

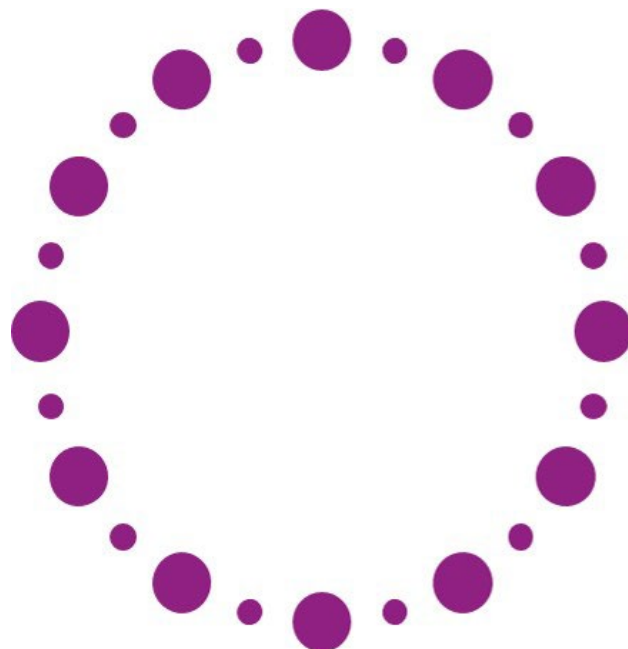
census
2021



Census 2021

Population and household estimates for Northern Ireland: Quality assurance report

24 May 2022



nisra.gov.uk/Census2021

Table of contents

| | |
|--|----|
| 1. Introduction..... | 3 |
| 2. Phases of quality assurance..... | 4 |
| 2.1. Quality assurance – planning..... | 4 |
| 2.1.1. Address register..... | 4 |
| 2.1.2. Electronic questionnaire | 6 |
| 2.2. Quality assurance – collection | 7 |
| 2.3. Quality assurance – data transfer | 8 |
| 2.4. Quality assurance – special enumeration | 10 |
| 2.5. Quality assurance – accuracy..... | 11 |
| 2.6. Quality assurance – process..... | 12 |
| 2.6.1. Data consolidation (consolidating responses, removal of false persons and reconciling of multiple responses)..... | 12 |
| 2.6.2. Data transformation | 13 |
| 2.6.3. Coverage adjustment using administrative data (Census Under Enumeration (CUE)) including an adjustment for very young children | 14 |
| 2.6.4. Coverage estimation and adjustment using statistical methods | 15 |
| 2.7. Quality assurance – outputs | 16 |
| 3. Quality assuring the census estimates | 17 |
| 3.1. Comparison against administrative data sources | 17 |
| Figure 1: All usual residents on census day..... | 19 |
| 4. Conclusion..... | 20 |

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.

It is essential to ensure the quality of census data given the key role it plays. In January 2021, the Northern Ireland Statistics and Research Agency (NISRA) published the [Census 2021 quality assurance strategy](#) which gives an outline of the quality assurance work that will be undertaken in order to ensure that the outputs released from Census 2021 are robust, reliable and meet user needs.

As outlined in the strategy, quality assurance of Census 2021 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;
- deliver census results as soon as possible;
- minimise the risk of errors in the census estimates;
- enhance the overall credibility of the statistics; and
- leave a legacy of methods, tools and skills for the quality assurance of future population statistics.

This report details the phases taken as part of the quality assurance process to ensure the Census 2021 statistics are, based on sound data and methods, as well as being trustworthy and of a high quality.

2. Phases of quality assurance

2.1. Quality assurance – planning

Years of planning go into a census to design the content and agree questions to be asked. As part of this planning a number of activities took place:

- a review of the successes and lessons learnt from the 2011 Census and [2011 Census quality survey](#);
- the [Census 2021 topic consultation](#) carried out between September and December 2015;
- a programme of testing and research to better understand the public's perception of the census, question wording, questionnaire design and response channels;
- the [census test in 2017](#), which was a large-scale, voluntary census test conducted in autumn 2017;
- topic expert meetings which took place between 2016 and 2019; and
- a full end-to-end [census rehearsal](#), which took place in autumn 2019.

This work helped to get the questionnaire design right and test the processes ahead of Census 2021.

