Quality and methodology information (QMI) for Census 2021

Details the strengths, limitations, uses, users and methods used for Census 2021, England and Wales.

Table of contents

1 . Output information

- National Statistic: yes

- How compiled: Census 2021

- Frequency: decennial

- Geographic coverage: England and Wales

2 . About this Quality and Methodology Information report

This Quality and Methodology Information (QMI) report contains information on the quality characteristics of the data, including the European Statistical System's five dimensions of quality (PDF, 916KB). It also contains information about the methods we at the Office for National Statistics (ONS) used to create the data and outputs.

The information in this report will help you to:

- understand the strengths and limitations of the data

- understand quality considerations of the data

- learn about existing uses and users of the data

- understand the methods we used to create the data

- decide suitable uses for the data

- reduce the risk of misusing the data

Help us improve our Census 2021 content by completing our survey.

3 . Important points

- We, the Office for National Statistics (ONS), held Census 2021 in England and Wales on Sunday 21 March 2021.

- We achieved a very high census response rate, at 97% of the usual resident population of England and Wales and more than 88% in all local authorities (exceeding our targets of 94% overall and 80% in all local authorities).

- We take many steps to ensure that census products are high quality and trustworthy; this Quality and Methodology Information (QMI) report provides a summary, and our Maximising the quality of Census 2021 population estimates report provides further detail.

- Census 2021 is an important source of high-quality population data during the coronavirus (COVID-19) pandemic, but the circumstances may have affected some people's place of usual residence; Conducting a census during the coronavirus pandemic explains what this means for the data.

4 . Quality summary

Overview

The England and Wales census has happened every 10 years since 1801, except in 1941. It gives us detailed information about the characteristics of all the people and households in England and Wales. As per the Census Act 1920, it is a legal requirement for everyone in England and Wales to be counted in the census and to provide accurate information.

At the Office for National Statistics (ONS), we held the latest census in England and Wales on Sunday 21 March 2021. We ensured that our data and products would be high quality by setting and exceeding strategic aims for success in our Census White Paper, Help Shape Our Future. For example, we achieved a very high response rate, at 97% of the usually resident population of England and Wales and 88% in all local authorities. This exceeded our target of 94% overall and 80% in all local authorities.

This report describes the methodology we used to produce Census 2021 estimates and gives information about the quality of the census statistics.

Uses and users

Census data help a wide range of organisations plan for the future, and also underpin statistics such as gross domestic product (GDP), employment and coronavirus (COVID-19) rates.

Typical users of census data include:

- local authorities and other public bodies, for informing policy development, service provision and fund allocation

- businesses, for understanding customers and deciding where new stores should be located

- voluntary organisations, for learning about the communities they work in and supporting funding applications

- academics, for supporting research

Provisional Census 2021 data have also been used to meet demands for rapid and real-time data. This includes helping to inform our response to the coronavirus pandemic and support our humanitarian response to the Russian invasion of Ukraine.

Read our census stories on census.gov.uk to learn more about how different organisations use census information to plan services.

Strengths and limitations

Strengths

The census provides the most detailed picture of the entire population, with the same core questions asked to everybody across England and Wales. There is less margin for error in the census than with surveys based on a sample of the population, because the whole population is included.

The Scotland and Northern Ireland censuses ask the same core questions, making it possible to compare different parts of the UK. You can also compare the UK with other countries, as we align questions and classifications to international standards where possible.

The Office for Statistics Regulation has independently assessed the census estimates and checked compliance with the Code of Practice for Statistics. The UK Statistics Authority assigned National Statistics designation to Census 2021 outputs, providing assurance that these statistics are of the highest quality and value to users.

We undertook a rigorous and comprehensive quality assurance process, including comparing against the widest range of alternative and complementary data sources ever. In addition, for the first time, we invited local authorities to review provisional census estimates, drawing upon their local expertise, in parallel to our own quality assurance checks. Detailed information is provided in our Maximising the quality of Census 2021 population estimates report.

Census estimates are important for understanding the accuracy of other population estimates. For example, mid-year population estimates (MYEs) are based on the most recent census and adjusted for live births, deaths and migration, but the potential for error in MYEs increases over time between censuses. We are also using numerous data sources to produce more regular and timely national and local population estimates. Reports comparing the Census 2021 population estimates with the latest MYEs and admin-based population estimates (ABPEs), including explanations for any differences, are planned for publication later this year.

