User guide to Census 2021 microdata samples, England and Wales

Supporting information for microdata samples from Census 2021 in England and Wales, including how to access and use public, safeguarded, and secure microdata samples.

Contact:
Harriet Berner
census.customerservices@ons.gov.uk
+44 1329 444972

Release date:
7 June 2024

Next release:
To be announced

Table of contents

1. Overview

2. Protecting confidentiality within microdata samples

3. Comparing microdata samples with the census

4. You must select a population base

5. Variables and categories

6. Industry and Occupation variables

7. Sampling method

8. Public microdata teaching sample

9. Safeguarded microdata samples

10. Secure microdata samples

11. Quality considerations for Census 2021 microdata samples

12. Census 2021: Quality and methodology information

13. Workplace zones

14. Census microdata for major data linkage projects

15. Previous uses of census microdata

16. Related links

17. Cite this methodology

1 . Overview

This user guide helps you to use, analyse, and interpret Census 2021 microdata samples for England and Wales.

About microdata

Microdata are small samples of individual records from a single census and from which we have removed any identifying information. They contain a range of individual and household characteristics. This means that you can use them to carry out analysis not possible from standard census outputs, such as:

- creating tables using bespoke variable combinations
- investigating specific combinations of variables or categories in a high level of detail
- conducting non-tabular statistical analyses on record-level data

Who uses microdata

Many types of people and organisations use census microdata, for example:

- government
- academics
- local authorities
- research institutes
- market research organisations
- independent public interest groups
- commercial researchers

They can use microdata samples for a wide range of purposes.

Quality information

Census 2021 collected responses during the coronavirus (COVID-19) pandemic, a period of unparalleled and rapid change. International and national travel restrictions alongside other social distancing measures of the pandemic affected the information gathered for several topics. Take these factors into consideration when using Census 2021 microdata samples. The Quality and methodology information (QMI) for Census 2021 explains what this means for the data.

Products to suit different needs

To make data available as widely as possible, and to maximise benefits from the census, we have several different Census 2021 microdata products. These products strike a balance between detail and security, and ensure microdata are available for inquiring citizens through to expert analysts. Consequently, the release of Census 2021 microdata samples is across three settings.

Our public microdata teaching sample is available to download from our website with minimal conditions applied to their use as stated within the Open Government Licence (OGL). It provides an educational tool to assist with teaching of statistics and social sciences.

Read more about our Public microdata teaching sample.

Our safeguarded microdata samples are only available to data analysts through the UK Data Service in line with previous censuses. Data analysts must register with the UK Data Service and agree to the terms and conditions of the UK Data Service End User Licence.

Read more about our Safeguarded microdata samples.

Our secure microdata samples are only available to accredited researchers through the Integrated Data Service.

Our public microdata sample was released on 7 September 2023; our safeguarded samples were released on 18 October 2023. The Census 2021 secure microdata sample will be available to access in early 2024.

Census 2021 microdata samples range in size from 10% to 1% of households or individuals. The secure microdata samples provide the highest level of detail and have the largest sample size.

Both individual and household samples are available and contain variables relating to individual and household characteristics. Individual samples include person-level data for sampled residents within a household or communal establishment. Household samples include person and household-level data for each resident in sampled households; vacant households in the sample include only household-level data.

Household samples allow linkage between individuals in the same household. This is not possible using the individual sample. Our household samples enable researchers to understand individuals within their household context. The household and individual samples contain different persons to protect the confidentiality of individuals across public and safeguarded samples.

