.

4.2.9   Homeless People

Scotland's Census 2022 will collect data on a number of different communal establishments. The specific category of 'Hostel or shelter for the homeless' will include individuals who do not have a permanent settled address and who slept in temporary accommodation such as a hostel on census night. In addition, Homeless People Sleeping Rough is a further group captured by the census and refers to individuals who do not have a permanent settled address and who slept outdoors on census night. They will be enumerated if they attend a day centre where census enumerators will be present during the week prior to census day. Difficulties exist counting these two populations and there are few existing comparator data sources.

The Homelessness Monitor: Scotland (Crisis/Heriot Watt, 2019[6]) report estimated 650-800 individuals sleeping rough per night. A direct comparison check will be

---

[4] Amnesty International UK: https://www.amnesty.org.uk/
[5] Equality and Human Rights Commission (EHRC): https://www.equalityhumanrights.com/en
[6] Homelessness in Scotland 2019 | Crisis UK

made between this figure and the Census 2022 data. The Scottish Government Homelessness Statistics team are assisting in identifying further comparator sources and highlighting the difficulties in gaining an accurate estimate of the size of these populations both on census night and from any alternative sources.

4.2.10   Other Communal Establishments

There are a number of additional categories of communal establishments with low numbers of usual residents (e.g. hospitals, hotels). These populations lack robust comparator data and we do not anticipate being able to validate them, but will continue to investigate ways to check these population estimates.

4.2.11   Variables covered by the 'Hard to Count' Index

The 'Hard to Count' Index is made up of four variables, with low ranking corresponding to harder to count areas.
The four variables are:
1) proportion of dwellings (occupied or not) that are flats (corresponding to census question on accommodation type);
2) proportion of school pupils where English is an additional language (corresponding to census question on main language);
3) Scottish Index of Multiple Deprivation (SIMD) (we will validate whether population estimates vary in quality according to SIMD decile);
4) proportion of the population that is 16 to 29 years old (covered by population validation in priority one).

1) Accommodation type
Census 2022 will collect data on the type of accommodation that households reside in. The comparator data that will be used is Estimates of Households and Dwellings in Scotland. A proportion check will be used to compare the estimates of the percentage of dwellings that are flats, terraced, semi-detached or detached at Council Area and national level with the Census 2022 data.

2) Main language

Census 2022 asks about the respondents' main language. Response options are English or other with an option to write in the answer. For school-aged children the SG Pupil Census can be used to validate this topic at the national and Council Area levels, using the data on pupils' home language, grouped into English and non-English. Nationally and in the more populous Council Areas, the most common languages after English will also be able to be validated. The ONS Labour Force Survey asks a similar question (Specifically: What is your first language at home? 1 English 2 Welsh 3 Gaelic 4 Ulster Scots/Ullans 5 Other) which can be used to validate this topic for the adult population at national level. Validation of this topic may not be possible before first outputs due to the order in which census data will be processed.

3) Scottish Index of Multiple Deprivation (SIMD)

During the validation of the census population counts at 2011 Data Zone level as described in section 3.1.1, we will also aggregate the estimates by SIMD 2020 decile at national level and compare directly with the equivalent figures in the NRS Small Area Population Estimates aggregated to SIMD 2020 deciles.

4) Population of 16 to 29 year olds

Population estimates for 16 to 29 year olds will be validated as part of priority one, as described in section 3.1.1.

4.3   Priority three: Protected characteristics

4.3.1   Marital or civil partnership status

Census 2022 data on marital or civil partnership status will be aggregated to match the categories contained in the Scottish Survey Core Questions, which will be used as the comparator data source. These categories are 'Never married - single', 'Married/Civil Partnership', 'Separated', 'Divorced/Dissolved civil partnership' and 'Widowed/Bereaved civil partner' and will be compared for each Council Area using 95% confidence intervals as upper and lower bounds. The data will also be compared by 10-year age bands and sex at the national level.

4.3.2 Trans status or history

This is a new and voluntary question for Census 2022 with no established comparator data sources. A 2018 report from the Scottish Public Health Network[7] investigated a range of estimates and concluded an estimate of 0.5% of the population identified as trans. We will use this figure as a comparator, but we will not reject higher or lower estimates in the census data.