Our very high response rate and extensive online collection ensured that we have collected extremely high-quality data about the population and its characteristics on Census Day. It was particularly important to understand how the coronavirus (COVID-19) pandemic affected, and continues to affect, our population in a variety of ways (for example, health impacts, working from home). Census data and our ongoing transformation of our social statistics system will help us to both understand and to measure population change more effectively than ever before.

Limitations and mitigations

We only conduct a census every 10 years because of cost and burden. This means that the data can be updated less regularly than population statistics, which are estimated from other sources.

Census statistics are estimates rather than counts, and so have measures of uncertainty associated with them. We take numerous steps to minimise possible sources of error, as described in Accuracy.

The coronavirus pandemic may have affected some people's choice of usual residence on Census Day, for example, students and in some urban areas. These changes might have been temporary for some and more long-lasting for others. Conducting a census during the coronavirus pandemic explains what this means for the data.

No census is perfect – some people are inevitably missed or counted twice. Our Census Coverage Survey (CCS) enables us to estimate how many people have been missed or double-counted. We also have processes within the cleaning stage that check for and resolve multiple responses, allowing us to adjust the census counts accordingly. More information is provided in our Maximising the quality of Census 2021 population estimates report.

As with all self-completion questionnaires, some forms will have contained incorrect, incomplete or missing information about a person or household. We used editing and imputation strategies to correct inconsistencies and missing information. Further information will be provided in a more detailed report later this year.

Recent design and operation improvements

Since 2011, we have:

- conducted extensive stakeholder engagement, research and testing to inform the questions to be asked and their design on the census form; the changes since 2011 are summarised in Changes to questions asked in the census and described in more detail in our Help Shape Our Future White Paper

- adopted a digital-first approach to census completion and developed our search-as-you-type and address look-up functionality, helping us to achieve a very high response rate and improve the quality of the data

- improved our methodology for collecting, processing and quality assuring census data; we summarise these improvements in this report and our Maximising the quality of Census 2021 population estimates report, and will publish further information later in the year

Conducting a census during the coronavirus pandemic

Every census has unique circumstances. For Census 2021, the coronavirus pandemic in particular may have affected the data in different ways. It was important to understand the population and its characteristics during the pandemic. For example, early census data have already been used to inform our response to the coronavirus pandemic and to support our humanitarian response to the Russian invasion of Ukraine.

Census data users should be aware that total population statistics in the census first release will reflect circumstances in March 2021. For most of the population, the coronavirus pandemic would not have affected where they considered themselves resident.

For some students and in some urban areas, there is evidence that the coronavirus pandemic did result in changes to where people lived. These changes might have been temporary for some and permanent for others.

Students

The census counts students at their term-time address. Evidence of changes to the term-time population resulting from the coronavirus pandemic includes:

- analysis by the Higher Education Statistics Agency (HESA) showing a marked increase in the number of students who were not at term-time accommodation in the academic year 2020 to 2021, compared with the equivalent number in each year since the academic year 2016 to 2017

- a student hall survey we conducted shortly after the census to inform our production of census statistics, which found that some halls were below full occupancy

- insight from local authorities during quality assurance processes – for example, analysis from Bristol City Council and the University of Bristol found 60% to 70% of all students in the city were in residence, noting international students may never have come for the academic year

- analysis by the Home Office, which suggests the large increase in study visas for the year to March 2022 is because of students starting or resuming a deferred course, or changing from distance to in-person learning, after the pandemic-related restrictions on in-person attendance were lifted

Read more about how we ensured an accurate estimate of students.

Urban areas

The Greater London Authority (GLA) carried out analysis to understand population change in London during the pandemic. This analysis concluded that there had been a fall in London's population over the first year of the coronavirus pandemic, but that the population is likely to have started growing again since. Their analysis suggested this was attributable to:

- many young adults leaving London during lockdown, most likely linked to the temporary closure of the hospitality and tourism sectors

- higher mortality, mainly in those aged 75 years and over, and the continuation of a downward trend in the number of births

- an increased loss of other age groups to surrounding regions, as evidenced by house-price and registration data – seen as a potentially more persistent trend

Other areas, especially urban centres, may have also experienced similar effects.

The Greater London Authority analysis highlighted the potential temporary nature of this change. It pointed to evidence of many young adults returning to London during the spring and summer of 2021, following the recovery of the hospitality and tourism sectors. Several local authorities also referenced similar trends during the quality assurance process.

