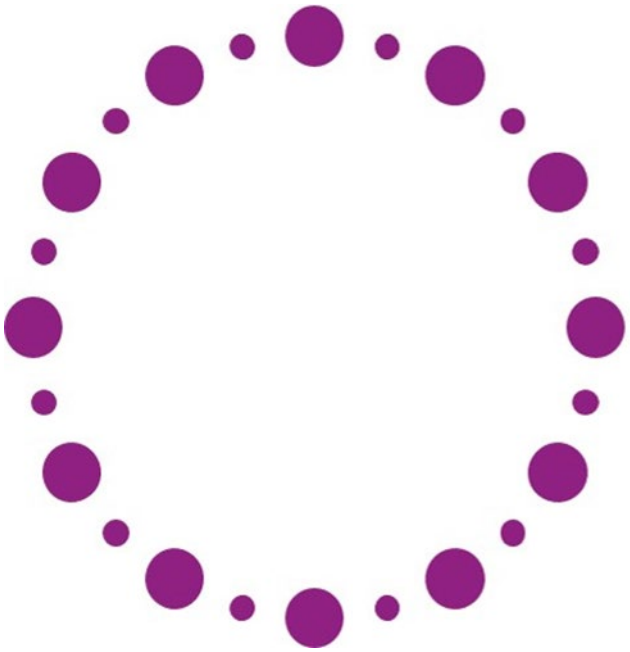
census 2021

Census 2021

Coverage information paper

25 May 2023



NISRA

nisra.gov.uk/Census2021

Table of contents

1. Introduction

A census usually takes place once every ten years and is the largest and most complex statistical exercise undertaken in Northern Ireland. Census statistics are a vital source of information and are widely used by government, public bodies, academia, commercial businesses and others to develop policies, allocate resources and help deliver services. The last census in Northern Ireland was taken on 21 March 2021.

A good census obtains complete and accurate responses directly from the vast majority of the population. However it is recognised that, while every effort is made to maximise response, no census will enumerate every single household and individual directly (this is referred to generally as undercoverage). To correct this, robust methodologies are used to enumerate as many people as possible and then estimate and adjust for any remaining undercoverage.

Invariably the census is also affected by a small amount of overcoverage – this is when responses are obtained from people who should not have been counted. As an example, unmanaged duplication of response can be viewed as overcoverage – where multiple responses are submitted for the same person in an unplanned way. To correct this, robust methodologies are implemented to avoid or remove overcoverage.

This paper will describe these approaches, and the quality assurance undertaken to validate the robustness of the data.

2. Maximising coverage in the collection phase

It is well established, and observed in previous censuses, that undercoverage is not distributed uniformly by geographic area and population demographic. Consequently, the design of the 2021 census sought to address this by focusing significant effort on specific areas and sub-populations in the collection phase to endeavour to make sure that no-one was missed.

This work included procedures to assist with collection for special populations such as students, people with no settled place of residence and persons sleeping rough (more information on operational measures that were introduced for Census 2021 to

help maximise the response rate can be found in the Census 2021 operational report).

As an example, a comprehensive process was implemented for capturing the characteristics of those living in approximately 1,500 communal establishments such as nursing homes, hotels, hospitals, boarding schools or prisons. This included gathering information from establishment managers and universities in the form of listings of who were registered as staying in relevant establishments at the time of the census. This information was used to ensure that the communal establishments' resident record listings were as complete as possible. It should be noted that this process was over and above the return of individual forms from residents in these establishments and thus this twin-track approach ensured as full an enumeration as possible.

Numerically, around 800 records were added to general communal establishments from this process, with a further 600 records added to student halls of residence (further information on the student enumeration can be found in the Census 2021 student information paper).

3. Identifying and removing overcoverage

There are two types of overcoverage in a census. One relates to duplication and can happen for various reasons, such as:

- returns made for children of separated parents at two different locations;
- returns made for people with more than one home;
- returns made where a household has moved home in the collection window;
- returns made on individual forms and continuation forms where the individuals were already included on the main questionnaire; and
- duplicates arising from responses made using both electronic questionnaire and paper questionnaire channels.

The other type is due to spurious or limited responses in the data. For example, people filling in responses for pets, or family members that no longer live with them, or making returns for which insufficient information was provided.

The duplication listed above stems from the standard census design and has been inherent in recent censuses due to the nature of the collection operation. However, in 2021 the instruction to students to include themselves at both their home and their term-time addresses, introduced the potential for more overcoverage than has previously been experienced. However, this extra overcoverage was specifically planned to maximise the coverage for students. As such, specific data processing steps were introduced to compensate for it (see step three below and Census 2021 student information paper for more details).

