

Coverage adjustment for Census 2021 in England and Wales

Methodology for the coverage adjustment of Census 2021 in England and Wales.

Contact:
Census customer services
Census.customerservices@ons.
gov.uk
+44 1392 444972

Release date:
19 December 2022

Next release:
To be announced

Table of contents

1. [Background and aims](#)
2. [Coverage adjustment strategy](#)
3. [Coverage adjustments methods](#)
4. [Results](#)
5. [Related links](#)
6. [Cite this methodology](#)

1 . Background and aims

Every effort is made to ensure everyone is counted in a census. However, no census is perfect, and some people are not captured. This under-enumeration does not usually occur uniformly across all geographical areas or across other sub-groups of the population, such as age and sex groups, so it must be accounted for using coverage estimation and adjustment processes.

[The coverage estimation process](#) uses a variety of statistical methods to provide estimated coverage-error corrected population totals at local authority level by key demographic characteristics.

The purpose of the census coverage adjustment is to amend the unit level census database so that it is consistent with the population estimates derived from the coverage estimation process. By adjusting the database so that it agrees with the coverage estimates, robust census population outputs can be obtained for lower-level geographical areas (such as census output areas) and these outputs will account for types of persons and households that were missed by the census. The coverage adjustment process is looking to obtain representative aggregate level population totals rather than an accurate unit level database.

To summarise, an adjusted database will provide outputs that:

- account for missed persons and households
- can be produced for small geographical areas or for more detailed characteristic breakdowns
- and will aggregate to any other outputs (for example, local authority or national level) as they are from the same database

2 . Coverage adjustment strategy

Overall strategy

The household and communal establishment (CE) populations are adjusted separately, but the overall strategy remains the same.

The adjustment process imputed persons, households, and CEs into the census database for the usual resident population. The 2021 adjustment strategy for the household and communal establishment populations consisted of two main stages. For the household adjustment, which was carried out at local authority level:

Stage 1:

1. Derive integer benchmarks from the coverage population estimates for person and household totals by key demographic characteristics that represent the missed households and persons within them.
2. Select donor households (and persons within them) from the census database to impute using the Combinatorial Optimisation (CO) method (see Section 3, "Combinatorial optimisation"), ensuring the benchmarks derived in step 1 are maintained as closely as possible.
3. Place the donor households in an appropriate small area.

Stage 2:

1. Impute remaining characteristic variables for the persons and households imputed in Stage 1. Imputation of the characteristic variables was completed using CANCEIS methods and software. [Read more about item editing and imputation for Census 2021.](#)

For CEs, the overall approach was the same but the method to select donor persons for CEs was different and is described in Section 3. Where large CEs were missed by the census, the adjustment process also added these prior to imputation of CE persons.

The Census Coverage Survey (CCS) sampled CEs according to size, so the adjustment was carried out separately for small and large CE populations. Small CEs are defined as establishments with fewer than 50 usual residents, and large CEs as establishments with 50 or more usual residents. The adjustment was carried out at Delivery Group (DG, groupings of one or more local authorities) level for small CEs and Unique Property Reference Number (UPRN) level for large CEs. More information will be released in January 2023 on the CE estimation process.

Main changes for 2021

The 2021 adjustment strategy was designed to address challenges with the 2011 coverage adjustment, and to make best use of the new strategy and outputs from the coverage estimation process. The main changes that were made were to the methods, to improve the quality and transparency of the outputs.

The 2011 coverage adjustment methodology worked well to provide an adjusted database that took account of the measured coverage. However, the implementation of the methodology was challenging and [an evaluation of the 2011 Coverage Assessment and Adjustment process in 2013](#) concluded that alternative approaches should be explored for the 2021 Census.

Household adjustment at local authority level

For 2021, the benchmarks for the adjustment process were produced at local authority level by the coverage estimation system, broken down for key characteristics associated with non-response. In 2011, the only population estimates available at local authority level were for five-year age-sex groups, other key characteristics were produced at the more aggregate Estimation Area (groupings of one or more local authorities) level.