Future developments

We recognise the population continues to change and that we need to understand these changes. Using a variety of data sources, we will be providing more frequent, relevant and timely statistics to allow us to understand population change in local areas in 2022 and beyond. The results from Census 2021 will therefore provide an important bridge from the past to the future.

As we release further census statistics, we will publish supporting information to consider the impact the coronavirus pandemic may have had. This will include how the pandemic may have affected data on labour market and travel to work.

Changes to sex question guidance

The census question "what is your sex?" (female, male) has not changed since 1801. We changed supplementary guidance for the Census 2021 sex question from 9 March 2021, part-way through the collection period, in line with a court order.

The guidance originally read: "If you are considering how to answer, use the sex recorded on one of your legal documents such as a birth certificate, Gender Recognition Certificate, or passport." It was changed to: "If you are considering how to answer, use the sex recorded on your birth certificate or Gender Recognition Certificate." The same change occurred on equivalent guidance pages in Welsh.

We have used guidance pageviews and duplicate response submissions to measure the potential effect of the guidance change on census data. Analyses indicate any potential impact was very small. We have found no evidence showing that the change in guidance affected the high quality of census sex data.

Access of sex question guidance

Guidance webpages went live with the online census on 22 February 2021. Online census collection ended 30 April 2021. The sex question guidance received around 3,320 recorded pageviews during this time. Of these, around:

- 860 pageviews (25.9%) were prior to 9 March

- 360 pageviews (10.8%) were on 9 March

- 2,100 pageviews (63.3%) were after 9 March, of which around

- 270 pageviews (8.2%) were on 21 March, Census Day

When comparing against online returns received during the same period, you can use data from our Designing a digital-first census article to show that:

* 14.5% were before 9 March

* 2.9% were on 9 March

* 82.6% were after 9 March

* 21.5% were on 21 March, Census Day

It is not possible to link guidance pageviews to submitted responses. So, it is also not possible to determine how many:

* people used the guidance while filling in their census form

* guidance pageviews were from those looking only out of interest

* people changed their sex response while or after viewing the guidance

* guidance pageviews were from individuals who had already submitted their census response

Changes from duplicate responses

There were cases where duplicate census responses were submitted, for example, when someone submitted a separate individual form as well as being included on their household form.

In total, there were fewer than 100 instances where sex responses differed between a form submitted before the guidance change and a duplicate for the same person after the guidance change. Even for these few instances, it cannot be known if these differences were because of the guidance change.

It is not possible to know the effect of the guidance change for people who only submitted one response.

Future developments

These analyses are part of broader census quality assurance investigations. We will publish further evaluations of the quality of data from individual census questions, including the sex question, later this year.

5 . Quality characteristics

This section outlines different measures of data quality, reflecting the European Statistical System's five dimensions of quality (PDF, 916KB) and other important quality considerations.

Relevance

Relevance refers to how much the output meets user needs.

At the Office for National Statistics (ONS), we conducted extensive consultations with the main users of the census, to seek feedback in numerous areas, including the:

- design and development of the Census 2021 questionnaire

- operation of the census

- statistical processes

- statistical output

We carefully evaluated all the responses that users submitted against published criteria, and acted upon those that had a strong and clearly defined user need. This ensured that Census 2021 collected relevant, reliable and accurate data, prioritising topics with no comparable and accessible sources of information able to meet the need. For more information, see the detailed reports of the six consultations we have conducted since 2011.

Accuracy

Accuracy is the degree of closeness between an estimate and the true value the statistics were intended to measure.

The data that a census collects will inevitably contain errors, however well the census is designed. Errors can arise at all stages of the data collection and production processes.

Types of error

Users should be aware of possible types of error, so that they can assess the usefulness of the data for their own purposes.

Coverage error

This is error that occurs from failing to obtain some or all of the information from a member of the population. This includes when a person or household fails to respond to the census, which is called person or household non-response. It also includes item non-response, which is when an answer to a question is missing, invalid or inconsistent with the rest of the completed questionnaire. Finally, it also includes overcoverage, such as duplicate returns or individuals counted in the wrong location. We correct for coverage errors during our estimation process.

Measurement error

This error occurs from failing to collect the correct information from respondents. Measurement errors are made by the respondents themselves and may include misunderstanding what is required, responding multiple times (duplicates) or responding at the wrong address.