To mitigate against all overcoverage, NISRA enhanced its already well-established procedures for identifying and removing overcoverage, by adding processes to examine the **full** census response dataset against itself. This work also included linking the census response dataset with administrative data to assist the processing. This process was referred to as **Data Consolidation** and is explained in the steps below.

Step one in this process was to consolidate all the different sources of census data and identify and remove any inconsistencies. All individual and continuation returns were associated with their household or communal establishment return.

Step two is the removal of false persons process. This process caters for spurious or limited responses in the census by removing any returns that were clearly not real people or did not contain sufficient information to be treated as a valid response. In order to decide if the response was valid and therefore not a “false” person, at least 2 of the following 5 variables had to be present (slightly different rules were applied to data collected from the paper questionnaire for points a and b):

a. Name on individual questions
b. Name on household members table (paper questionnaire only)
c. Date of birth
d. Sex
e. Marital status

For 2021 this process was enhanced with the incorporation of administrative data. Here, if the failing raw census data could be linked to an administrative record, the raw census record was enhanced with this data and therefore these raw census

records had a greater chance of being retained, rather than disregarded. Enhanced records were only retained where they didn't create a duplicate of an existing response. **In total the false persons process removed around 1.7% of all records**[1] (1.0% could not be enhanced or were spurious; 0.7% were enhanced but would have introduced duplicates)**.**

Step three is reconciling multiple responses (RMR). This process is the tried and tested census approach taken to identifying and removing duplication in the census dataset. RMR was enhanced for 2021 to include extra modules that match the **entire** Northern Ireland dataset to itself, allowing duplicates to be removed across the whole database. In the 2011 Census, only the removal of duplicate records within the same household or communal establishment was possible. This is a significant improvement in overall quality for 2021. **This process removed around 2.6% overcoverage** (of which 0.4% was attributable to student overcoverage).

Step four, student moves - this process identified all students in the dataset and utilised administrative data to assess both the coverage and location of the students. Where students had been enumerated at their home address and information was available on their term-time address, they were moved to their correct term-time address.

4. Identifying and adjusting for undercoverage

As a result of the comprehensive work carried out during collection, NISRA was confident that coverage had been maximised for Communal Establishments and that nobody was missed. As such, communal establishment residents were not subject to an adjustment to account for undercoverage as a result of statistical estimation. Therefore, the remainder of this section covers the estimation and adjustment processes relating to households only.

Census 2021 had a household response rate of 97%. Therefore there was 3% of people and households from which a census response was not received. This

[1]While extremely difficult to determine, a proportion of these records may have been from real people not found elsewhere on the census database. It is expected that the undercoverage methodologies recovered any shortfall created from this process.

estimate and the processes to correct for undercoverage were driven from two processes, run one after the other:

1. The Census 2021 Census Under Enumeration (CUE) project; and
2. Coverage estimation and adjustment using statistical methods based on a Census Coverage Survey (CCS).

The Census 2021 CUE project utilised information from administrative data records and looked at the household addresses from the field operation where Census Office believed a response should have been received. Within this group administrative data was used to “administratively enumerate” some of the non-responding households – in total this added 26,900 residents in 12,700 households to the dataset (this equates to around 1.4%).

Separately, an independent Census Coverage Survey (CCS) was taken immediately after the main census data collection. The CCS sample covered a host of postcodes from all across Northern Ireland, with around 16,000 addresses found and, where possible, re-enumerated – the CCS had a response rate of 88%. The CCS results were then matched to the 2021 Census and a capture-recapture statistical methodology was used to assess the numbers of households/individuals missed by the census enumeration.

The statistical analysis showed that a further CCS coverage adjustment of around 31,900 people and 9,200 households was needed to create a complete estimate of the population and households in Northern Ireland. This equates to an undercoverage of 1.7%.

More detail on both methods is given below.

4.1. Census Under Enumeration (CUE) project

For 2021, NISRA further developed the CUE project implemented in the 2011 Census. Through the field operation and access to high quality administrative data records, NISRA had evidence pointing to the presence of usual residents at a number of addresses with no census response.

As with the 2011 Census, this was only undertaken for completely non-responding addresses where NISRA considered that a response should have been received. The approach taken was cautious and is described in greater detail in the paper entitled 'Using an Administrative Primary Care Health Activity Indicator to Address Under-enumeration in the 2011 Census in Northern Ireland', published after the 2011 Census. It should be noted that this method is now internationally recognised and has been used by New Zealand in their 2018 Census.