Summary of Census 2021 microdata samples

Public microdata teaching sample:

- sample size: 1% of individuals
- statistical unit: persons
- lowest geography: Wales, regions within England, Inner and Outer London
- number of variables: 19, low detail

Safeguarded individual microdata sample at region level:

- sample size: 5% of individuals
- statistical unit: persons
- lowest geography: Wales, regions within England, Inner and Outer London
- number of variables: 89, medium detail

Safeguarded individual microdata sample at grouped local authority level:

- sample size: 5% of individuals
- statistical unit: persons
- lowest geography: grouped local authority
- number of variables: 87, low detail

Secure individual microdata sample:

- sample size: 10% of individuals
- statistical unit: persons
- lowest geography: local authority
- number of variables: 189, maximum detail

Safeguarded household microdata sample:

- sample size: 1% of households, includes only those with household size under nine persons
- statistical unit: households and all persons within sampled households
- lowest geography: Wales; regions within England; Inner and Outer London
- number of variables: 56, low detail

Secure household microdata sample:

- sample size: 10% of households
- households and all persons within sampled households
- lowest geography: local authority
- number of variables: 194, maximum detail

UK microdata

Our Census 2021 microdata samples cover England and Wales. Northern Ireland Statistics and Research Agency and National Records of Scotland produce similar products for Northern Ireland and Scotland.

In England, Wales and Northern Ireland the census was conducted in March 2021. In Scotland the decision was made to move the census to March 2022 because of the impact of the coronavirus pandemic. This difference needs to be considered when performing any UK-wide analyses using census microdata.

Comparability with the 2011 Census

Every effort has been made to maintain high levels of comparability between Census 2021 and the 2011 Census. Where possible, variables and output categories within our microdata samples have been harmonised to allow comparisons.

Expert help from within and outside the ONS

A range of internal and external stakeholders helped to design, create and disseminate Census 2021 microdata samples so we could best meet our users' needs.

Internal members of our working group included topic experts on:

- migration
- travel to work
- demography and census transformation
- population estimates
- statistical disclosure control

External members of our working group included:

- National Records of Scotland
- Northern Ireland Statistics and Research Agency
- Welsh Government
- UK Data Service
- local authorities
- academia
- market researchers
- commercial researchers
- community group

2 . Protecting confidentiality within microdata samples

Our microdata samples are designed to protect the confidentiality of individuals and households. We do this by applying access controls and removing information that might directly identify a person, such as names, addresses and date of birth.

Record swapping is applied to the census data used to create the microdata samples. This is a statistical disclosure control (SDC) method, which makes very small changes to the data to prevent identification of individuals. The microdata samples use further SDC methods such as collapsing variables and restricting detail.

The samples also include records that have been edited to prevent inconsistent data and contain imputed persons, households, and data values. To protect confidentiality, imputation flags are not included in any Census 2021 microdata sample. This was also the case for 2011 Census microdata.

The number of records that are within the sample and unique within the census database was measured as a proportion of the number of records that are unique within the sample. The level for this proportion was set to determine sufficient uncertainty.

Find out more about protecting personal data in Census 2021 results. There is also lots of information on disclosure control methods for a variety of statistics on our disclosure control page.

3 . Comparing microdata samples with the census

All our microdata samples have been subject to quality assurance processes. For comparability of Census 2021 microdata samples with the full Census 2021, open our Comparing microdata samples with Census 2021 dataset.

In the document, we compare category proportions for selected microdata sample datasets with the equivalent datasets from the full census to show that the samples are representative. Some of the datasets have a single variable (univariate) and some have multiple combined variables (multivariate). Take care when generalising findings to a wider population, as reliability of results could be affected.

4 . You must select a population base

Before you carry out any analyses using our Census 2021 microdata samples select the population base for the analyses and filter the microdata sample accordingly using the variable “USUAL_SHORT_STUDENT”.

Our Census 2021 microdata samples include data from the total England and Wales population, which includes:

- usual residents
- non-UK born short-term residents staying 3 to 12 months
- students living at an alternative address during term-time

The main population base for many published datasets is the usual resident population. A usual resident is anyone who on Census Day, 21 March 2021, was in the UK and had stayed or intended to stay in the UK for a period of 12 months or more or had a permanent UK address and was outside the UK and intended to be outside the UK for less than 12 months.

To conduct analysis on usual residents within our microdata samples, filter the variable “USUAL_SHORT_STUDENT” to include only “is a usual resident” responses.