Modelling uncertainty

The population estimates are based on combining the census and Census Coverage Survey (CCS) data with logistic regression models to estimate how likely individuals and households are to respond. The models share the inherent limitations of all statistical models, in that they are constrained by their assumptions and cannot perfectly predict the outcome. Also, because the CCS is based on a sample rather than the entire population, the CCS data also contain sampling error, that is, the difference between the sample and the population.

Processing error

These are errors that happen during data processing before we produce the final estimates. They include errors in:

- geographical assignment

- data capture

- coding

- data loads

- editing

- coverage assessment and adjustment

How we reduced error

It is not possible to precisely calculate every form of error, however, we took various measures to ensure the effects of errors were minimised.

Questionnaire design and testing

The census questionnaire itself was carefully designed and tested.

Online census form

Our digital-first approach to census completion further reduced the risk of non-response and measurement error, as the online form requires respondents to use the specified response format and answer all compulsory questions. Details of the extensive quality control procedures we used throughout data collection to reduce the risk of measurement and processing error, will be available later this year.

Targeted field follow-up

Our digital-first approach, together with a questionnaire tracking system to monitor return rates in real time, enabled us to deploy field staff to areas with lower response rates for targeted follow-up. This further reduced non-response error. Our overall response rate of 97% confirms the effectiveness of these strategies. Further detail on how we estimated and adjusted the degree of both overcoverage and undercoverage is in our Maximising the quality of Census 2021 population estimates report.

We will publish details on national and subnational confidence intervals for the population estimates, which can be used as a measure of accuracy, later in the year.

Coherence and comparability

Coherence is the degree to which data derived from different sources or methods, but which refer to the same topic, are similar. Comparability is the degree to which data can be compared over time and domain, for example, at geographic level.

Comparing census population estimates with other data sources

Our Maximising the quality of Census 2021 population estimates report details our quality assurance process, including comparisons of census estimates with admin-based statistics for each local authority. We will also publish a comparison of census estimates with other population estimates, such as the mid-year estimates (MYEs) and admin-based population estimates (ABPEs), later this year.

Furthermore, we will publish topic-specific comparisons between census data and other data sources as part of our topic analysis. For example, we will analyse how labour market census data compare with data from the Labour Force Survey.

Changes to questions asked in the census

The majority of questions remained the same as in the 2011 Census, so analysis of trends over time is possible for most topics.

Some changes were made since 2011 to meet user needs and improve data quality. For example, in Census 2021, a new question was asked about past service in the UK armed forces, and those aged 16 years and over were also asked new voluntary questions on sexual orientation and gender identity. The census included these topics because user feedback showed they were highly important for equalities monitoring and service provision, and because no suitable alternative data sources existed.

We removed two questions from the census form between 2011 and 2021. The question on the number of rooms in a household was removed because suitable administrative data are available from the Valuation Office Agency, and the question on year last worked was removed because of high respondent burden.

Certain questions also had updated response options for 2021. For example, the question on central heating was updated to include options for renewable energy, and district or communal heat networks.

More detail on the changes to the questions asked in the census since 2011 is in the Help Shape Our Future White Paper. You can read more about question development and testing on our question development page.

Finally, we also undertook an evaluation of the Census 2021 questionnaire to understand the potential impact of the coronavirus (COVID-19) pandemic on how respondents answer questions. We then updated the question guidance accordingly to help respondents understand how questions should be answered in light of the pandemic. More information is in Updates to Census 2021 online questionnaire guidance.

UK comparability

The ONS carries out the census and produces outputs for England and Wales. The Northern Ireland Statistics and Research Agency (NISRA) and the National Records of Scotland (NRS) carry out separate censuses for Northern Ireland and Scotland, respectively.

The censuses in all of the UK constituent countries have been designed to be undertaken in a consistent manner and feature significant overlap in terms of topics and questions. This is laid out in the statement of agreement (PDF, 164KB) between the National Statistician of the ONS and the Registrars General for Scotland and Northern Ireland.

However, there are small differences between the three censuses, including when they were held. The censuses in England, Wales and Northern Ireland were held in March 2021, whereas Scotland held their census in March 2022. All UK census offices are working closely to understand how this will affect the production and comparability of UK-wide data. We will explain all differences in future publications on UK harmonisation.

Read more on our UK census data page.

Accessibility and clarity

Accessibility is the ease with which users can access the data, also reflecting the format in which the data are available and the availability of supporting information. Clarity refers to the quality and sufficiency of the release details, illustrations and accompanying advice.

We are developing the ONS website to accommodate users' needs for accessible online statistics. We want everyone who visits our website to have a positive experience, and to easily find and use the information they need. More information is in our accessibility statement.