2.1.1. Address register

Central to the operational design of Census 2021 was the development of a comprehensive address register of all occupied or occupiable addresses in Northern Ireland (domestic and communal establishments). The development of the address register for the 2021 Census built on lessons learned from the 2011 Census and recent tests/rehearsals undertaken in the interim period.

To help determine which addresses were to be included in the address register, a scoring methodology was developed by analysing address performance/inclusion on a range of datasets; this included:

- POINTER¹ characteristics;

¹ POINTER is the address database for Northern Ireland and the standard address for every property. Land and Property Services (LPS) maintains the database with help from local councils and Royal Mail. More information on POINTER is available on [NIDirect website](#) and the POINTER technical specification is also available from [Spatial NI website](#).

- domestic data from the valuation list²;
- Royal Mail postal address file;
- housing associations data;
- other administrative data such as vehicle test data;
- data from Northern Ireland Electricity on properties under supply; and
- data from local household surveys.

A program of work was then developed to enhance the selection process, which included:

- working closely with the Land and Property Services (LPS);
- reviewing a sample of addresses in each address score (for example, postal checks);
- desk based address checking (using web searches and aerial photography);
- linking more address databases to POINTER;
- on-the-ground activities (to check progress of new builds from information provided by local planning offices);
- engagement with all the Local Government Districts (LGD) in Northern Ireland.

The work resulted in a pre-census day address register that included 843,000 domestic addresses and 1,500 communal establishment addresses. These addresses were the primary basis for Census 2021. In creating this address register, Census Office aimed to maximise coverage by including lower scoring addresses, which would be subsequently followed up by field staff. This address register had multiple functions during the census operations:

- it facilitated the posting out of letters/questionnaires to households and communal establishments;
- it facilitated the process for following up non-responding households by identifying addresses for which no return had been received;
- it provided address 'lookups', for users of the online questionnaire.

² In line with The Rates (Northern Ireland) Order 1977, housing stock is defined as a count of properties which are valued as domestic or mixed for the purposes of rating. This refers to properties in the valuation list which are used for the purposes of a private dwelling; and excludes caravans, domestic garages, domestic stores and car parking spaces. The Northern Ireland valuation list housing stock statistics are available from [Annual housing stock statistics webpage](#).

The address register was maintained throughout the census period and when new addresses were identified they were added to the register, with invalid addresses being removed. Further details on addresses will be published in a technical paper in due course.

2.1.2. Electronic questionnaire

One of the keys to a successful census is to ensure that the completion process is straightforward and easily accessible to all. For Census 2021 this was greatly facilitated by the use of a bespoke electronic questionnaire. This system was tested in the census rehearsal in 2019 and further developed over the period between the rehearsal and census day. The electronic questionnaire was developed with automated routing to only show respondents the questions they needed to answer. Drop down lists and verifications were built in as part of the completion process, to help respondents fill out the questionnaire accurately. In addition, a range of accessibility products and guidance were produced to help respondents understand the questions and then complete their questionnaires. Alongside this, frequently asked questions, website help, videos and translation booklets were provided to maximise the quantity and quality of responses from respondents.

In total, just over 80% of returns for households were made online³. Online responses tend to have a higher data quality due to automated routing and lower levels of 'missing items', compared to returns made on paper questionnaires. Analysis of the first five questions on the 2021 form (names, date of birth, sex and marital status) showed the electronic questionnaire had a 0.5% level of 'missing items' while the paper questionnaire had a 3.6% level of 'missing items'. Given the high level of online returns, the overall quality of the collected data in the 2021 Census was significantly higher than in the 2011 Census.

³ In total, for persons 85% of returns were made online. The increase from household online percentage (80%) to person online percentage (85%) is due to households that contain more people tending to respond online. This position was to be expected, as households that contain less people tend to be older and thus more likely to respond on paper.

2.2. Quality assurance – collection

A major focus of Census Office in the run-up to Census 2021 was developing the systems and services required to collect data securely and efficiently from the general public.