The census database for 2021 was therefore adjusted at a lower geographical level compared with 2011, so small area outputs from the adjusted database will better account for non-response.

Whole household imputation only

The adjustment strategy for 2021 selected donor households to impute from local authority responses that represented missed households (and persons within them) to meet the benchmarks for both households and persons. A three-stage approach rather than two-stage was used in 2011, where an additional stage was included prior to imputing wholly missed households that imputed missed persons into counted households.

The justification for simplifying to just select donor households was that persons missed in counted households would be implicitly corrected through the selection, and placement, of donor households that account for both individual and household characteristic benchmarks. So, for example, if a man is missed from a household with a woman and child, instead of imputing a man into an existing two-person household we impute a three-person household and omit imputation of a two-person household to compensate.

This approach maintained the structure and relationships of counted households, and because the selection of households was of whole households, it ensured new households added by adjustment still captured non-response using households with typical structures in the local area. It also reduced the complexity of the methodology, for example it removed the need to determine the relationship of the new persons with the remainder of a counted household. It therefore reduced the risk of problems such as those experienced in 2011 and improved the processing time of the adjustment.

New household donor selection method

The Combinatorial Optimisation (CO) method, as described in a [Methodological Assurance Review Panel paper on coverage adjustment strategy for the Census 2021 \(EAP106\)](#), was used to select donor households for imputation. CO is an integer programming method which, in terms of the coverage adjustment problem, involves searching for the best combination of households that best fit the estimated benchmarks from many possible combinations of households. See Section 3, "Combinatorial optimisation" for more information on how it was used.

The CO method provided a simpler and more transparent method for selecting households and persons that best capture the estimated non-response. It performed well when benchmarking to multiple characteristic benchmarks in one step, and successfully adjusted the Census 2021 household population (see Section 4). Previously, in 2011, donor households were selected by comparing a cumulative total of calibrated coverage weights to the census return counts. The calibration did not always work well when constraining to both individual and household benchmarks simultaneously.

Usual resident population

For Census 2021, the coverage adjustment was carried out for the usual resident population only. This approach was taken so that the process could prioritise consistency with the [first Census 2021 population and household estimates release](#).

3 . Coverage adjustments methods

Household population

Benchmarks

Coverage population estimates were produced for person and household totals by key demographic characteristics associated with non-response, as the response may have varied by category. For example, more people may have been missed from one age group compared to another.

These estimates are used to calculate the shortfall for each population subgroup, which is the difference between the census return count and the estimate. The shortfall provides the target level of adjustment, also referred to as a benchmark. Table 1 provides an example of the shortfall benchmark for just one population subgroup for fictional local authorities. For this age-sex group example, a shortfall of 150 for LA 1 means that the adjustment process will target imputing 150 persons of that age group into that LA.

Table 1: Fictional example for local authority counts of men aged 20 to 24 years

| Local Authority (LA) | Census return count | Population estimate | Shortfall |
|----------------------|---------------------|---------------------|-----------|
| LA 1 | 650 | 800 | 150 |
| LA 2 | 400 | 500 | 100 |
| LA ... | ... | ... | ... |

Source: Office for National Statistics

The population estimates were available for the following characteristics, for both the household and person in household populations:

- age-sex groups (person), in five-year age groupings
- activity last week (person)
- ethnic group (person)
- household size (household)
- tenure (household)
- hard to count index (household)

For more information on the hard to count index, see the [Hard to Count index for the Census 2021 Methodological Assurance Review Panel paper \(EAP123\)](#).

For every local authority in England and Wales, this means that there were over 50 shortfall benchmark totals, including 35 different groups organised by age and sex, that the adjustment process tried to adjust for. To ensure the adjustment process could impute a selection of households and persons within them that best meet all these totals, where possible, the combinatorial optimisation (CO) method was used.