We are also committed to meeting the needs of different users by releasing a range of supporting products, including:

- statistical bulletins

- digital content articles

- data visualisations

- supporting information reports

We have published a Census 2021 dictionary to provide more information about variables, definitions and classifications. These products will help users to understand and interpret the census data. Where feasible, products will be available in Welsh as well as English.

Timeliness and punctuality

Timeliness refers to the lapse of time between publication and the period to which the data refer. Punctuality refers to the gap between planned and actual publication dates.

The breadth and depth of census statistics means that Census 2021 data will be released in stages. The timetable is planned around user need and our aim was to ensure that statistics are released as soon as they are ready. Read more about our Census 2021 release plans.

We published the first Census 2021 estimates on 28 June 2022. This first release consisted of rounded population and household estimates for local authorities in England and Wales, including a breakdown of the population by five-year age bands and by sex. Our Census 2021 – the count is done, the data is in, so what happens next? blog outlines the extensive work undertaken between Census Day and the publication of the first results.

Definitions

We have published standard and derived variables, classifications and datasets in our Census 2021 dictionary.

The definitions used for Census 2021 aimed to be consistent with the international definitions where possible. We also work closely with NISRA and NRS to harmonise definitions across the UK censuses, where possible; read our UK Census Data webpage for more information.

Geography

Maintaining stability in small area statistical geographies to allow comparisons over time and across England and Wales was an important part of the design for Census 2021. However, in areas where Census 2021 indicated significant population change since 2011, changes to some 2011 Output Areas (OAs), Lower layer Super Output Areas (LSOAs) and Middle layer Super Output Areas (MSOAs) have been necessary.

It is expected that around 5% of 2011 OAs may change (through splits and mergers) so that all 2021 OAs (comprising unchanged 2011 OAs and new 2021 OAs) remain within established population and household thresholds. A smaller proportion of LSOAs and MSOAs are expected to change than for OAs. We will publish more information about the changes to small area statistical geographies between 2011 and 2021 later this year.

Census 2021 estimates for output geographies are aggregations of whole OAs, best-fitted to the geographies that were current as at the time of publication. This is the method used to produce all Census 2021 and other statistics, so that different statistics produced for the same geography are consistent, comparable and non-disclosive. The only exception is the Census 2021 results for national parks. These reflect the population within each park, rather than aggregations of OAs, as OA best-fit estimates were considered to be inappropriate for this largely rural geography.

Our Census 2011 report on An overview of best-fitting explains the methodology that will also be used for Census 2021 estimates.

Output quality trade-offs

Trade-offs are the extent to which different dimensions of quality are balanced against each other.

This output is subject to the two following trade-offs.

Improvement versus consistency

As with previous censuses, we made some changes to the census questions in 2021 in order to continue to meet user needs, as described in Changes to questions asked in the census. However, this necessarily limits comparability over time.

Quality assurance versus timeliness

We invited local authorities to quality assure provisional census estimates for their areas by comparing them with alternative data sources, in parallel with our own quality assurance methods. This is the most comprehensive quality assurance we have ever undertaken, but it necessitated a trade-off with timeliness.

Why you can trust our data

The ONS is the UK's recognised national statistical institute and its largest independent producer of official statistics. Our data policies detail how we collect, secure and use data in the publication of statistics. We treat the data that we hold with respect, keeping it secure and confidential, and using statistical methods that are professional, ethical and transparent.

You can read about how we ensure the results are of high quality and fit for purpose in our Statistical design for Census 2021, England and Wales article. This design was independently approved through an external review process. For more information, refer to our Methodological Assurance Review.

Trustworthiness is further assured through ongoing accreditation assessment. In June 2022, Census 2021 in England and Wales received National Statistics accreditation from the Office for Statistics Regulation (OSR), the regulatory arm of the UK Statistics Authority. This confirms that the published statistics adhere to the Code of Practice for Statistics and are of public value, high quality, and produced by trustworthy people and organisations. For more information, see How the ONS is ensuring Census 2021 will serve the public good.

6 . Quality considerations affecting the data

We advise users to review quality considerations when interpreting the Census 2021 data. These considerations affect all topics. Topic-specific quality considerations are linked in each of the published topic summaries.

Census statistics relate to standard definitions of, for example, who is counted as living in an area and what a household is. It is important to understand these definitions when comparing census statistics with other sources that may use different definitions.