4.3.3 Sexual Orientation

The comparator source for these data will be Scottish Survey Core Questions, which contains data on sexual orientation split into 'Heterosexual' and 'LGB and Other' for each Council Area. The data are published with 95% confidence intervals. At national level, a comparison will also be made by 10-year age bands and sex. The Census 2022 question about sexual orientation is voluntary. When comparing with the Scottish Survey Core Questions data, the rate of non-response (in both the census and Scottish Survey Core Questions) will be taken into account.

4.3.4 Disability

Census 2022 will ask about long-term conditions, which have lasted, or are expected to last, at least 12 months. The data collected at this question can be used as a proxy for health conditions and disability. The percentage of responses in each response category will be compared with data from the Scottish Health Survey. A comparison of the type of longstanding illness for adults, split by 10-year age band and sex, and for children at a national level will be used to validate the Census 2022 data.

Census 2022 will also ask whether day-to-day activities are limited because of a long-term health problem or disability. Scottish Survey Core Questions contains comparable data on this for adults, which will be compared with the Census 2022 data split by 10-year age band and sex at national level and overall by Council Area.

---

[7] Health Care Needs Assessment of Gender Identity Services

The SG Pupil Census contains data on the numbers of pupils with additional support needs and information on what these needs are. Where possible, we will match these to the census categories for long-term health conditions (e.g., learning difficulty, blindness, mental health condition, etc.) as an additional check on long-term health conditions in the school-aged population.

4.3.5   Religion

Census 2022 data on religion will be compared with data from Scottish Survey Core Questions for the adult population. The comparator data contains estimates of adherents to each religion, with 95% confidence intervals at Council Area level. Categories used in the SSCQ are None, Church of Scotland, Roman Catholic, Other Christian, Muslim and Other. The census data will be aggregated to these categories so the data can be compared. A check will also be performed at national level, where more detailed data is available in Scottish Survey Core Questions, comparing by 10-year age bands and sex. When comparing with the Scottish Survey Core Questions data, the rate of non-response (in both the census and SSCQ) will be taken into account.

4.4   Priority four: remaining topics

4.4.1   General health

Census 2022 will ask about general health, with response options of 'Very good', 'Good', 'Fair', 'Bad' and 'Very bad'. Scottish Survey Core Questions contains comparable data on self-assessed general health for adults, which will be compared with the Census 2022 data split by 10-year age band and sex at national level and overall by Council Area.

Children are asked this question in the Scottish Health Survey, but not in the Scottish Household Survey or the Scottish Crime and Justice Survey (the other two surveys that provide data for the Scottish Survey Core Questions). The Scottish Health

Survey data will be compared with the Census 2022 data for the population aged under 18, split by age band and sex at national level.

4.4.2 Household and family relationships

This Census 2022 topic will provide data on the relationships between members of the same household. This data will be reported in categories, including married couple, family with stepchildren, etc.

Data from the Scottish Survey Core Questions will be used as the comparator source; the types of household categories in the Scottish Survey Core Questions: Large Adult, Single Parent, Small Family, Large Family, Single Pensioner and Older Couple. We will collapse Census categories to match those available from SSCQ compare the proportion of households in each category with those in the SSCQ. The data will be compared at national and Council Area level.

4.4.3 Address one year ago

The 2022 Census data will be compared with data from the Scottish Household Survey (SHS) on length of time at current address. The percentage of people responding to the SHS that they have been at their current address for a year or more will be compared with the percentage in the 2022 Census data that report having the same address one year ago. The comparison will be done at national and Council Area level.

4.4.4 Tenure

The Scottish Survey Core Questions contain an estimate of the percentage of households of each type of tenure, with 95% confidence intervals at Council Area and national level. The census categories will be compared to the categories available from the Scottish Survey Core Questions namely: Owned Outright, Mortgaged, Social Rented and Private Rented.

4.4.5 Landlord

Census 2022 asks 'Who is your landlord?' with options: Council (Local Authority) or Housing Association/Registered Social Landlord OR Private landlord or letting agency. The data from Scottish Survey Core Questions on tenure (described above, we will use 'Social Rented' and 'Private Rented' to match the two census categories) will be used to validate these household estimates at national and Council Area levels.

