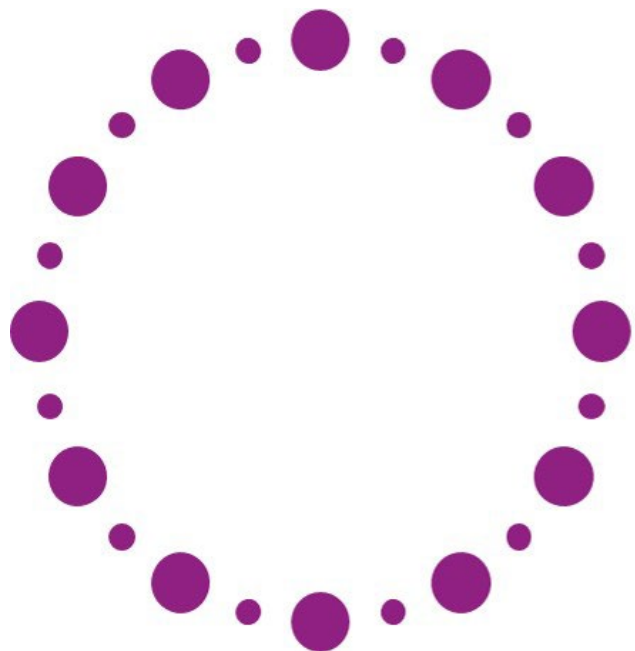


census
2021

Census Quality Survey

Information paper

May 2023



nisra.gov.uk/Census2021

Contents

1. Executive summary	3
2. Introduction	4
3. Methodology	5
3.1. Survey approach	5
3.2. Data capture arrangements	5
3.3. Preparing the data for analysis	6
3.4. Coverage and weighting of the data	7
4. Important information regarding the results	9
4.1. Definitions	9
4.2. Sampling Error	10
4.3. Points to note about question recall	11
5. Results	11
5.1. Household questions	12
5.2. Individual questions	12
6. Conclusion	18

1. Executive summary

The 2021 Census Quality Survey (CQS) was a voluntary survey conducted independently of the census in order to provide an insight into the overall quality of the information collected from Census 2021.

The CQS was conducted seven months after Census 2021 (i.e. October 2021) and involved a random sample of households that had responded to Census 2021. Participants were asked the majority of the Census 2021 questions and their responses were then matched to, and compared with, those provided in the census. The results, based on 1,351 households and 3,204 residents, are presented in the form of 'Agreement Rate' tables that show the extent to which the information provided in the census aligned with that collected through the CQS. At the outset, it was accepted that the census and CQS responses might not align precisely for several reasons, including the differing dates between the census and the CQS (e.g. there may have been changes to circumstances in that time period), and who in the household actually provided the responses.

Overall, the CQS shows that for the questions asked, there was a high degree of similarity between the surveys, with 27 of the 40 questions having agreement rates above 85% which indicates high levels of quality. For the 13 questions where the agreement rates were lower than 85%, a number of factors may have impacted on the agreement rates – not least changes in circumstances between the surveys and people's ability to recall detailed information seven months on from the census.

Taking all of this into account, and considering that all but one of the questions produced agreement rates that were above 70%, this provides reassurance that the information collected in Census 2021 is of high quality and fit-for-purpose.

2. Introduction

The 2021 Census in Northern Ireland was taken on 21 March 2021, in line with arrangements in England and Wales. The planning, development, testing and operational aspects of the census were underpinned by the following key strategic objectives:

- to provide high quality, value-for-money, fit-for-purpose statistics that meet user needs, and which are consistent, comparable and accessible across the UK;
- to protect, and be seen to protect, confidential personal census information;
- to deliver a high quality 2021 online census data collection operation;
- to maximise overall response rates and minimise differences in response rates in specific areas and among particular population groups;
- to secure public and user confidence in the final results and deliver them in a timely manner; and
- to encourage wider use and exploration of census results to facilitate greater benefits from census outputs.

In keeping with these key strategic objectives, a raft of user consultation, questionnaire design and question testing work was undertaken by NISRA (in conjunction with the other UK Census Offices) to optimise both the design of the Census 2021 questionnaire and the topic/question content.

The primary aims of this work were to ensure that:

- the census would collect the information required by users as identified through the consultation exercises;
- the questions would be readily understood by the public and would be easy to complete; and
- the information provided could be captured electronically to a high degree of accuracy.