The main areas and processes included:

- rigorous refinement and testing of the electronic questionnaire (eQ);
- the management of the printing and posting out of :
 - door-drop postcards (in total around 2.5 million postcards were sent);
 - initial contact letters (in total around 675,000 initial contact letters were sent);
 - initial contact paper questionnaires (in total 168,000 initial contact paper questionnaires were sent);
 - reminder letters (in total around 250,000 reminder letters were sent over three waves);
 - reminder paper questionnaires (in total around 80,000 reminder paper questionnaires were sent);
 - questionnaires requested by the public during the operation (in total the public requested 50,000 paper questionnaires);
- the delivery of paper questionnaires to households and communal establishments;
- targeted follow up of non-responding households by the field staff.

In conjunction with this, there was a contact centre and online help, which assisted the public with making their census returns. In total, over 160,000 phone calls were received by the contact centre.

A wave of contact timetable and materials were developed to ensure households knew when the census was happening, how to complete their questionnaires and to enable additional reminder letters could be sent to non-responding households. Full details can be found in the [Census 2021 operational report](#).

A central system for the management of data collection was developed which allowed live tracking of the status of each census return throughout the operation; this included interactions with, or requests from, householders. This system acted as the basis for management information and informed the decision making process

around changes to the interaction with householders during the collection period. In particular, the daily management information helped inform key decisions in relation to reminder letters, as well as developing new approaches to drive up responses such as targeted letters to holiday homes and student areas. Data on this central system was also updated with information on vacant or uninhabitable properties provided by the Northern Ireland Housing Executive (NIHE) and housing associations. Various administrative data sources were also used to look for signs of activity, to help decide if a response was likely from particular addresses.

A final and vital element in the collection operation was a targeted 'push' on the doorstep of non-responding households/addresses. In total over 375,000 visits were carried out by field staff during the follow-up period. These visits were targeted using data from the central system mentioned above to drive a fieldwork management tool used by all field staff.

Full details of visits and return rates can be found in the [Census 2021 operational report](#).

All the collection approaches resulted in an overall household and person response rate of 97% and a communal establishment response rate of over 99%. This is the highest census response rate since 1991, the corresponding figures were 94% household and 92% person in 2011. Further information on how response rates are calculated can be found in the [Census 2021 outputs definitions](#).

2.3. Quality assurance – data transfer

Working with the Office for National Statistics (ONS) and NISRA, the contractor prepared and dispatched paper questionnaire packs, and then securely managed, captured, digitised and coded the responses for the 2021 Censuses in England, Wales and Northern Ireland. Once processed, the data relating to scanned paper questionnaires were transferred to ONS.

Following receipt of this data, ONS carried out a large reconciliation exercise to ensure that data for all paper questionnaires were received. ONS merged this data with data from the online system prior to it being delivered to NISRA. There were established processes for checking the data, both prior to transfer from ONS and following receipt of data by NISRA. This ensured that the data satisfied the specified requirements and included checks such as:

- reconciling expected numbers of records against data files delivered (by response type and response mode);
- variable range checks;
- variable format checks;
- analysis of variable distributions to identify unusual patterns;
- geographical distribution of responses.

A range of testing was carried out on the paper questionnaire and online data supply pipelines including test cases being input to cover all response options for each question both paper and online. These test cases were then followed the whole way through the various pipelines to validate all the pre-processing steps. This helped to check the accuracy of capture and coding of the information prior to its delivery to NISRA.

Once the data was received and verified, a number of further processes took place, in order to refine the data prior to onwards processing by NISRA. These included:

1. removing test and spurious returns – all the test records that had been entered were removed along with a small number of obvious spurious persons/households. In addition, a separate exercise was run to resolve those returns that did not provide enough information on name, date of birth and sex;
2. removing those born after census day – some responses included persons born after census day, these persons were disregarded;
3. removing those deceased prior to census day – administrative death records were used to ensure that the population recorded was as accurate as possible for 21 March 2021;
4. a small number of paper questionnaires were returned without being linked to an address, so additional work was undertaken to link these returns to the relevant address;
5. reconciling placeholder returns – ensuring a record was created for every non-responding valid address so that the dataset included all domestic addresses in Northern Ireland. Placeholder forms were removed for addresses from which census responses were subsequently received;

6. blank questionnaires – a very small number of paper questionnaires were recorded as blank. Each of these questionnaires were examined and, where possible, information on residents or visitors was retrieved;
7. inclusion of late returns – some questionnaires were received after the processing system had closed. In some cases, the householder telephoned the local census office and provided information over the phone. In total 400 questionnaires were added during this process;
8. reconciling communal establishment manager listings with individual census returns – managers at communal establishments were required to provide the name, date of birth and sex of all residents in their establishments. Residents were also asked to fill out a separate individual questionnaire. This meant two sources of data in many cases for the same person; this information was reconciled to ensure that there was a complete list of residents for each communal establishment.