Students and schoolchildren in full-time education studying away during term-time are counted as usually resident at their term-time address.

Students were asked to provide basic demographic information only (name, sex, age, and marital status) for their non-term time (home or vacation) address. Data for students at their non-term time address are available by filtering the variable “USUAL_SHORT_STUDENT” to include only “Is a student living at an alternative address in term time”.

If filters are not applied to the microdata samples using the “USUAL_SHORT_STUDENT” variable, then basic demographic variables for some individuals may be duplicated. This is because individuals can be present in the sample as both a “usual resident” and a “student living at an alternative address in term-time”.

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. Find out more about what this means for the data in conducting a census during the coronavirus pandemic.

5 . Variables and categories

You can find all the variables contained in our microdata samples and the categories included in our Microdata sample codes: Census 2021 dataset. This file is designed to be machine-readable; filters can be applied to show the information for each sample separately.

The following geographical variables are based on geography boundaries as of 31 December 2022:

- "REGION"
- "LA"
- "GLTLA22CD": Grouped local authority
- "IOL22CD": Inner and Outer London former residence indicator
- "MIGRANT_LA"
- "FM_IOL22CD": Inner and Outer London former residence indicator
- "WORKPLACE_LA"
- "SECOND_ADDRESS_LA"

6 . Industry and Occupation variables

The variables of Industry and Occupation are available in all Census 2021 microdata samples for people aged 16 years and over who have ever worked. To conduct analysis only on those in current employment in the week before the census, one of the following filters needs to be applied:

- set “ACTIVITY_LAST_WEEK” to include only those who were working
- set “ECONOMIC_ACTIVITY” to include only those who were economically active in employment

You can find more information on variables included in standard census outputs, some of which are used within the census microdata samples, in our Census 2021 dictionary.

7 . Sampling method

When considering inferences to be investigated, analysts should account for the sample design.

Census 2021 microdata products follow similar design principles to those used for 2011. The principles are that:

- smaller, less disclosive microdata are sampled from microdata at the next level up the security hierarchy; first we sample the secure individual microdata, then we sample the safeguarded individual microdata from the secure individual sample – this also applies to the household microdata

- our household microdata samples do not contain any of the persons who are contained within our individual microdata samples

- our safeguarded individual samples at region and at grouped local authority level do not overlap – they do not contain any of the same people

- sample sizes are consistent with the 2011 Census and are based on statistical disclosure risk assessments

When sampling, the source dataset was first sorted by local authorities, then Output Areas (OAs) within local authorities. This matches the 2011 approach and ensures the samples are evenly spread across England and Wales. Small additional improvements to the 2011 approach have been implemented; these are considered to have a negligible impact on the comparability with 2011 Census microdata samples.

For individual microdata samples the source data were further sorted by age and sex within each OA to maximise the representativeness of these samples in relation to age and sex.

For household microdata samples the source data were further sorted by household size to maximise the representativeness of these samples in relation to household size. For the safeguarded household sample, to protect confidentiality, households containing more than eight persons were removed from the sample frame as these were found to increase the disclosure risk beyond that acceptable for safeguarded data. For the individual microdata samples, systematic random sampling was used to select people at regular intervals. For the household samples, selection of households was at regular intervals. These intervals were determined by the size of the samples and the sampling frame.

As in 2011, the sampling method results in equal probabilities of inclusion for all individuals and households, except for high-risk records, which are removed to protect confidentiality. Sample weights for households and individuals in each sample are therefore equal and relate to the sample size. Consequently, sample weights are not provided.

Variable estimates from the microdata samples are subject to sampling error, which arises from the fact that they are based on a sample rather than a full census of the population. Sampling error is the difference between the estimates derived from a sample and the true population values. Take care when generalising findings from any sample to a wider population, with due consideration to the reliability of results.