Combinatorial Optimisation (CO) method

CO is an integer programming method which involves finding the best combination (solution) from a finite set of combinations for a problem, as described in [an evaluation of the combinatorial optimisation approach to the creation of synthetic microdata by Voas and Williamson](#). In the context of the coverage adjustment, CO involves the selection of a combination of households (and persons within them) from the census database that best fits the shortfall benchmarks. The CO approach is essentially an integer re-weighting exercise where most of the households in the census database are assigned zero weights and positive integer weights are assigned to a combination of households.

As the household coverage adjustment was carried out at local authority level, the census respondents in the local authority were considered as donors for the adjustment. In other words, the local authority responses formed the donor pool for CO. Local authorities were large enough to provide a diverse donor pool of responses, and likely contained households and persons similar to those that were missed. The CO method selected households, and the persons within them, to meet the benchmarks according to the high-level steps below.

1. A random selection of N households, N being the total shortfall of households, is selected from the donor pool.
2. The selection is assessed using the overall total absolute error (OTAE), the description for this measure is provided below these steps.
3. One household is swapped out of the selection and the current selection of household is reassessed using the OTAE.
4. The household swapped in goes through an acceptance process. A simple approach would be to accept the household into the selection if it reduces the OTAE, which corresponds to the hill climbing algorithm. The adjustment process instead used simulated annealing, where swaps which adversely affect the fit might be accepted to avoid getting trapped with a sub-optimal selection of households.
5. This swapping step is repeated millions of times (iterations), or until the OTAE reaches zero.
6. When CO finishes all iterations, a final selection of donor households that the method has determined best meet the benchmarks is available for the local authority.

The most straightforward measure of assessment for CO is the TAE, which makes up the OTAE. It is calculated simply as the sum of the absolute differences between estimated and observed counts for all the benchmarks of a characteristic. The estimated is the coverage population estimate, and the observed is the count after the CO selection of households and persons have been added (imputed). For each benchmark characteristic:

$$TAE = \sum_i O_i - E_i$$

where O_i is the observed count for category i of the characteristic and

E_i is the estimated number for category i of the characteristic.

As the estimated and observed numbers converge, the TAE will tend towards zero. The OTAE is the sum of the TAE values for each characteristic, providing an overall performance measure for the CO method. An optimal solution will have a OTAE of zero, however, a higher figure is acceptable if it is the best solution available using CO and when the confidence intervals of the population estimates are considered. More information on confidence intervals can be found in section 5 of [our measures showing the quality of Census 2021 estimates](#).

TAE is easy both to calculate and to understand, but one of its disadvantages is that the magnitude of TAE depends on the total error in any given table. A relative measure was also used to assess the CO results for a local authority, where the TAE was divided by the number of households being imputed.

The adjustment process repeated the application of CO 50 times for each local authority, where variation is introduced when the first step selects a different random selection of donor households. This provided 50 solutions, in other words selections of households, to meeting the benchmarks. The best selection was determined according an automated procedure that used measures based on the Total Absolute Error (TAE) or OTAE, and the level of reuse of household donors in the selection. The top priority was minimising TAE or relative TAE for five-year age-sex groups.

Placement of new households

The donor households selected by CO needed to be placed in appropriate locations within their local authority. Ideally this would be somewhere like where they would have been missed from, and where they won't significantly disrupt the existing characteristics of a small area.

There is limited information about census missingness below local authority level, but the census collected dummy forms at addresses where there was no response. The forms collected information on the reason for completion, for example, an absent household. These forms are strong indicators of where usual resident households could have existed, and forms where the reason indicates that a household may not have usual residents (for example holiday residences) are not included in this step of the process. Other placeholders, aside from dummy forms, can be used to determine where households have been missed.

The donor households were placed in the dummy form locations, so they took on the postcode, grid reference etc. of the dummy form. They were paired using a scoring approach that considered all possible combinations of donor households and dummy forms. The score determined how likely a donor household was to be placed, several factors fed into the score around how similar the characteristics were (for example, accommodation type, of the donor household were to the dummy form) and how similar the census returns were in the small area of the dummy form.

Some donor households were not allocated to dummy form, this is usually because there not being any remaining dummy forms left that are appropriate or provide a high scoring pair. The remaining donors were allocated a postcode within their small area that contained other census household returns with similar characteristics.