As a result of this work, the vast majority of the 13 Household and 43 Individual questions included in the Northern Ireland Census 2021 questionnaire could be

answered by simply ticking a box (or combination of boxes) that the respondent considered best described their particular circumstances on Census Day.

To assess the overall quality of the information reported through Census 2021, NISRA conducted an independent voluntary survey of a random sample of households – called the Census Quality Survey (CQS). The aim of the CQS was to ask the Census 2021 questions again to provide some insight into how well the Census 2021 questionnaire had been understood by the public (see Annex A for a list of the questions asked and their corresponding Census 2021 question number).

This paper presents the background to the 2021 CQS methodology and the findings from it. In particular, it outlines the extent to which the responses provided in the CQS, to both the household and individual questions, agreed with those captured in the census.

3. Methodology

The CQS was an independent voluntary online survey taken seven months after Census Day and was targeted at a random sample of households that had responded to Census 2021 using the online platform. An online response approach was chosen for the CQS to mitigate against the risk of low response rates given the ongoing COVID-19 pandemic at the time. Communal establishments such as Care Homes and Student Halls of Residence were not included in the sample. In total, 2,500 households from across Northern Ireland were randomly selected and asked to take part in the CQS.

3.1. Survey approach

The CQS fieldwork commenced in late October 2021 with the issue of a letter to ‘The householder’ in each of the selected households. The letter invited them to take part in the CQS and included a secure access code for the online system. The online platform also advised that if recipients did not live at that property on Census Day then they should not complete the survey.

A total of 1,511 households participated in the CQS giving a response rate of 60.4%.

3.2. Data capture arrangements

The CQS online platform covered all the relevant household and individual questions that were asked in Census 2021.

Respondents were asked the same questions as in the census, using exactly the same wording and following the same routing. However, they were not asked to remember what answers they had provided in the census, rather they were asked to reflect their circumstances as on Census Day (21 March 2021).

The online platform also included a link to the Census Privacy Notice that reassured respondents that any information provided through the CQS would be treated in strict confidence, in line with the comprehensive Information Assurance protocols that underpinned the census.

3.3. Preparing the data for analysis

A total of 3,936 people were captured in the 1,511 households that responded to the CQS.

To prepare the data for comparison, these 3,936 people needed to be matched to the individuals captured in the same 1,511 households in the census. While the matching process was comprehensive, a full match rate was not expected given the interval between the census and the CQS (seven months). This meant that there was a possibility that some of the people captured in the CQS may not have lived at these 1,511 addresses on Census Day, and conversely some of the residents captured at those addresses in the census may have moved elsewhere since Census Day.

The matching began by cleansing the data to ensure that each record contained sufficient information to be included in the analysis (i.e. demographic information for matching and variable information for weighting – see section 3.4). If insufficient information was provided then these records were disregarded from the matching and analysis. This yielded a person level linkage rate of 81% (or 3,204 out of the 3,936 CQS respondents) and resulted in 1,351 household records (160 were disregarded as they contained zero valid residents after linkage).

3.4. Coverage and weighting of the data

In terms of coverage, the sample was drawn to ensure that there was as much geographic spread as possible. Spatial analysis of the 60.4% of households that responded to the CQS confirmed that there was appropriate coverage in all local government districts – therefore adjustment of the household data was not considered necessary.

The distribution of the resident data was then examined to ensure that it reflected, as far as was reasonably possible, the distributions in the census.

The profile of the 3,204 CQS people was compared to those in the census to see if any sub-group was under or over-represented. Three key characteristics were checked, namely:

- age (0 to 17 years; 18 to 64 years; 65 years or older);
- sex (Female; Male); and
- religion or religion brought up in (Roman Catholic; Protestant and Other Christian; and Other including None).

Figures 1 to 3 show the distribution of these characteristics for both the CQS people and those captured in the census. They show that the distributions of age, sex and religion differed between the datasets and that data weighting was needed to make the CQS dataset more representative of the census.