2.4. Quality assurance – special enumeration

To ensure maximum participation in Census 2021, a dedicated team was established to oversee the process for capturing the characteristics of those living in approximately 1,500 communal establishments such as hotels, hospitals, boarding schools or prisons. This team worked with the managers of the communal establishments to ensure a response was made for each one. Additional procedures were also adopted for some special populations. This included people with no settled place of residence and persons sleeping rough.

A number of measures were introduced for Census 2021 to help drive up the response rate of special population groups, these included:

- designing a new manager questionnaire for communal establishments which captured the number of residents along with their name, date of birth and sex – this was accessible online and on paper and was included in the [Census 2021 legislation](#);
- additional engagement with local universities enabled emails, with instructions on how to complete the census, to be sent to all students. The universities provided data on students enrolled and their term-time addresses;
- the NIHE, community liaison representatives and voluntary organisations were contacted and helped to provide information on Irish Traveller sites;

- the NIHE provided information on the number of homeless people housed in local hotels around census day;
- the NIHE and housing associations provided information on social housing which was vacant or uninhabitable;
- students were asked to fill in their census questionnaires at their term-time address, and their parents/guardians were also asked to complete details on them at their family home. This facilitated a more accurate count of students, given the circumstances of the pandemic.

Engagement took place with the responsible public bodies (NIHE, Department for Communities and Department of Health) to ascertain the number of persons sleeping rough around census day.

Further details on how students were enumerated will be published in a technical paper in due course.

2.5. Quality assurance – accuracy

A number of approaches were used to ensure the captured and coded data from Census 2021 questionnaires was of the highest quality. These included:

- internal quality assurance was carried out by ONS and the supplier to ensure that the captured and coded data met the agreed specification and service level agreements targets;
- to improve the accuracy of statistical coding, Artificial Intelligence (AI) coding techniques were successfully introduced into the overall coding process for the first time;
- independent validation was done by the local NISRA census team to ensure that the data met agreed specification and service level agreements targets;
- there was a separate NISRA internal review and, where possible, coding of all ‘uncodeable’ responses;
- an additional manual review of all records was undertaken of specific age groups. As an example, all centenarians (100+ year olds) were individually checked against administrative data and records held by NISRA. This check was extended to all people aged 90 to 99 years, to ensure that the estimate of the very elderly was as accurate as possible.

2.6. Quality assurance – process

Invariably the census is affected by a small amount of over-coverage (for example, duplication) and a small amount of under-coverage (for example, missed people and households). To correct this, robust methodologies were implemented to assess and remove over-count, and estimate and adjust for undercount. Given the complexity of these processing steps, it is necessary to have controls in place to assure the processes are doing what they were designed to do, and to measure and manage quality throughout.

Data processing involves many steps, the four main stages applied to the data for the first release are:

1. data consolidation (consolidating responses, removal of false persons and reconciling of multiple responses);
2. data transformation;
3. coverage adjustment using administrative data (Census Under Enumeration (CUE)) including an adjustment for very young children;
4. coverage estimation and adjustment using statistical methods.

2.6.1. Data consolidation (consolidating responses, removal of false persons and reconciling of multiple responses)

- Step one in this process was to consolidate all the different sources of data and identify and remove any inconsistencies. All individual and continuation returns were associated with their household or communal establishment return.
- Step two, the removal of false person process, removed any census returns that 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:
 - a. Name on individual questions*
 - b. Name on household members table (paper questionnaire only)*
 - c. Date of birth
 - d. Sex
 - e. Marital status

*Slightly different rules were applied to data collected from the paper questionnaire.

- Step three (reconciling multiple responses), identified and removed duplicate census responses. This step worked across the whole database and is only possible with the increased computational power available today. This process of removing duplicates was not undertaken at all in the 2001 Census and only carried out for duplicate records within the same household or communal establishment in 2011. This is a significant improvement in overall quality for 2021.