Communal establishment (CE) population

Benchmarks

Owing to the sampling of CEs according to size by the Census Coverage Survey (CCS), population estimates were produced for each size type, and the adjustment carried out separately for these populations too. More information will be released in January 2023 on the CE estimation process.

For small CEs, the coverage population estimates were available at Delivery Group (DG, groupings of one or more local authorities) level by nature of establishment and five-year age-sex groups. Large CE estimates were available for individual CE properties (Unique Property Reference Number (UPRN) level) and by age-sex groups, but custom groups were determined by the nature of establishment. For example, for student accommodation one age group was aged 18 to 21 years, to better capture the population. Some large CEs were missed by the census, so person population estimates were also provided for these, and the adjustment process added in the CE record too.

Shortfall benchmarks are produced as in the household adjustment example, except at the required level based on the CE population estimates.

Record selection

Like the use of CO for the household population adjustment, a donor pool was formed at the level of the population estimates. The census respondents in each CE formed the donor pool for large CEs, and the respondents in the DG for small CEs. Persons were selected from these donor pools to meet the shortfall benchmarks, carried out separately for the small and large CE populations.

A simple approach was used for the selection of donors within a subgroup for both the small and large CE adjustment processes. For each CE (large CEs) or nature of establishment group within each DG (small CEs):

1. The shortfall was divided by the number of donor persons available for the subgroup.
2. The resulting integer number m was used to select all available donors in that subgroup m times and these were added to the selection of donor persons to impute.
3. Any remainder from the shortfall in step 1 were randomly selected from the available donors and also added to the final selection of donors to impute.

This approach roughly replicated existing distributions within the age-sex group of the level of adjustment. Where large CEs were missed by the census, the adjustment process added these prior to imputation of CE persons.

For large CEs where there were no donors (either for entirely new CE or an age group in an existing CE where there were no responses) or where donors were going to be repeated five or more times, the donor pool was expanded to other CEs of same establishment nature. This minimised high reuse of donor persons, where expanding successfully found more donors, and allowed for new CEs that the process added in to be imputed into using existing records from other similar CEs within a DG. This was an adapted approach compared to 2011 to improve meeting the benchmarks.

The selected CE donor persons were imputed into the CE they were selected from, rather than another location. This assumed that the most appropriate location for CE persons is the same type of communal establishment and local area that they were selected from. An exception is when donors were selected from an expanded donor pool, the donor persons from other CEs were placed within the large CE they were selected for.

Record completion

The result of the household and communal establishment adjustment processes was a set of records that were added to the census database. However, the household and person adjustment records were partial records, they only contained key characteristics information. Not all the information from the donor record was carried across to the imputed record.

The [post-adjustment editing and imputation process](#) followed the coverage adjustment process. It took these incomplete records and imputed the remaining characteristics and information using the Canadian Census Edit and Imputation System (CANCEIS) methods and software.

4 . Results

Performance

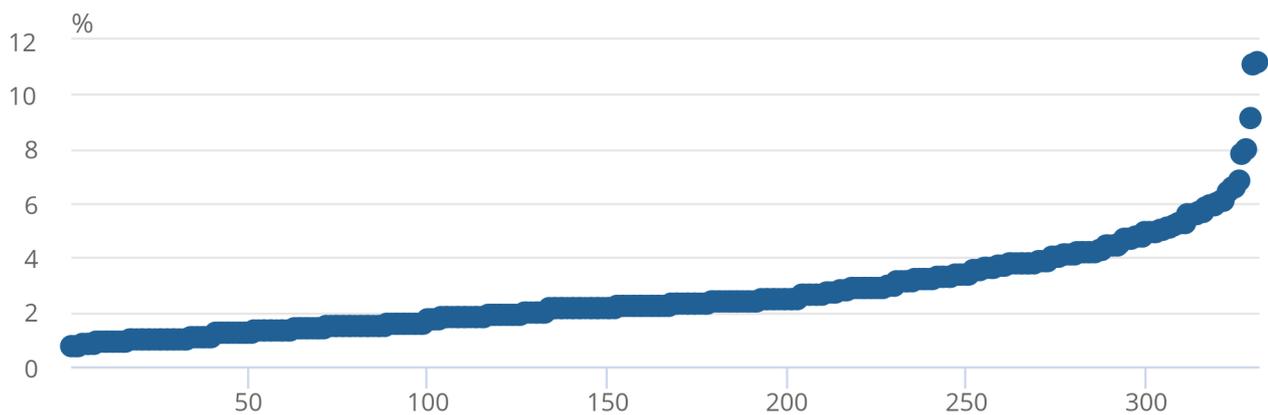
The coverage adjustment for Census 2021 was successfully completed, achieving a fully adjusted census database from which consistent outputs can be produced.