8 . Public microdata teaching sample

Our public microdata teaching sample consists of a random sample of 1% of person records from Census 2021 for England and Wales; it includes records for 604,351 persons. Download our Public microdata teaching sample, England and Wales: Census 2021 dataset.

The primary purpose of our public microdata teaching sample is as an educational tool to:

- encourage wider use of census data by providing a way of examining census data beyond the standard tables
- introduce the detail, metadata and data formats included in microdata products in order to give users the skills and information necessary to make use of the more detailed products available
- assist with the teaching of statistics and geography at GCSE and higher levels

Data made available under the public mechanism have the following characteristics. Data can be downloaded from the ONS website for any purpose with minimal conditions applied to their use, as stated within the Open Government Licence (OGL). Disclosure risk is managed primarily by the design of the dataset, with the conditions of the OGL ensuring that no attempt is made to re-identify the data.

Given the relatively small sample size and the limited detail on each of the characteristics, our public teaching sample may not always be the most appropriate dataset for analysis. Please consider whether creating a custom dataset, topic summaries or alternative census outputs covering the entire population would better meet your requirements.

The Open Government Licence, which applies to our public microdata teaching sample, allows unrestricted use of government data as long as the source is acknowledged, and the data is not misrepresented.

Users reproducing ONS content without adaptation should include a source accreditation to: “ONS: Source: Office for National Statistics licensed under the Open Government Licence v.3.0”.

Users reproducing ONS content which is adapted should include a source accreditation to: “ONS: Adapted from data from the Office for National Statistics licensed under the Open Government Licence v.3.0”.

9 . Safeguarded microdata samples

Both our safeguarded individual microdata samples consist of a random sample of 5% of person records from Census 2021 for England and Wales. Our safeguarded household microdata sample consists of a random sample of 1% of households and contains records for all individuals within these sampled households. By design, all our safeguarded samples include a different set of people.

Our safeguarded Census 2021 microdata samples contain the following numbers of records:

- individual sample at region level: 3,021,455 persons
- individual sample at grouped local authority level: 3,021,611 persons
- household sample at region level: 263,729 households, 606,210 persons; for households containing no enumerated persons, records in the sample will only include household-level data

Our safeguarded Census 2021 microdata samples are only available to users who have registered with the UK Data Service and who have agreed to the terms and conditions of the UK Data Service End User Licence. Commercial use of the safeguarded data is subject to licensing and each project incurs administrative fees.

Data made available under the safeguarded mechanism have the following characteristics:

- data can be downloaded from the UK Data Service into the researcher's local environment
- researchers can only use data for statistical research under a set of conditions that limit and control purpose and behaviour; conditions of use are set out within the End User Licence
- disclosure risk is managed by the combination of the licence agreement with the researcher and the disclosure controls applied within the design of the dataset

Further information on what safeguarded data can be used for is available in section 5.2.1 of the Research data handling and security guide.

The citation that should be used when publishing research based on our safeguarded microdata samples is specified in the accompanying metadata on the UK Data Service website.

Safeguarded individual microdata samples

We have two individual safeguarded microdata samples, each at a different geography level. The sample at region level provides less detail for geography but a greater level of detail in the other variables. The sample at grouped local authority level provides greater geographic detail but consequently provides slightly less detailed output categories for variables.

Grouped local authority is a geography created specifically for our microdata samples. It comprises groups of local authorities, or single local authorities where the population reaches at least 120,000 persons. A list of grouped local authority groups is available from the open geography portal. Local authority groups used for 2021 have been made as similar as possible to those used for 2011. Some local authorities have encountered boundary changes between 2011 and 2021 and the nine-digit geography code has changed as a result. Details of changes are available from the open geography portal.

Safeguarded household microdata sample

Our safeguarded household microdata sample includes records for all individuals within sampled households. It enables linkage between individuals in the same household. Following the 2011 Census, this was only possible using the secure household sample. This new product for Census 2021 has been created following user feedback from the 2011 Census.