Census data are adjusted to reflect estimated non-response. This is so that the published results relate to the entire usually resident population as it was on Census Day (21 March 2021), not just to people who completed a census questionnaire. More information can be found in how we process the data.

Census results are estimates including some uncertainty from the statistical models, which are used to estimate non-response, and a small effect as described in Statistical Disclosure Control. For this reason, they should not be interpreted as being an exactly correct count of the population for an area.

Uncertainty in the census estimates because of the estimation of non-response can be described using confidence intervals. Estimation for non-response is primarily conducted for five-year age groups and uncertainty will be greater for single years of age than for those combined ages. We take numerous steps to minimise possible sources of error, as described in Accuracy.

More information on how to correctly interpret data on students is described in Students.

19-year-olds

Single age figures are subject to uncertainty when disaggregating age-group estimates (which are themselves subject to a degree of uncertainty) to single year of age. This is described in more detail in the Modelling uncertainty section of Quality characteristics. This age group was particularly affected by adjustments made around students to account for those living in Halls of Residence.

Looking at the numbers and some characteristics of this group suggests a potential very slight overestimate. This would not affect the reliability of most analyses of the population but care should be taken if comparing derived rates for 19-year-olds with surrounding ages. This is particularly the case in areas with larger student populations.

Changes in population estimates since the first release

The first results from the 2021 Census were published in June 2022. These were estimates of numbers of households and of the population by five-year age-group for each local authority district.

These were produced before the final stages of processing the census data were completed. There are small differences between the estimates in that release and the final estimates provided in all subsequent releases.

The largest differences are in the age distribution of the population aged 80 years and over. The number of people estimated to be in the 80 to 84 years age-group has decreased by a rounded figure of 2,000. This reduction is counterbalanced by an increase of around 200 in the estimate for the 85 to 89 years age-group and around 1,700 for the 90 years and over age-group.

7 . Methods used to produce the data

How we collect the data, main data sources and accuracy

At the Office for National Statistics (ONS), we made every effort to ensure Census 2021 counted everyone. We used AddressBase Premium to ensure every household in England and Wales received invitations to complete their census questionnaire. We supplemented this with other data sources identifying communal establishments such as halls of residence and care homes. We ensured address data were as accurate as possible by working with GeoPlace to update our address frame prior to Census Day.

We gave individuals and households multiple ways to request a census questionnaire and add their address to our frame if they had not been contacted in advance. Alongside different strategies for counting people living in communal establishments, we put in place measures to contact population groups with no fixed abodes – such as caravaners, boaters and rough sleepers – to enable them to participate.

We encouraged participants to fill in an online census form rather than a paper questionnaire, where possible. Across England and Wales, 89% of households received a letter with an online access code, with an option to request a paper form if needed. Of the households in these "online first" areas, 94.2% completed their census form online. Read more in our Designing a digital-first census article.

The remaining 11% of households in England and Wales were in "paper first" areas. These were where we anticipated online completion rates to be low. We sent paper questionnaires, which also included an online access code, to all households in "paper first" areas. Overall, 46.4% of households in "paper first" areas completed their census form online.

During the main census collection period, we monitored local return rates. Census field officers visited households who had not responded after Census Day, offering support and providing paper forms or online access codes where needed. We also offered support in completing questionnaires through Census Support Centres, completion events and our contact centre.

How we process the data

Data capture, coding and cleaning

The challenge for this processing phase was making data collected through all of the response channels fit for later statistical processes and analyses. We achieved this in three stages.

The aim of Stage 1 was to transform the raw data into a standard format and carry out essential cleaning and validation tasks. We:

- scanned, captured and formatted data from paper questionnaires to match the electronic questionnaire data, which was captured automatically

- removed unwanted characters, such as emojis, from all written text responses

- checked the address on every questionnaire for accuracy and resolved any inconsistencies using the ONS Address Index Matching Service (AIMS)

Stage 2 was designed to consolidate and standardise responses collected from each of the different response channels. We:

- resolved multi-tick errors and differences in responses between paper and electronic questionnaires using a detailed set of data capture and coding rules

- gave numerical values to all written text responses using rule-based and statistical coding and classification methods; this aligned the data with standardised national coding frames, such as the Standard Industrial Classification (SIC) and Standard Occupational Classification (SOC)

Stage 3 focused on resolving three specific census data issues likely to have an adverse impact on census estimates and ongoing analyses.