2.6.2. Data transformation

There are three main stages to data transformation, all designed to correct for inconsistencies and 'missing items' within household and/or individual responses. Firstly, inconsistencies can arise from people not following the routing on the questionnaire, (this was only possible on the paper form). A process called 'filter rules' was applied to the data to correct any inconsistency or ambiguity relating to this issue. For example:

- the language questions wouldn't have been displayed for any persons aged 2 and under. Therefore, a filter standardises online and paper questionnaire responses by setting all responses to 'no code required' for any person recorded as aged 2 years and under;
- persons aged under 5 are not shown the carer question (question 22). Therefore, a filter standardises online and paper questionnaire responses by setting all responses to those questions to 'no code required' for persons recorded as aged under 5 years;
- for those who are not currently working or studying, or who work or study mainly from home, the question "How do you usually travel to your main place of work or study (including school)?" was set to 'no code required'.

Secondly, some rules were applied to the data to correct "invalid" scenarios. These rules were applied to address and issues found – as an example one rule corrects for a parent being younger than a child.

Thirdly, any missing items in responses were statistically imputed using a donor imputation methodology designed and developed by the Canadian Census Bureau

called CANCEIS⁴. This approach is an international standard and used in censuses in Great Britain and in other countries.

2.6.3. Coverage adjustment using administrative data (Census Under Enumeration (CUE)) including an adjustment for very young children

As was done in the 2011 Census, the 2021 Census has used high quality administrative data records to add a small number of people into the census dataset, to account for domestic addresses where the systems indicate no response had been made. This approach has been used since the 2011 Census in other countries (Canada 2016, New Zealand 2018) and is now an international standard in terms of helping to address any undercount found.

As with the 2011 Census, this was only done for completely non-responding addresses where Census Office 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](#).

The Census 2021 Census Under Enumeration (CUE) process added approximately 27,000 residents to just under 13,000 households.

It is acknowledged that people can “forget” to include very young children on their completed census questionnaires. This is an issue in censuses around the world. 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 form. In order to address this issue a number of young children (0 to 2 years old) were added to the final census database. The returns were only added where there was clear evidence that the mother of the baby (from the birth registration) was included in the census at an address in Northern Ireland, and where there were no other children of a similar age included at the same address on the mother’s completed census questionnaire. During this process just under 1,000 young children (aged 0-2) were added to the final census database.

⁴ More information on CANCEIS (CANadian Census Edit and Imputation System) is available in the online journal article - [Efficient methodology within the Canadian census edit and imputation system \(CANCEIS\)](#)

2.6.4. Coverage estimation and adjustment using statistical methods

While completion of a census questionnaire is a legal requirement, the reality is that a very small number of individuals will not comply fully with the census. As has been done since 2001, statistical processes (sometimes called capture-recapture techniques) are used to assess and address the coverage issues that arise as a result of this, in order to make adjustments to the census counts to ensure that the outputs provide the best estimate of the true population.

The Census Coverage Survey (CCS) is the statistical methodology used to assess under-coverage. It involves a re-capture of selected census information from a representative sample of households across Northern Ireland. Once captured, the CCS information is then matched to the census results to estimate which household and/or individuals have been missed. Estimates are calculated from this process using robust statistical methodologies that extrapolate the results across the entire population. This is in addition to the CUE process which added information from high quality administrative data records⁵. Once estimated, the records required to adjust for under-enumeration are then added to the census data with minimal socio-demographic information. These records are then subsequently fully populated using the CANCEIS process.

The CCS is an independent interviewer-led survey carried out immediately after the census fieldwork is completed. For the 2021 CCS, interviewers were in the field from 12 May 2021 to 29 June 2021.

The CCS interviewers were NISRA interviewers who normally work on the official social surveys. This meant they were already experienced interviewers, so training could focus on the purpose of the CCS, the schedule and equipment, the high response rate needed, and the key questions that needed accurate answers.

The Northern Ireland CCS sample included approximately 16,000 households and used a subset of census questions to collect basic demographic characteristics (such as age, sex, marital status, religion and economic activity). As response to the CCS was voluntary, maximising response rate was a strategic aim during planning and development of the 2021 CCS. The response rate for the 2021 CCS was 88%.