Our safeguarded household sample size is smaller than the individual safeguarded samples and it contains fewer variables and less detail for variable classifications. This is a consequence of providing details on entire households, which have greater disclosure risk when characteristics on their individuals are combined, compared with sole individuals.

In our safeguarded household sample, disclosive individual variables have been replaced by household equivalents providing functional but limited detail on full households compared with the safeguarded individual samples. For example, the ethnic group of individuals has been replaced with whether household members have the same ethnic group or if there are combinations of multiple ethnic groups within the household. To protect confidentiality, the sample does not include households that contain more than eight people.

Our safeguarded individual microdata samples contain many more variables that cannot be included in the safeguarded household microdata sample because of the increased risk of disclosure posed by the household file. Use the individual samples where linkage between individuals in the same household is not required. The secure household file is available in cases where the safeguarded household file does not meet user needs. Users should consider their needs and select the most appropriate sample for their research.

10 . Secure microdata samples

Our secure individual microdata sample consists of a random sample of 10% of person records from Census 2021. Our secure household microdata sample contains a random sample of 10% of households. The household file allows linkage between individuals in the same family and the same household. Our household and individual samples contain different people.

Our secure Census 2021 microdata samples contain the following numbers of records:

- individual sample: 6,204,787 persons and empty household spaces
- household sample: 2,641,775 households, 6,097,307 persons and empty household spaces; for households containing no enumerated persons, records in the sample will only include household-level data.

Our secure microdata samples represent the products with the highest level of detail and the largest sample size. As a result, they are only available under the secure mechanism and have the following characteristics:

- data are protected in law
- data cannot be distributed outside of the controlled environment; access to data will only be made available to accredited researchers, working on accredited projects
- in addition to data being protected in law, researchers can only use data for statistical research under a set of conditions that limit and control purpose and behaviour
- can only be accessed by researchers who have up-to-date training on how to work within a controlled, secure environment
- all outputs from the controlled environment are checked for disclosure risk before they are made available to the researcher
- disclosure risk management is mainly built into the access mechanisms and not the dataset

Because of the disclosure risk, secure microdata samples are only available to accredited researchers. The data are only available to access via the Integrated Data Service (IDS), a highly secure environment from which no data can be exported without specific approval.

Access to our secure microdata samples in the IDS is possible via the Safepod Network (SPN). SafePods are primarily based at universities and research institutions across the UK. The SafePod Network is funded by the Economic and Social Research Council and is part of the Administrative Data Research UK programme.

Access to our secure microdata samples is also possible via Assured Organisational Connectivity (AOC). This is an agreement between your organisation and the ONS to directly allow access to the IDS from your organisation or your home office space. All AOC agreements must be approved by the ONS, and a successful application will satisfy the requested evidence concerning how your organisation meets the required physical and system security standard.

Our secure microdata samples include indices of multiple deprivation (IMD) along with indices of deprivation (IoD), which were updated in 2019. There are separate indices for:

- England: English indices of deprivation
- Wales: Welsh Index of Multiple Deprivation

IMDs use a combination of administrative and census data since not all domains are covered by the census.

11 . Quality considerations for Census 2021 microdata samples

Important detail on known quality information impacting topics covered by Census 2021 in England and Wales is linked in this section. This information should inform your use of Census 2021 microdata samples, ensuring data are fit for purpose and results are interpreted correctly, and covers:

- impacts related to the coronavirus (COVID-19) pandemic
- changes in question wording affecting comparability with the 2011 Census
- changes in definitions, classifications or the derivation of specific variables impacting comparability over time

Topic quality reports for Census 2021:

- Demography and migration quality information for Census 2021
- Labour market quality information for Census 2021
- Travel to work quality information for Census 2021
- Health, disability and unpaid care quality information for Census 2021
- Ethnic group, national identity, language and religion quality information for Census 2021
- Housing quality information for Census 2021
- Education quality information for Census 2021
- UK armed forces veterans quality information for Census 2021
- Welsh language quality information for Census 2021
- Sexual orientation quality information for Census 2021

12 . Census 2021: Quality and methodology information

The census is the most complete available source of information on the population. However, despite efforts to reach everyone and obtain the most accurate information possible, no census is perfect, and some people are inevitably missed.