Overall, the objectives of the process were met and lessons from 2011 addressed. The adjustment imputation methodology worked well to provide a census database that was fully adjusted to take account of the measured coverage, adding wholly missed households and persons within them, wholly missed CEs, and persons within any CEs. The coverage adjustment process added in 1.8 million usual resident people to the Census 2021 household and communal establishment population.

Figures 1 and 2 present the percentages of the total census population imputed by the coverage adjustment process for local authorities in England and Wales and by five-year age-sex groups, respectively.

Figure 1: Percentage of people imputed by the coverage adjustment process into local authorities in England and Wales, local authorities ordered by imputed size rank

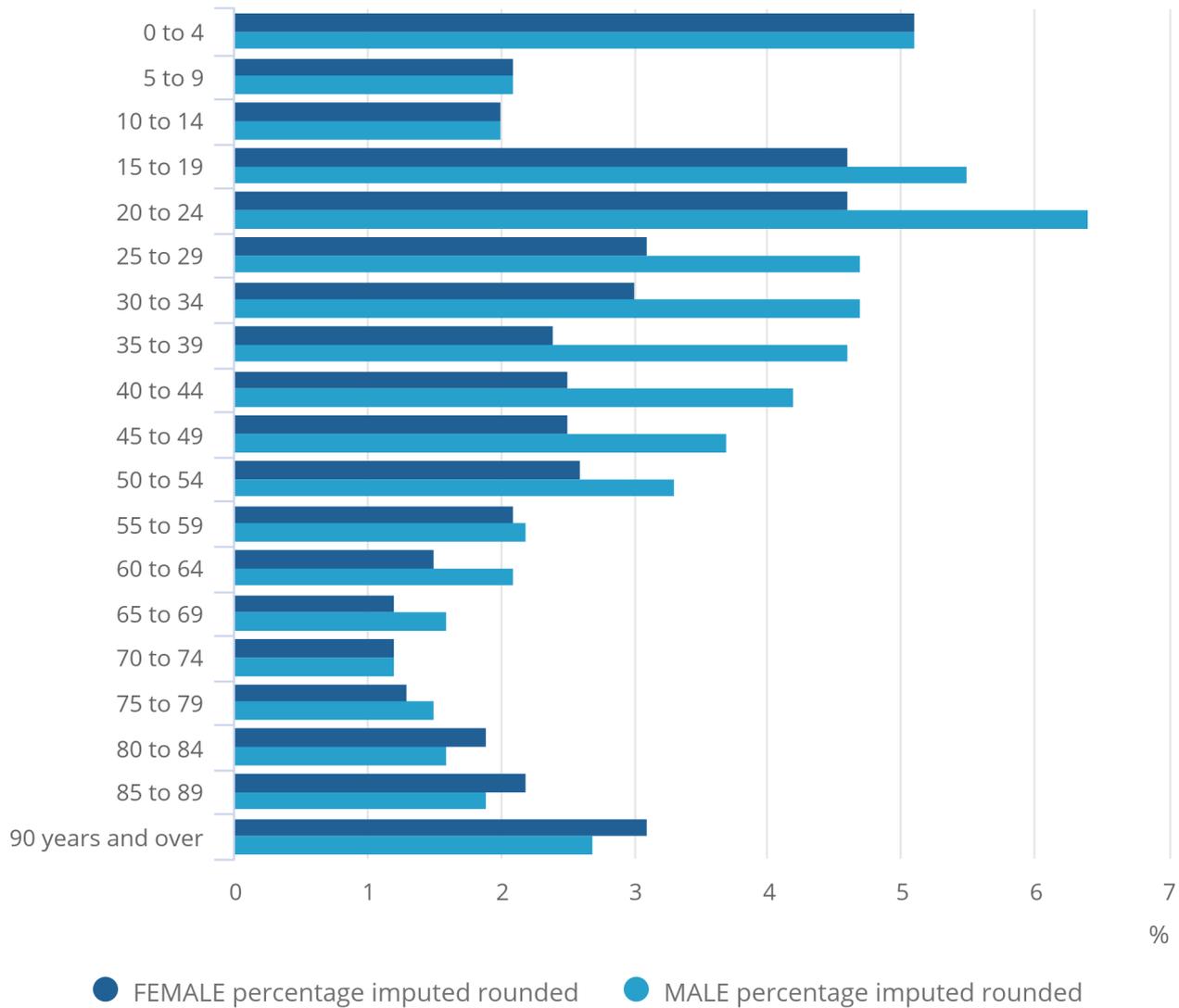
Figure 1: Percentage of people imputed by the coverage adjustment process into local authorities in England and Wales, local authorities ordered by imputed size rank