Figure 1: Comparison of age band (CQS vs Census 2021)

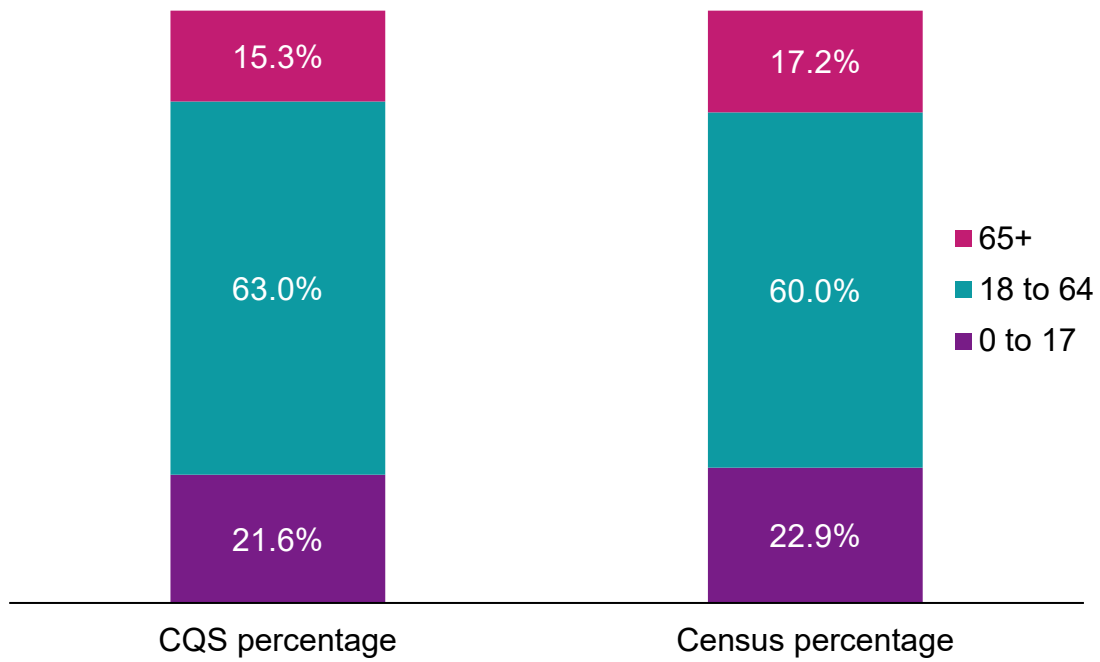


Figure 2: Comparison of sex (CQS vs Census 2021)

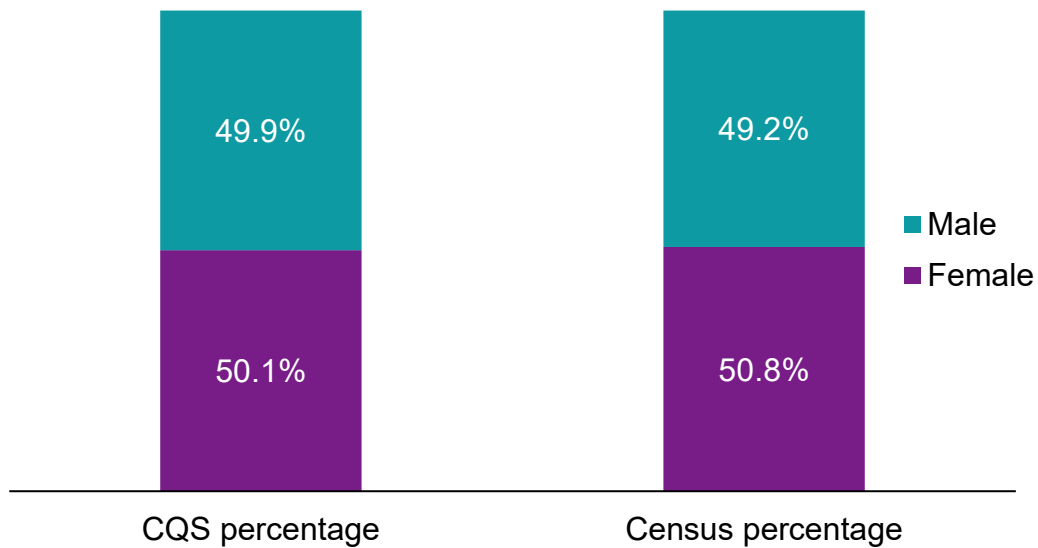
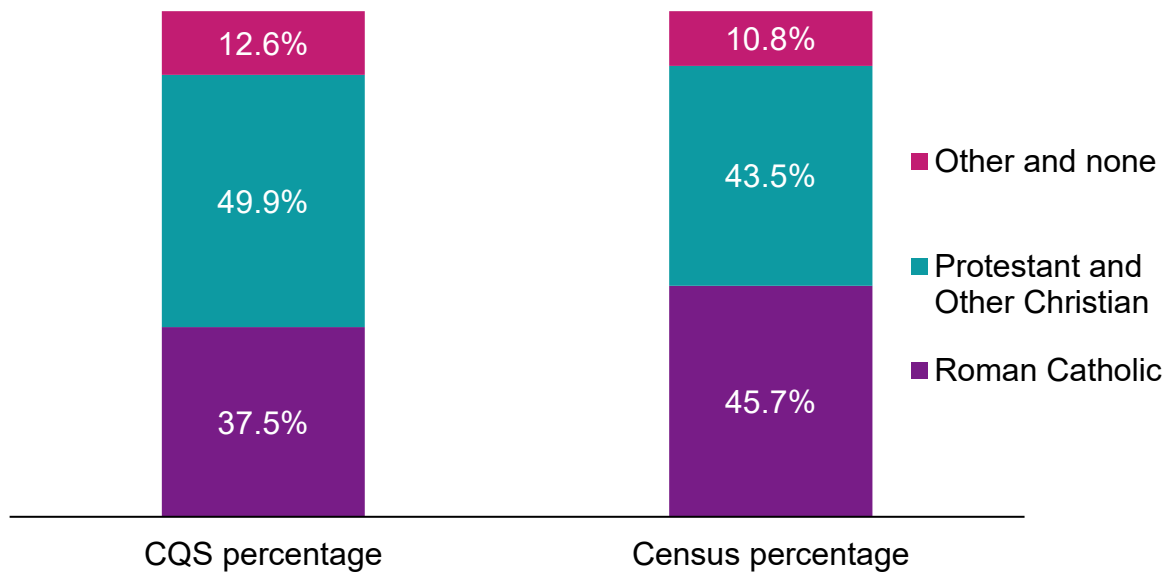