⁵ Note the adjustment for very young children mentioned above was applied after the CCS process.

Analysis of the CCS 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. As the CCS is a sample survey it is subject to sampling error. In overall terms, the effect of this sampling error on the census day population estimate has been calculated as equivalent to a 95% confidence interval about the population estimate (1,903,100) of +/- 4,700, or roughly 0.3% of the estimated population.

Further details on the whole coverage estimation process will be published in a technical paper in due course.

2.7. Quality assurance – outputs

In addition to the CCS, a Census Quality Survey (CQS) was also carried out, between October and December 2021. NISRA invited 5,000 households to take part online and they were asked questions from the census, again focusing and asking about their situation on 21 March 2021. The survey closed when the target sample size was achieved. The CQS will be used to estimate how accurately the Census 2021 questionnaire was completed by the general public. Further information will be published in due course that will help users to understand any strengths or limitations in Census 2021 data, based on the CQS responses.

An internal team was used to assess the quality of the results and to identify any issues that needed rectifying.

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

3. Quality assuring the census estimates

3.1. Comparison against administrative data sources

Census 2021 estimates were compared against a variety of statistical and administrative datasets to benchmark across all age groups.

The [March 2021 population estimate](#) produced by NISRA, is an estimate of the population at 21 March 2021. The March 2021 population estimate was derived using the standard cohort component method, and is in line with the standard mid-year estimate series. In summary, the [2020 mid-year estimates](#) have been used as a starting point, the population has been aged on 9 months (from July 2020 to March 2021), with the number of births in the 9 months added and the number of deaths in the 9 months removed. An adjustment has also been made for migration.

Figure 1 shows that the Census 2021 estimates closely align with the March 2021 population estimate, with the census estimate of 1.903 million just 0.2% higher than the March 2021 estimate of 1.899 million. There is more variation by age band between the Census 2021 estimate and the March 2021 estimate, with those aged 5 to 9 years 2.3% higher than the census estimate, and those aged 65 to 69 years 1.6% lower than the census estimate.

The Census 2021 estimates were compared against the active medical cards dataset from the Business Services Organisation Information Unit. The dataset is a count from April 2021 of anyone registered in Northern Ireland for a medical card who has some type of activity (collection of a prescription, changes to registration details or treatment by a dentist or optician) in the previous two years. Overall the Census 2021 estimate is 6.1% higher than the active medical card count of 1.787 million. The largest differences were seen in those aged 20 to 24 (12.0%) and 25 to 29 years (11.1%), however, the lower use of health services particularly by young males is widely acknowledged, making the number of active registrations a conservative or under-estimate of the population.

The school census, from the Department of Education, provides a count of those aged 5 to 14 years from the 2021/22 school census. The figures are very similar to

the Census 2021 estimates with the school census⁶ figure for 5 to 14 year olds being around 1,000 children higher (0.4%).