The three issues were that:

- some questionnaires were returned with too little, or no, information

- by design or in error, some people submitted more than one response, leading to multiple questionnaires and duplicate individual data at a given address, which were not always consistent

- because of the coronavirus (COVID-19) pandemic, many students responded from their home address but also provided details of a term-time address

We resolved these issues by:

- removing the questionnaires with too little, or no, information from the data

- using a complex resolution method based on rules and statistical logic to combine all available data from multiple questionnaires and duplicate individual data into a discrete and coherent response

- putting in place a method to copy students to their term-time address, if they had not provided a return for that address

Overall, the capture, coding and cleaning process resolved a wide range of data issues and errors, improving the quality and utility of the data.

Edit and imputation

Editing and imputation strategies are first used on completed census forms that supply incorrect, incomplete or missing data.

For example, respondents may have accidentally missed a question or chosen to skip a question if they did not know or want to provide the answer. They may also have provided invalid responses that were inconsistent with other values on the questionnaire. Some of these inconsistencies can be identified within one person's record, for example, if a person gave their age as 5 years old and said they had a university degree. Others can be identified between two or more records in the same household, for example, a parent being younger than their child.

The digital-first data collection approach reduced these types of mistakes, as paper forms are more likely to contain incorrect values than online forms. This is because the latter includes in-built rules that require the respondent to answer compulsory questions and to use the expected range of responses before they can submit the online form.

Nonetheless, some returned census forms did still contain incorrect, incomplete or missing data. Leaving these mistakes in the data would make the census statistics look wrong, damaging trust in this valuable and important dataset. So, we used well-tested and world-renowned item-level edit and imputation strategies to correct inconsistencies and impute missing items, while preserving the relationships between census data characteristics. The observed data were passed through the Canadian Census Edit and Imputation System (CANCEIS) and any missing or inconsistent data were imputed. This meant that the census dataset was complete and consistent.

We will include more information about the strategy in our upcoming "Non-response, edit and imputation" article.

Census Coverage Survey

It is also necessary to impute whole missing persons or household records that were estimated to have been missed by Census 2021. To support this adjustment of population and household estimates, the Census Coverage Survey (CCS) estimates the degree of undercoverage.

The format of the 2021 CCS was broadly similar to the 2011 CCS. We conducted the CCS in the weeks following Census Day, sampling around 350,000 households in England and Wales in a randomly-selected sample of postcodes within Output Areas (OAs) arranged by local authority area. We used the Hard-to-Count index (Word, 1.28KB) to identify and include areas where people were less likely to complete the census form. Participation in the CCS was voluntary.

We will provide more information on how the CCS contributed to data processing work in future publications.

Coverage assessment and adjustment

The coverage assessment and adjustment (CAA) process uses the results of the CCS to identify and adjust for the number of people and households not counted, those counted more than once, and those counted in the wrong place in Census 2021. Firstly, we estimated undercoverage and overcoverage in the collected data by:

- matching the CCS records with those from Census 2021 using automated and clerical matching

- using the matched census and CCS data within a Dual System Estimation technique to estimate the number of people and households missed by the census

- searching the Census 2021 database for duplicates and using the CCS to estimate the level of overcount in the census

- estimating populations for each local authority by age, sex, ethnic group, economic activity and other important characteristics, balancing over-estimates and under-estimates, using a combination of statistical regression and small area estimation techniques

- assessing and correcting for biases

We then adjusted the data by:

- imputing whole households, including the people missing from the census count

- selecting "Donor" households through a process called Combinatorial Optimisation, considering a range of demographic and household characteristics – this included people with similar characteristics to those missing from the census

- adding individual and household characteristics to the imputed records

- taking measures to try and find the best possible match between donor and missing households, including the use of administrative data

- imputing characteristics using CANCEIS methods

We will provide more information on coverage assessment and adjustment processes, including specific adjustment models, in future publications.

How we quality assure and validate the data

We outlined our quality assurance strategy in our statistical design for Census 2021. Our strategy ensures that:

- we produce high-quality data

- we provide confidence to our users

- our estimates accurately reflect the population of England and Wales

We implemented various checks to reduce the possibility and impact of errors that could emerge during collection and processing of census data, as discussed in Quality characteristics. We did this by validating responses, monitoring the success of processes and making corrections to missing or incomplete data.