Figure 3: Comparison of religion or religion brought up in (CQS vs Census 2021)



4. Important information regarding the results

4.1. Definitions

The **Response Rate** for a question is the percentage of participants who were expected to answer the question who actually went on to answer it. As an example, if 240 people were routed to a particular question (based on their earlier responses) but only 180 actually responded then the response rate would be 180 out of 240 = 75%.

The **Agreement Rate** is the percentage of CQS responses to a question that agreed with the census responses to that same question, noting that we compared only records where there was a substantive response on both the CQS and the census.

The response rates and agreement rates are shown in Table 1 and Table 2 in Section 5. Table 1 shows the results for the household questions and Table 2 shows the results for the individual questions.

4.2. Sampling Error

Given that the CQS is based on a single random sample of households, the agreement rates shown are an estimate of the overall agreement rate for that question for all of Northern Ireland. As such, if another CQS sample was randomly drawn, then it might generate slightly different estimates of the Northern Ireland agreement rates. The spread of these estimates is known as the 'sampling variability' and we use a Confidence Interval (CI) to present the extent of the sampling variability.

The 95% CI given is a range (from the Lower Bound to the Upper Bound) within which the true Northern Ireland agreement rate would fall for 95% of all possible samples that could have been selected. This is the accepted way of expressing the statistical accuracy of a survey-based estimate. If based on a sample the estimate of an agreement rate has a large error level (called the Standard Error) then the corresponding Confidence Interval for that estimate will be wider.

The Standard Error (SE) of the agreement rate (AR) for N valid responses is given by:

$$SE = \sqrt{\frac{AR(1 - AR)}{N}}$$

The 95% Confidence Interval (CI) for the AR is then given by:

$$95\% CI = AR \pm 1.96 * SE$$

In this report, the CI is presented in terms of a Lower Bound and an Upper Bound for the agreement rate.

4.3. Points to note about question recall

The main purpose of the CQS is to provide insight into the overall quality of the information reported from Census 2021 by presenting the level of agreement between what respondents reported in Census 2021 and what they subsequently reported in the CQS.

However, it is accepted that the CQS and Census responses may not align precisely for reasons such as:

- the time interval – for practical reasons the CQS was conducted some seven months after Census 2021, which may have limited the participant's ability to recall exactly what their situation was on Census Day. In addition, in the intervening period, the participant may have taken a different perspective in terms of which response option best reflected their circumstances; and
- who actually completed the response - the responses provided to the individual questions in the census may not have been provided directly by the individual concerned (i.e. someone may have completed the census questionnaire on their behalf) and the CQS response may have been provided directly by the individual concerned or by someone different.