Further information on the conducting of Census 2021, the treatment of missing data, the quality assurance process and the census quality survey are available:

- Quality and methodology information Census 2021
- How we assured the quality of Census 2021 estimates
- Census Quality Survey agreement rates, England, and Wales: Census 2021

13 . Workplace zones

ONS Geography are considering whether to update 2011 workplace zones (WPZs) using Census 2021 and/or other data sources.

Given the impact the coronavirus (COVID-19) pandemic had on how respondents answered questions related to their workplace, consideration is being given to how and whether 2011 WPZs should be updated. The ONS Data Science Campus has produced some experimental modelled travel to work matrices, incorporating:

- 2011 Census travel to work data
- Census 2021 employment data
- National Travel Survey data
- the Department for Transport's National Trip Ends Model

It is hoped that an outcome of this work may be updated UK WPZs in the future.

We have been unable to include any variables relating to WPZs in the microdata samples. WPZs for Census 2021 were not available when our microdata samples were created. The disclosure risk resulting from changes in workplaces overtime means it was also not possible to include 2011 WPZs.

14 . Census microdata for major data linkage projects

As census microdata samples do not contain information that can identify individuals, researchers will not be able to use these for data linkage purposes.

Where accredited researchers can prove the samples do not provide sufficient utility for major data linkage projects or projects which show major public policy impact, 100% Census 2021 microdata exist within the Integrated Data Service (IDS). This dataset includes personal identifiers and is available for data linkage across other administrative and survey data. Access and usage of these data are very strictly controlled since they contain identifiable information.

Researchers can make a request for datasets to be linked by the ONS if they can show that there is a wider research benefit beyond their individual project. If a request is agreed, the dataset is linked internally and then the de-identified version would be transferred to the IDS so that all accredited researchers are able to make a project application to use it. Researchers wanting to make a request for a linked dataset should email adrcuration@ons.gov.uk.

15 . Previous uses of census microdata

Mapping 2011 Microdata using R

P. Troncoso and J. Wathan (2017) Guide to mapping 2011 Census Microdata using R (PDF, 3.49MB). UK Data Service, University of Manchester. This shows how users can create their own bespoke variables and analyse these under the assumption that census microdata samples are representative of the entire census database.

The prevalence and characteristics of children growing up with relatives in the UK

D. Wijedasa (2018) The prevalence and characteristics of children growing up with relatives in the UK (PDF, 686KB). Economic and Social Research Council, University of Bristol. Microdata from the 2011 Census were analysed to provide nationally representative, reliable statistics and maps on the distribution and characteristics of kinship care households in the four countries of the UK.

The impact of limiting long term illness on internal migration in England and Wales

S. Wilding, D. Martin and G. Moon (2016) The impact of limiting long term illness on internal migration in England and Wales: New evidence from census microdata. This project provides an example of where census microdata have been used for multi-level modelling.

SelfEmployment amongst Migrant Groups in England and Wales

K. Clark, S. Drinkwater, and C. Robinson (2015). Self-Employment amongst migrant groups in England and Wales: New evidence from census microdata (PDF, 311KB).

16 . Related links

Microdata samples
Webpage | Last updated 7 September 2023
What Census 2021 microdata samples are available and how to access them.

Census products
Webpage
Learn about the different products we will produce from Census 2021 data and when we plan to release them.

Create a custom dataset
Tool
Tool to create a custom dataset from Census 2021 data.

Census customer services
Webpage | Last updated 12 August 2022
Get help and advice about Census 2021 and all previous censuses back to 1801.

17 . Cite this methodology

Office for National Statistics (ONS), released 7 September 2023, ONS website, methodology article, User guide to microdata samples for Census 2021, England and Wales