4.4.6   Number of bedrooms

Census 2022 will collect data on the number of bedrooms available for use exclusively by each household, including all rooms built or converted for use as bedrooms. The Scottish Household Survey also reports data with 95% confidence intervals on the number of bedrooms in the property, including rooms being used for other purposes and also only includes those available to the household where applicable. Therefore, the data is directly comparable with the Census 2022 data and a proportion check will be done to check the level of difference between the two sources at Council Area and national level.

4.4.7   Type of central heating

Data from three questions asked in the Scottish House Condition Survey will be used to validate the Census 2022 data on the type of central heating that the household's accommodation has. The questions in the Scottish House Condition Survey that will be used are: 'What is the main form of heating in the dwelling?', 'Extent of heating system' and 'Main heating fuel'. 'Extent of heating system' will be used to provide data on those without any central heating, while the other two questions in combination will be used to provide comparable data on to the census question categories, which are: Mains gas, Other gas (including liquid petroleum gas and biogas), Electric (including storage heating), Oil, Solid fuel (excluding wood), Wood or biomass (logs, pellets, chippings), Other renewable energy source (including electric and air heat pump systems), District or communal heat system, Other. The data will be compared at national and Council Area level.

4.4.8   Qualifications held

We will use the Annual Population Survey to validate the proportions of 16-64 year olds with qualification at or above National Vocational Qualification (NVQ) levels 1-4. We will convert the census data so that the reported categories match those in the Annual Population Survey. A comparison will be made at both the national and Council Area levels. Scottish Survey Core Questions also contains information on the proportion of the population by highest achieved qualification so this data source will be used as a second comparator data source.

4.4.9   Country of birth

NRS publishes information on the Scottish population by country of birth based on data from the Annual Population Survey. These data are available at the national and Council Area levels split into UK, EU and non-EU with 95% confidence intervals around the estimates and will be used as the primary comparator source for this question. Scottish Survey Core Questions produces estimates for Scotland's population by country of birth, with 95% confidence intervals around the estimates. The 2022 Census data will be compared with this, grouping the country of birth responses in the following categories: "Scotland", "Rest of UK", "EU" and "Rest of World" in order for the Scottish Survey Core Questions data to have a large enough sample to make a valid comparison . This will be compared overall by Council Area and at national level by 10-year age band and sex.

4.4.10   Date of arrival in the UK

We have yet to identify a robust comparator data source, but are investigating the use of Census 2011 data to validate this question.

4.4.11   English Language

This question asks about respondents' proficiency in English language and provides a grid with response options of 'Very well', 'Well', 'Not well' or 'Not at all'. The SG Pupil Census collects data on pupils with English as an additional language in the following categories: New to English, Early acquisition, Developing competence,

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

Competent, Fluent and English as 'First Language.' We will attempt to match these categories with those in the census as a rough validation of the proportions of the population with differing English language proficiency nationally and at the Council Area level. We will be mindful that the data may not be directly comparable.

4.4.12   Scottish Gaelic and Scots Language

These questions ask 'Can you understand (spoken), speak, read, write Scottish Gaelic [or Scots]' or 'No skills in either.' The SG Pupil Census records the number of students receiving Gaelic education, learning Gaelic or with no Gaelic, but it may be hard to convert this information into a robust comparator data source.

4.4.13   British Sign Language

This question asks whether respondent's can use British Sign Language (BSL). We have not been able to identify a robust comparator data source for this question.

4.4.14   National Identity

Data from the Annual Population Survey will be used to compare with Census 2022 data on the national identity of respondents. The data will be compared using the categories reported in the Annual Population Survey: Scottish, British, Irish, English, Welsh and Other. An overall comparison will be made at national and Council Area level.

4.4.15   Ex-service status

A new question is being added to three Scottish surveys - the Scottish Household Survey, Scottish Health Survey and the Scottish Crime and Justice Survey. The question will be Have you previously served in the UK Armed Forces? Data from these surveys should be available in early 2023 and we will use these data to validate the corresponding census question at the Scotland and Council Area levels.