5. Results

This section presents the response rates and agreement rates for each question asked in the CQS. Agreement rates of 90% or more indicate very high levels of quality; those between 85% and 90% indicate high quality; those below 85% are considered further.

5.1. Household questions

Table 1 shows the response rates and agreement rates (including lower and upper bounds of the agreement rate) for the seven household questions asked in the CQS.

Table 1: Response Rates and Agreement Rates – Household questions

Question	Response Rate	Agreement Rate	Lower Bound	Upper Bound
Renewable energy systems	99.4%	98.5%	97.9%	99.2%
Tenure	99.9%	96.4%	95.4%	97.4%
Type of accommodation	99.8%	95.2%	94.0%	96.3%
Number of cars or vans	99.9%	88.4%	86.7%	90.1%
Landlord	99.7%	86.8%	83.0%	90.7%
Household adaptations	98.4%	86.6%	84.8%	88.4%
Central heating	99.9%	78.3%	76.1%	80.5%

The response rates show that almost all CQS respondents provided answers to these questions. Looking at the agreement rates, all questions apart from central heating have levels above 85% indicating high quality, with renewables, tenure and type of accommodation having levels above 90% - indicating very high quality.

Central heating (78.3%)

This was a 'Tick all that apply' question where respondents could select more than one of a relatively large number of tick boxes available – this in itself could have generated a vast number of response combinations. That level of complexity can be removed by looking at the agreement rates of each central heating type captured for households whether available solely, or in combination with others. Doing this shows agreement rates which are generally much higher. For example, the agreement rates for captured central heating types (whether solely or in combination with others) were 98.3% for 'Mains gas' and 89.3% for 'Solid fuel'.

5.2. Individual questions

Table 2 shows the response rates and the agreement rates (including lower and upper bounds of the agreement rate) for the 33 individual questions asked in the CQS.

Table 2: Response Rates and Agreement Rates – Individual questions

Question	Response Rate	Agreement Rate	Lower Bound	Upper Bound
Student's term-time address	100.0%	100.0%	100.0%	100.0%
Waiting to start work	98.3%	99.9%	99.6%	100.0%
Country of birth	99.9%	99.8%	99.6%	99.9%
Ethnic group	100.0%	99.5%	99.3%	99.7%
Students	100.0%	99.0%	98.6%	99.3%
Main language	99.7%	98.5%	98.1%	99.0%
Marital or civil partnership status	100.0%	98.3%	97.8%	98.7%
Looking for work	99.6%	97.0%	95.8%	98.1%
Employment status in main job	94.5%	96.4%	95.5%	97.3%
Sexual orientation	99.1%	94.8%	93.9%	95.6%
Apprenticeship	97.5%	94.6%	93.7%	95.4%
Passports held	99.7%	92.9%	92.0%	93.8%
Ability in Irish	99.7%	90.8%	89.8%	91.8%
Provision of unpaid care	99.2%	89.6%	88.5%	90.7%
Ability in Ulster-Scots	99.5%	89.3%	88.3%	90.4%
Long-term health problem or disability	99.7%	89.3%	88.2%	90.3%
Religion belong to	100.0%	88.7%	87.6%	89.8%
Employment last week	98.2%	88.6%	87.3%	89.8%
Available for work	100.0%	88.2%	77.4%	99.1%
Method of travel to main place of work or study	95.6%	87.1%	85.4%	88.8%
Supervision of other employees	98.8%	85.8%	84.2%	87.4%
Frequency of speaking Ulster-Scots	100.0%	83.6%	73.9%	93.4%
Religion brought up in	100.0%	81.2%	77.8%	84.6%
Other activity last week	99.8%	80.9%	78.4%	83.5%
Highest qualification	98.7%	79.3%	77.7%	80.9%
Year arrived to live in Northern Ireland	99.8%	77.4%	73.6%	81.1%
Long-term health conditions	99.6%	76.6%	75.2%	78.1%
Hours worked	94.6%	76.3%	74.3%	78.3%
National identity	100.0%	72.2%	70.6%	73.7%
Frequency of speaking Irish	99.5%	71.3%	63.7%	78.9%
General health	99.7%	71.3%	69.8%	72.9%
Proficiency in English	100.0%	70.8%	61.7%	79.9%
Ever worked	99.8%	52.9%	49.6%	56.1%

The response rates show that almost all CQS respondents provided answers to these questions. Looking at the agreement rates, 21 of the 33 questions have levels above 85% indicating high quality, within which, well over half (13) have levels above 90% - indicating very high quality.