The Census 2011 CUE project utilised records from the Health Card Register, adding a total of 67,000 persons in 30,000 households. NISRA has built on this work to use multiple administrative data sets for the Census 2021 CUE project. A linkage framework was used to determine which records belong to the same person across all datasets.

Following Census 2021 fieldwork and data consolidation, those household addresses where a response was expected but none had been received were identified. Any addresses where field staff had identified that the address was vacant, a second residence or a holiday home were not included, leaving only addresses where an enumerator had found evidence of occupancy. This resulted in a list of nearly 25,000 household addresses eligible to be assessed against administrative data.

A Census 2021 CUE dataset was created by extracting individuals from administrative data, where the individuals were recorded at these non-responding addresses. Records were excluded if they met any of the criteria below:

1. Linked to a census resident (i.e. they had completed a census return at a different household or communal establishment address);
2. Resident in a child-only household (the maximum age within the address was under 16);
3. Born after Census Day;
4. Aged 90 or more and not active on the Health Card Register; or
5. More than 10 records were located at the address on administrative data.

The final Census 2021 CUE dataset contained 26,900 residents in 12,700 households. These records were added to the census database with a limited

number of known characteristics taken from administrative data, including: Names, date of birth, age, sex, address, country of birth, year of arrival, one year ago address, marital status, in full time education, place of study and ethnicity, where available.

4.2. Census Coverage Survey (CCS)

Separately, NISRA ran an independent CCS immediately after the main census data collection. The CCS is an independent interviewer-led survey, and for the 2021 CCS, interviewers were in the field from 12 May 2021 to 29 June 2021.

Experienced interviewers from NISRA's Central Survey Unit rather than census field staff were used, thus ensuring independence from the census enumeration, an important part of the methodology. In addition, interviewers created their own address listing for sampled postcodes based on investigation of their allocated postcode areas, rather than relying on the Census Address Register.

The CCS results were matched to the 2021 Census and a capture-recapture statistical methodology was used to assess the numbers of households/individuals missed by the census enumeration. This capture-recapture approach is a well-established method of estimating the true extent of a population, and this is the third census in which such a methodology has been successfully applied in Northern Ireland.

4.2.1. Statistical methodology

The CCS sample included approximately 16,000 households (or around 2% of residential addresses in Northern Ireland). Postcodes rather than addresses were selected for inclusion in the CCS sample using a random sample of Small Areas, stratified by Northern Ireland's 11 Local Government Districts, together with a Hard to Count (HtC) Index.

The HtC Index is a five way measure of the anticipated difficulty of enumeration of each postcode. A 2021 HtC Index was developed for the 2021 census, and classified each of Northern Ireland's 4,537 Small Areas into one of five strata, with '1' being expected to be easiest to enumerate and '5' being the most difficult. The model used

response patterns from the 2011 census to help predict which characteristics of an area are associated with response, for example, deprivation, and accommodation type.

The CCS questionnaire used a subset of census questions to collect basic demographic characteristics. It included only those questions which were required for data linkage (name, age, sex) or for estimation processing (economic activity, religion, household tenure). The CCS asked respondents to focus on their household situation at the time they completed their census return.

The response rate for the 2021 CCS was 88%, which was higher than the 85% achieved in 2011, and provided a sound basis on which to create statistical estimates.

The data collected by the CCS were processed in a way that closely mirrored that from the census to reduce potential bias from processing methods (consolidating responses, removal of false persons and reconciling of multiple responses).

4.2.2. Census to CCS data linkage

Once both the census and CCS datasets had been processed to remove duplicate responses, persons and households from the CCS were matched to census responses using both automated matching processes developed by NISRA, and clerical matching for records that were not auto-matched. In addition, a percentage of automated matches were subject to clerical review. These were found to have a high degree of accuracy, with less than 0.1% false positives found by clerical checking.

4.2.3. Coverage estimation

The linked census-CCS data allowed residents and households to be classified into three categories, namely:

- captured by both census and CCS;
- captured only by the census; and
- captured only by the CCS.

This formed the basis of the estimation calculations that allowed the calculation of the likelihood of a person of a given age and sex, or a household with a particular tenure, to be missed from the census. This was used to calculate the true size of the population by age and sex, and the population distribution by Local Government District.

The 2021 coverage estimation process followed methodology developed for the 2001 and 2011 Censuses with only minor differences. Improvements in the technical capacity of digital processing allowed Northern Ireland to be treated as a single area for estimation in 2021, rather than being split into three areas as was the case in the two previous censuses. This allowed for more statistically robust estimates to be generated.