We also used three further main quality assurance methods:

- comparing Census 2021 population estimates with rolled-forward 2020 mid-year estimates (MYEs) and admin-based population estimates (ABPEs), to identify and correct potential errors in census data

- inviting local authorities to assist our quality assurance by reviewing provisional census estimates, and using their feedback in our validation process alongside the investigations already under way

- using the small-sample voluntary Census Quality Survey (CQS), carried out several weeks after Census Day, to estimate the level of error by comparing census and CQS question responses

More information on these and other quality assurance processes is included in our Maximising the Quality of Census 2021 population estimates report. We will publish further quality evaluations later this year.

How we disseminate the data

Outputs and analysis release schedule

Proposals for the release of data, analysis reports and other products are shown on our release plans webpage. The order in which data and analytical products are being released has been carefully considered, incorporating user feedback given during the Census 2021 outputs consultation.

The first results from Census 2021 will be followed by a series of topic summaries. These are data and commentary grouped into topic-based themes, based on "univariate" datasets, which provide breakdowns of one variable only. After this, we will release "multivariate" data, which are datasets combining two or more census variables, and more detailed census analysis. Read more about our proposals for the census analysis release schedule.

As with previous censuses, Census 2021 data will be hosted on both the ONS website and on Nomis, which also hosts data from earlier censuses. The data will also be directly accessible to third parties through an Application Programming Interface (API). All standard outputs are free under the Open Government Licence. Further information is also available through the:

- ONS's terms and conditions for data

- National Archives' copyright and reuse of published data

- UK Statistic Authority's conditions for pre-release access to data

Although we will continue to release some ready-made datasets, for Census 2021 we have also developed a new approach to enable users to make their own datasets by specifying a combination of variables, classifications and geographies.

If users require datasets that cannot be accessed or produced using the ONS website, they are able to commission datasets, which will incur a charge. For more information, please see our guidance for making a Census Commissioned Table request.

Microdata

In addition to aggregate census data, we will also provide access to samples of anonymised census microdata, which provide record-level data for individuals and households.

The conditions under which users can access microdata samples depend on the level of detail in each file, for example, the geographic information attached to each record. Everyone can access the publicly available microdata sample. The "safeguarded" Census 2021 microdata samples will only be available to data analysts through the UK Data Service (UKDS), having agreed to standard UKDS terms and conditions approved by the ONS. The most detailed "secure" microdata samples are stored within our Secure Research Service and are only available to approved or accredited researchers. Our website provides more information about our plans for Census 2021 microdata samples.

We aim to make as much census microdata publicly available as possible, where it is non-disclosive to do so.

Origin-destination data

We will also release origin-destination census datasets, which measure patterns of travel-to-work and migration between areas. These too will be classified as public, safeguarded, or secure data, depending upon their level of detail.

You can read more about how to access Census 2021 origin-destination data on our website, which includes a summary of how the coronavirus (COVID-19) pandemic and EU exit may have affected origin-destination data.

Statistical disclosure control

Statistical disclosure control is used to protect the confidentiality of census respondents.

With the exception of microdata, all census data are aggregated by [geography](). After aggregation, there could be a risk of individuals and businesses being identified if a characteristic, or combination of characteristics, they possess is uncommon within an area – for example, if they are the only person employed in a specific occupation. The risk of disclosure is heightened at smaller geographies, where people may be well known within their community.

Our [Statistical Disclosure Control (SDC) for 2021 UK Census (Word doc, 257 KB)]() report sets out the strategies used to lower the risk of disclosure. Targeted record swapping is used to move individuals or households with a higher risk of identification to different geographical areas. These individuals and households are swapped with others, matching on some basic characteristics so that the total population count is unchanged. In addition, to protect against disclosure by differencing, a small amount of noise is added to some cells in published datasets through the "cell-key method", also meaning that users would be less certain about whether small counts relate to the records of real individuals. More information about the processes used in SDC will be provided in future publications.

8 . Other information

How to cite this documentation

To cite this documentation, please use the following format:

Office for National Statistics (ONS), released 2 November 2022, ONS website, methodology, [Quality and methodology information (QMI) for Census 2021]()

Other useful links

For other information on the quality or methodology of census statistics, please see:

[Maximising the quality of Census 2021 population estimates report]()
Methodology | Released 28 June 2022
How we maximised the quality of Census 2021 population estimates during the processing and quality assurance of the final statistics.

[Compare age-sex estimates from Census 2021 to areas within England and Wales]()
Methodology | Released 28 June 2022
An interactive tool to compare local authorities in England and Wales using age-sex estimates and quality assurance information.