The remaining 12 questions that have agreement rates below 85% are each considered below.

Frequency of speaking Ulster-Scots (83.6%)

The question on how frequently Ulster-Scots was spoken was asked only of those who indicated that they could speak Ulster-Scots – which was a relatively small group of respondents. Therefore the agreement rate can be disproportionately affected by small changes in the pattern of response. In addition, it could be difficult for respondents to recall their use of the language at the time of the census if asked seven months later, so it is possible that the interval between the surveys has also had an impact. Given this, an agreement rate of 83.6% is considered very good.

Religion brought up in (81.2%)

The question on 'Religion brought up in' was asked only of those respondents who had replied 'none' to the 'Religion belong to' question – which was a relatively small group of respondents. Therefore the agreement rate can be disproportionately affected by small changes in the pattern of response. Looking at where the responses differed, the majority of those were as a result of people reporting having a religion in the census to reporting no religion in the CQS. This could be because of changing circumstances or due to the voluntary nature of the CQS. Given this, an agreement rate of 81.2% is considered very good.

Other activity last week (80.9%)

The question on other activity last week was asked only of those who indicated that they were not working in the week prior to the census – which was a relatively small group of respondents. Therefore the agreement rate can be disproportionately affected by small changes in the pattern of response. In addition, it could be difficult for respondents to recall their economic status in the week prior to the census if asked seven months later, so it is possible that the agreement rate has been affected by the interval between the surveys. Given this, an agreement rate of 80.9% is considered very good.

Highest Qualification (79.3%)

This was a 'Tick all that apply' question where respondents could select more than one of a relatively large number of tick boxes available – this in itself could have generated a vast number of response combinations. That level of complexity can be removed by looking at the agreement rates of each qualification captured for individuals whether available solely, or in combination with others. Doing this shows agreement rates which are generally much higher. For example, the agreement rates for captured qualifications (whether solely or in combination with others) ranges from 93.1% for 'Degree level or above' through to 85.0% for 'NVQ or equivalent'.

It is also noted that it could be difficult for respondents to recall their qualifications gained at the time of the census if asked seven months later, particularly given the CQS took place after the summer months where, typically, educational qualifications are gained. Given this, an agreement rate of 79.3% is considered very good.

Year arrived to live in Northern Ireland (77.4%)

The question on year of arrival was asked only of those who indicated that they were not born in Northern Ireland – which was a relatively small group of respondents. Therefore the agreement rate can be disproportionately affected by small changes in the pattern of response. Where responses differed, they were much more likely to fall into neighbouring years suggesting that some respondents may have had difficulty recalling exactly what year they arrived. Given this, an agreement rate of 77.4% is considered very good.

Long-term health conditions (76.6%)

This was a 'Tick all that apply' question where respondents could select more than one of a relatively large number of tick boxes available – this in itself could have generated a vast number of response combinations. That level of complexity can be removed by looking at the agreement rates of each health condition captured for individuals whether available solely, or in combination with others. Doing this shows agreement rates which are generally much higher. For example, the agreement rates for captured conditions (whether solely or in combination with others) ranges from

99.7% for 'A mobility or dexterity difficulty that requires the use of a wheelchair' through to 88.7% for 'None'.

It is also noted that it could be difficult for respondents to recall their precise health status at the time of the census if asked seven months later, so it is possible that the interval between the surveys has also had an impact. Given this, an agreement rate of 76.6% is considered very good.

Hours worked (76.3%)

This question asked respondents to calculate their weekly hours worked if they were currently in employment or had ever worked. It is noted that it could be difficult for respondents to recall their basis for this calculation if asked seven months later, so it is possible that the agreement rate has been affected by the interval between the surveys. Given this, an agreement rate of 76.3% is considered very good.

National Identity (72.2%)

This was a 'Tick all that apply' question where respondents could select more than one of the tick boxes available, and could also provide written-in responses for identities that weren't covered by the tick boxes – this in itself could have generated a vast number of response combinations. In addition, the proportion of people who wrote in a National Identity rather than ticking a box was a relatively small group of respondents. Therefore the agreement rate can be disproportionately affected by small changes in the pattern of response. Given this, an agreement rate of 72.2% is considered very good.