The statistical analysis showed that a coverage adjustment of around 31,000 people and 9,000 households was needed to create a complete estimate of the population and households in Northern Ireland.

4.2.4. Undercount in young children adjustment

During quality assurance of coverage estimates, the number of very young children were compared to administrative data sources as it is acknowledged that people can “forget” to include very young children on their census questionnaires.

The main known reasons for this include a new baby born on or just prior to Census Day and the return has already been made, or that the respondent believes that young babies do not need to be included on the questionnaire – this can especially be the case if the birth has not as yet been registered or no name has been decided for the baby.

To assess the extent of the issue, NISRA compared the census database with birth registrations in Northern Ireland in recent years and also linked responding mothers in the census to the parental information provided within those birth registrations. This identified a number of children aged 0 to 2 years that could be inserted into the census database. To be introduced into the census data, a birth record had to pass all the following criteria:

- baby/child with a date of birth between 22 March 2019 and 21 March 2021;
- baby/child not recorded as deceased prior to 21 March 2021;
- mother from birth registration record found in the completed census return; and
- baby not found on the mother's completed census return.

As a final step, babies were removed from the eligible records if:

- the associated mother was resident in the census in a communal establishment; or
- the babies' details on the birth registration form had a "forename" or "year of birth" which was the same as any other person resident on the form in the mother's return.

As a result of this analysis just under 1,000 young children (aged 0-2) were identified to be added to households in the census database as a "very young children" adjustment.

5. Adjusting the dataset to reflect final estimates

Once estimates of the population had been calculated, subjected to quality assurance, and a "very young children" adjustment agreed, the census dataset could be adjusted by adding in residents and households. The 2021 adjustment process followed the methodology applied in 2011, modelling the probability of a resident with a given set of characteristics (e.g. age and sex) being missed from the census, either from an enumerated household, or missed from a household that did not respond (i.e. a missing household).

This in turn allowed the selection of donors with the required characteristics from the census database. Donor residents with suitable characteristics to impute into enumerated households, and whole donor households with suitable characteristics to impute into empty addresses. Limited variables are carried forward from the donor.

The addresses that would be eligible for imputation of whole donor households were identified using information gathered by field staff, coupled with address-based activity measures developed by NISRA using administrative data.

Imputation of residents and households was examined in detail at lower geographic levels to look for spikes in imputation and compare to other available data sources prior to acceptance of the adjusted population.

In total 31,900 residents and 9,200 households were added to the NI population by statistical methods to reflect the final estimates (including the very young children adjustment). These records were added to the census database with limited variables.

6. Quality assurance

Census 2021 population estimates were compared against a variety of statistical and administrative datasets to benchmark across all age groups, and further information is provided on the NISRA website in the Census 2021 quality assurance report (PDF 275 KB).

The census outputs were also independently reviewed by a panel of four external experts, with experience of census and population estimates, and who reported to the Registrar General. This team asked important and probing questions regarding the estimates and helped to shape the quality assurance report.

7. Response rate

On Census Day 2021 there were an estimated 1,903,200 usual residents living in Northern Ireland. Of this 1,844,500 people filled out a census return. The difference (58,700 people) was estimated using information from the field operation, administrative data, statistical modelling and the Census Coverage Survey. Thus it is estimated that 58,700 people were not included on a census return. This gives a final full census person response rate of 97% (1,844,500/1,903,200).

In 2011 the person response rate was 92% and in 2001 the person response rate was 95%. Thus the overall response rate to Census 2021 was the highest in recent times. As census users expect statistics to reflect the full population, Census Offices

across the UK include modelled data in the 2021 Census to account for people who did not respond[2]. Thus locally 58,700 census non-response records (3%) have been modelled and the final results reflect the full estimated usual resident population.

The response rate can also be viewed in terms of households; on Census Day there were estimated to be 768,800 households in Northern Ireland, of which 746,900 provided a response with an estimated 21,900 non-responding households. This gives a household response rate of 97% (746,900/768,800). In 2011 the household response rate was 94% and in 2001 the household response rate was 95%.

7.1. Confidence intervals on estimates

Population level confidence intervals were generated using a "bootstrapping" technique that created 2,000 alternative CCS samples by re-sampling CCS data with replacement. This process gave a 95% confidence interval of 0.25% overall, that is +/- 4,700 residents of the 1,903,200 total population.

[2] Modelled data for 'census not returned' was implemented in the 2001 and 2011 Censuses in all parts of the UK. For Northern Ireland 158,000 modelled records were added to the 2011 Census and 82,000 modelled records were added to the 2001 Census. In total 58,700 modelled records were added to the 2021 Census.