Source: Office for National Statistics

Figure 2: Percentage of people imputed by the coverage adjustment process into age groups by sex in England and Wales

Figure 2: Percentage of people imputed by the coverage adjustment process into age groups by sex in England and Wales



Source: Office for National Statistics

Quality assurance adjustments

During live processing, challenging adjustment cases were quality assured through internal working groups and alternative adjustment approaches were agreed on. Some adaptations that were applied to a few cases.

To improve coverage adjustment of CEs, the process for 2021 added whole CEs into the census population. The adaptation in approach to expand the donor pool for captured CEs where there was high reuse or no appropriate donors, described in Section 3, could then be applied to these CEs too.

For approximately 15 local authorities the best CO selection of households was manually changed to an alternative CO selection. For the majority of those cases, this was carried out to improve the quality of the distribution of single year of age for a local authority post-coverage adjustment. The adjustment process benchmarked to the available population estimates for age-sex groups (person), in five-year age groupings. Other cases where an alternative CO selection was chosen was to improve meeting the benchmarks for each characteristic such as five-year age-sex groups, which was a priority, or household size, which typically had the largest TAE. Also, to reduce the reuse of households.

5 . Related links

[An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata](#)

Article | Released November 2000

Describes the application of the Combinatorial Optimisation method to a database method and evaluation approaches.

[Assessment and Adjustment Evaluation](#)

Article | Released March 2013

An evaluation report for the 2011 Census coverage adjustment process.

[Coverage estimation for Census 2021 in England and Wales](#)

Methodology | Released 9 November 2022

Methodology for coverage estimation of Census 2021 in England and Wales.

[Hard to Count index for the 2021 Census](#)

Article | Released May 2019

A paper presented to the Methodological Assurance Review Panel on the hard to count index for Census 2021.

[Item editing and imputation process for Census 2021, England and Wales](#)

Methodology | Released 8 November 2022

The methods for resolving item non-response and item inconsistencies in Census 2021 data, including deterministic editing, nearest neighbour donor imputation, and manual imputation.

[Population and household estimates, England and Wales: Census 2021](#)

Article | Released 28 June 2022

Census 2021 rounded population and household estimates for local authorities in England and Wales, by sex and five-year age group.

[The 2021 Census Coverage Adjustment Strategy](#)

Article | Released September 2018

A paper presented to the Methodological Assurance Review Panel on the change in strategy to the coverage adjustment process for Census 2021.

6 . Cite this methodology

Office for National Statistics (ONS), released 19 December 2022, ONS website, methodology, [Coverage Adjustment for Census 2021 in England and Wales](#)