That level of complexity can be removed by looking at the agreement rates of each national identity captured for individuals whether available solely, or in combination with others. Doing this shows agreement rates which are generally much higher. For example, the agreement rates for captured identities (whether solely or in combination with others) are 90.8% for 'Irish', 88.2% for 'British' and 81.9% for 'Northern Irish'.

Frequency of speaking Irish (71.3%)

The question on how frequently Irish was spoken was asked only of those who indicated that they could speak Irish – which was a relatively small group of respondents. Therefore the agreement rate can be disproportionately affected by small changes in the pattern of response. In addition, it could be difficult for respondents to recall their use of the language at the time of the census if asked seven months later, so it is possible that the interval between the surveys has also had an impact. Given this, an agreement rate of 71.3% is considered very good.

General health (71.3%)

It is noted that it could be difficult for respondents to recall their precise health status at the time of the census if asked seven months later, so it is possible that the interval between the surveys has had an impact. That said, collapsing the responses to from five to three categories (i.e. 'Very Good/Good', 'Fair' and 'Bad/Very Bad') increases the agreement rate to 89.1%.

Proficiency in English (70.8%)

The question on proficiency in English was asked only of those who indicated that English was not their main language – which was a relatively small group of respondents. Therefore the agreement rate can be disproportionately affected by small changes in the pattern of response. In addition, it could be difficult for respondents to recall their proficiency in English at the time of the census if asked seven months later, so it is possible that the interval between the surveys has also had an impact. Given this, an agreement rate of 70.8% is considered very good.

Ever worked (52.9%)

The question on Ever Worked was asked only of those who indicated that they were not working in the week prior to the census – which was a relatively small group of respondents. Therefore the agreement rate can be disproportionately affected by small changes in the pattern of response. In addition, it could be difficult for respondents to recall their employment information if asked seven months later, so it

is possible that the agreement rate has been affected by the interval between the surveys.

However, this agreement rate is relatively low and potentially points to an issue with the understanding of the question. As such, there are concerns over the accuracy of the information collected in this question and therefore it requires further investigation. At the time of publication the investigation has not concluded – further details will be published on the [Census 2021 quality notes, issues and corrections](#) page on the NISRA website when available.

6. Conclusion

Overall, the CQS shows that for the questions asked, there was a high degree of similarity between the surveys, with 27 of the 40 questions having agreement rates above 85% which indicates high levels of quality. For the 13 questions where the agreement rates were lower than 85%, a number of factors may have impacted on the agreement rates – not least changes in circumstances between the surveys and people's ability to recall detailed information seven months on from the census.

Taking all of this into account, and considering that all but one of the questions produced agreement rates that were above 70%, this provides reassurance that the information collected in Census 2021 is of high quality and fit-for-purpose.

Annex A

The table below shows each question topic referred to in this report and its associated question number within the [Census 2021 Questionnaire \(PDF, 692 KB\)](#).

Question Type	Question	Census Question Number
Household	Type of accommodation	H7
Household	Household adaptations	H8
Household	Central heating	H9
Household	Renewable energy systems	H10
Household	Tenure	H11
Household	Landlord	H12
Household	Number of cars or vans	H13
Individual	Marital or civil partnership status	4
Individual	Students	5
Individual	Student's term-time address	6
Individual	Country of birth	7
Individual	Year arrived to live in Northern Ireland	8
Individual	Passports held	10
Individual	National identity	11
Individual	Ethnic group	12
Individual	Religion belong to	13
Individual	Religion brought up in	14
Individual	Main language	15
Individual	Proficiency in English	16
Individual	Ability in Irish	17
Individual	Frequency of speaking Irish	17
Individual	Ability in Ulster-Scots	18
Individual	Frequency of speaking Ulster-Scots	18
Individual	General health	19
Individual	Long-term health problem or disability	20
Individual	Long-term health conditions	21
Individual	Provision of unpaid care	22
Individual	Sexual orientation	24
Individual	Highest qualification	25 and 26
Individual	Apprenticeship	27
Individual	Employment last week	28
Individual	Other activity last week	29
Individual	Looking for work	30
Individual	Available for work	31
Individual	Waiting to start work	32
Individual	Ever worked	33
Individual	Employment status in main job	35
Individual	Supervision of other employees	40
Individual	Hours worked	41
Individual	Method of travel to main place of work or study	43