4.4.16   Unpaid Care

Scottish Survey Core Questions contains estimates for the proportion of adults providing unpaid care in total at local authority level and the data is also available by 10-year age bands and sex at national level. The Census 2022 data on unpaid care will be compared with these categories for the adult population. For unpaid child carers, we will compare estimates to those in Census 2011.

4.4.17   Occupation/Job Title

Census 2022 data on occupation will be compared with the Annual Population Survey, which contains estimates of the number of people whose occupations are in each of the nine Standard Occupation Classification 2010 Major Groups[8] with 95% confidence intervals. This check will be performed at both national and Council Area level.

4.4.18   Economic Activity

The Annual Population Survey will be used as the comparator for Census 2022 data on economic activity. The data will be compared at national and Council Area level on the percentage of 16-64 year olds who are economically active and inactive.

The Census 2022 will also be validated by category of the economically active population. At national level, the proportion of the population classed as employees, self-employed and unemployed will be compared with the Annual Population Survey data.

A similar comparison will also be done for the economically inactive population, split into categories of students, looking after family/home, temporary sick, long-term sick, discouraged, retired and other.

4.4.19   Supervisory status

---

[8] Listed here: Standard Occupational Classification: SOC2010 | HESA

Census 2022 data on whether an individual formally supervises other employees as part of their job will be validated using Scottish Household Survey data on the proportion of individuals with supervisory roles. This topic will be validated at the National and Council Area levels.

4.4.20   Industry/Employer Activity

Census 2022 data on the industry in which individuals are employed will be compared with the ONS Business Register and Employment Survey. This survey contains estimates of the number of jobs in each industry (for businesses registered for Pay As You Earn). The data will be compared by five digit Standard Industrial Classification code[9] and by full-time and part-time employees at national and Council Area level. The ONS Business Register and Employment Survey Base is not directly comparable with the Census 2022 data as it is based on the location of the business, rather than where the employee lives. For example, if a business has 10 employees, they will all be counted in the area where the business is, even if they live in a different area. This will be taken into account when comparing the data.

4.4.21   Ever worked

Census 2022 questions on employment will allow data on whether respondents have ever worked to be produced. This data will be compared with the Annual Population Survey, specifically the Regional Employment Patterns in Scotland[10] publication. A direct proportion check will be conducted between the Census 2022 and Annual Population Survey data on the number of people who have ever worked.

4.4.22   Hours worked

Census 2022 data on hours worked will be validated using data from the Scottish Household Survey. We anticipate validating the proportion of individuals that work part-time (i.e. 30 hours or less) or full-time (i.e. 31 hours or more). We will also

---

[9] Listed here: UK SIC 2007 - Office for National Statistics (ons.gov.uk)
[10] Regional employment patterns in Scotland: statistics from the Annual Population Survey 2018 - gov.scot (www.gov.scot)

investigate using the ONS Annual Survey of Hours and Earnings as a secondary comparator data source, however this survey usually reports median and mode hours worked rather than number of part-time and full-time workers.

4.4.23 Looking for or available to work

The Annual Population Survey will be used as the comparator source for the Census 2022 data on respondents who are looking for or available for work. The Annual Population Survey has estimates of the number of economically inactive people who want a job and do not want a job. We will be able to derive the same data from the Census 2022 questions around employment. A proportion check will be made between the two sources at national and Council Area level.

4.4.24 National Statistics Socio-economic Classification (NS-SeC)

The National Statistics Socio-economic Classification (NS-SeC) is a derived variable that combines a person's occupation title with information about their employment status, whether they are employed or self-employed and whether or not they supervise other employees. It is An Office for National Statistics standard classification[11]. Full-time students are recorded in the 'full-time students' category regardless of whether they are economically active or not. Although all people aged 16 and over are assigned to a NS-SeC category, statistical outputs for this variable are generally based on the 16 to 74 years old population due to statistical disclosure control considerations. Because Census 2022 does not ask a question about the number of employees at a person's workplace, the reduced method of deriving NS-SeC (which does not require this information) is used. We will validate estimates of the number of people in each NS-SeC category at national and Council Area levels using data from the Annual Population Survey.