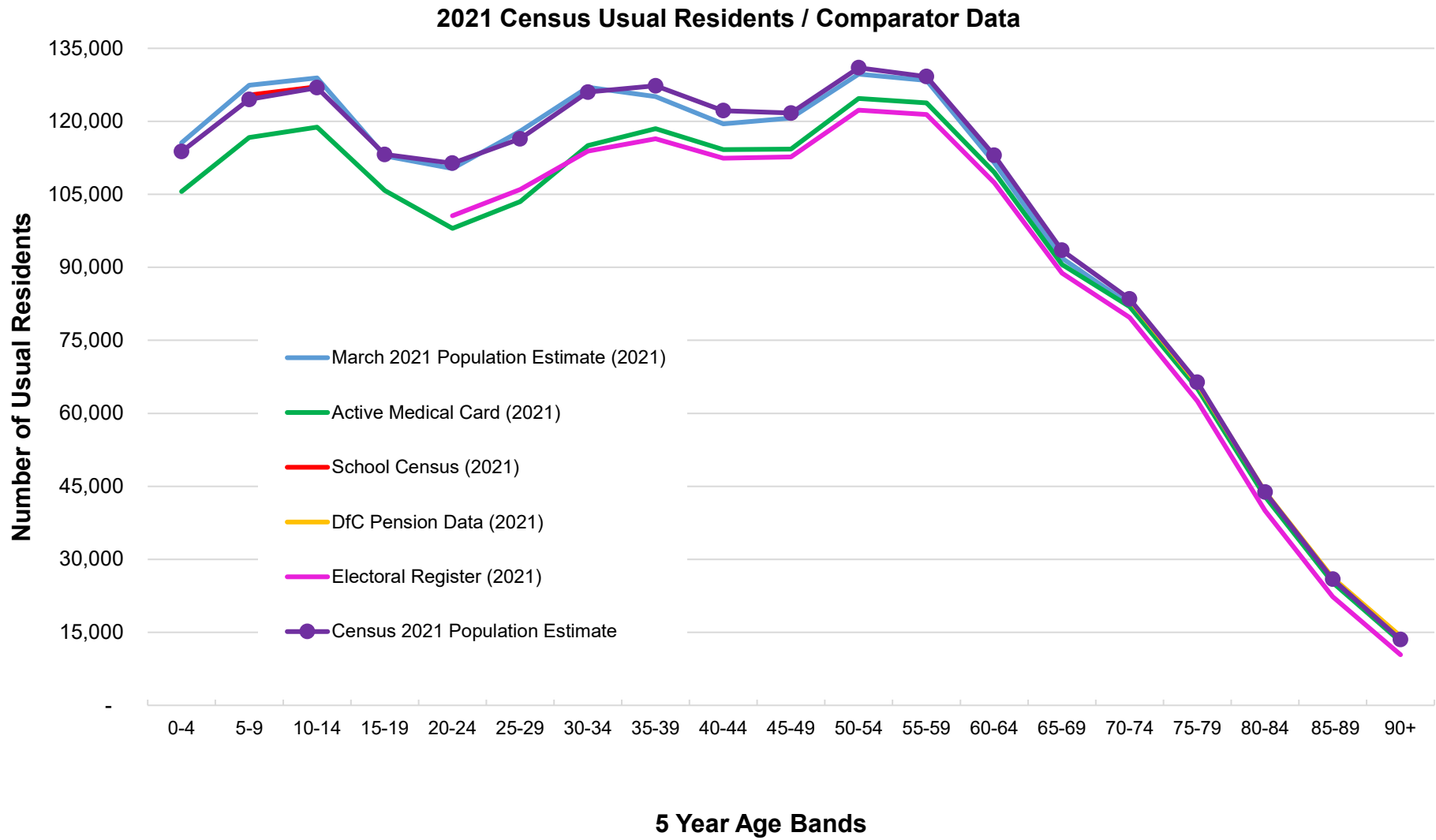
Data from the Department for Communities, on the number of people aged 70 (and over) claiming a pension⁷ showed very similar figures to the census estimates for those up to age band 85 to 89. Overall differences for those aged 70 plus showed that the pension data was just 0.3% higher than the census estimate.

Finally, the Census 2021 estimates were compared against the electoral register, supplied by the Electoral Office for Northern Ireland, and based on counts in December 2021. Overall there was a 7.6% difference with lower figures on the electoral register across all age bands.

⁶ It should be noted that the school census figures will contain a small number of children who although registered at a Northern Ireland school will usually live outside Northern Ireland – examples include children living in the Republic of Ireland, children registered for exam purposes only, etc.

⁷ Pension figures include all claimants that have provided a current Northern Ireland address regardless of whether they are resident.

Figure 1: All usual residents on census day



4. Conclusion

The census is the most complete source of information about the Northern Ireland population available, with significant effort made to include everyone. Quality has been at the forefront of all decisions taken throughout Census 2021.

A high response rate (97%) with the majority of person responses online (85%), coupled with a high response rate to the CCS (88%) were the result of a very successful data collection operation, with the vast majority of the Northern Ireland population engaging and completing their returns.

International standards were used for processing and to estimate those who did not respond, or who missed particular questions, to ensure that the results were representative of the entire population. The resulting Census 2021 Northern Ireland population estimates by age and sex align closely with comparator datasets.

One of the key uses of the decennial census is to provide a benchmark for the estimated population. NISRA plans to re-calibrate the 2011 Census based population estimates series on the basis of the 2021 Census data, and will publish the results in due course.

Census Office

24 May 2022