4.4.25 Cars or vans

---

[11] The National Statistics Socio-economic classification (NS-SEC) - Office for National Statistics (ons.gov.uk)

Census 2022 will collect information on the number of cars or vans owned, or available for use by the household. This will be directly compared with data from the Scottish Core Survey Questions which also collects information on the number of cars or vans available for use by the household. The comparator data is grouped into categories of 0, 1, 2 and 3 or more, with 95% confidence intervals around these estimates. A proportion check will be conducted to compare the two sources at national and Council Area level.

4.4.26 Method of transport to work or study

The method of transport used to travel to a person's main job or course of study (including school) will be validated using comparator data drawn from Transport Scotland's Transport and Travel in Scotland statistics. Census categories will be aggregated to match the Transport and Travel in Scotland categories of walking, car or van, bicycle, bus, rail and other. For children travelling to school, the bus category is split by school bus, service bus. The data will be compared by sex and 10-year age bands at national level.

If restrictions relating to COVID-19 are still in place around March 2022, then the data from this question may be particularly affected due to larger than usual numbers of people working from home, this will be taken into account when comparing with the Transport Scotland data.

5. Conclusion

We currently anticipate being able to validate population and topic estimates for the vast majority of census questions, including breakdowns by Council Area and by age and sex, where appropriate. Work is ongoing to secure access to the most up to date comparator data sources and to identify additional data sources for topics that currently lack robust comparators. Comparator data sources are a mix of administrative and survey data and expert opinions. We are developing methodologies to determine how to assess discrepancies between census and comparator data, adjusted for the timeliness and completeness of the comparator data.

6.   Annex

6.1   Glossary

Term	Definition
Administrative data	Administrative data refers to information collected primarily for administrative (not statistical or research) purposes. This type of data is collected by government departments and other organisations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.
Council Area (Local Authority)	There are 32 councils (local authorities) in Scotland, the administrative units of local government.
Communal establishment (CE)	A communal establishment is typically managed residential accommodation where there is full-time or part-time supervision of the accommodation.
Comparator data	2011 definition A range of data was used to help quality assure 2011 Census estimates. These data included administrative sources (such as the number of people registered with a doctor) and other official NRS products (such as the mid-year population estimates).
Data Protection Impact Assessment (DPIA)	A Data Protection Impact Assessment (DPIA) is a process to help you identify and minimise the data protection risks of a project.

Data Zone	Data zones are the key geography for the dissemination of small area statistics in Scotland and are widely used across the public and private sector (Full definition [here](#))
Main language	ONS 2011 Definition This is a person's first or preferred language.
National Statistics Socio-economic Classification (NS-SeC)	The National Statistics Socio-economic Classification (NS-SeC) provides an indication of socio-economic position based on occupation.
NRS	National Records of Scotland
NISRA	Northern Ireland's census is run by the Northern Ireland Statistics and Research Agency (NISRA).
ONS	The Office for National Statistics (ONS) is the UK's largest independent producer of official statistics and the recognised national statistical institute of the UK. The census in England and Wales is run by the ONS
Pay as you Earn	Pay As You Earn (PAYE) is the system employers and pension providers use to take Income Tax and National Insurance contributions before they pay wages or pensions to employees and pensioners. This publication relates to employees only and not pensioners. It was introduced in 1944 and is now the way most employees pay Income Tax in the UK.
Protected characteristics	It is against the law to discriminate against someone because of age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex and sexual orientation. These are called protected characteristics.

Quality assurance	Quality Assurance (QA) is about identifying, anticipating and avoiding the problems that can arise from our data inputs or the methods and processes we use to calculate statistics.
Quality Assurance of Administrative Data (QAADs)	A process that will be produced to help statistical assessors review the quality assurance arrangements of the address register used to produce census statistics.
Scottish Index of Multiple Deprivation (SIMD)	The Scottish Index of Multiple Deprivation is a relative measure of deprivation across 6,976 small areas (called data zones). If an area is identified as 'deprived', this can relate to people having a low income but it can also mean fewer resources or opportunities. SIMD looks at the extent to which an area is deprived across seven domains: income, employment, education, health, access to services, crime and housing.
Validation of Population Estimates (VoPE)	The VoPE process compares census data with existing data sources in order to verify that the census data are expected given the comparator data. This process focuses on geographic areas, population groups and topic areas where there are inconsistencies or need for further analysis.