

Collecting and processing data on gender identity, England and Wales: Census 2021

Methodology for collecting and processing data on gender identity in Census 2021.

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

Release date:
19 June 2023

Next release:
To be announced

Table of contents

1. [Main points](#)
2. [Gender identity and Census 2021](#)
3. [Response to the gender identity question](#)
4. [Coding responses](#)
5. [Adjusting for non-response](#)
6. [Glossary](#)
7. [Data sources and quality](#)
8. [Related links](#)
9. [Cite this methodology](#)

1 . Main points

- We included a question on gender identity in the census for England and Wales for the first time in 2021, reflecting a newly identified user need for data on this topic.
- The gender identity question went through a rigorous development and testing process which involved trans and non-trans people and, through the 2019 Census Rehearsal, people who did not have English as their main language before the final wording was confirmed through legislation.
- Evaluation of the question during its development concluded that the final version met the requirements of public acceptability, being understood by respondents and providing the data needed by users.
- The question was voluntary, but was answered by 94% of people responding to the census.
- Coding write-in responses to this question was complex, but uncertainties and small inconsistencies in coding do not have a practical effect on the use of the data; write-in responses which did not refer to a gender identity were disregarded and treated as if the question had not been answered.
- Statistical methods used to adjust the census data for undercoverage - that is, people not included on a census return - have been applied correctly.
- An adjustment to improve the estimation of students at their term-time address may have resulted in a small number of people who would have described themselves as trans being recorded in the census statistics as not having answered the question, but this does not have a substantial impact on the usefulness of the data.

2 . Gender identity and Census 2021

Census 2021 was the first census in England and Wales to collect data on people's gender identity (see [Section 6: Glossary](#)). We published our first census results for this topic in our [Gender identity, England and Wales: Census 2021 bulletin](#).

That release was accompanied by our [Sexual orientation and gender identity quality information for Census 2021](#) to help users understand the quality of the census statistics for this topic. We are adding to this information with this report, which provides information on how the Census 2021 data on gender identity were collected and processed and what this means for the quality of the published estimates. We plan to publish further research into other aspects of quality of census data on this topic in summer 2023.

Developing the gender identity question

In 2015 we consulted on what topics to include in Census 2021. The [Public consultation on proposed topics for the 2021 Census questionnaire in England and Wales](#) identified a clear need among data users for data about gender identity. This was particularly to help in planning services for, and allocating resources to interventions to support, the trans community in England and Wales.

As a result of that consultation, we set out our [Gender identity testing and research plan \(PDF, 799KB\)](#) to investigate how best to meet the need for these data. The results of that research fed into the 2018 white paper [Help Shape our Future \(PDF, 967KB\)](#), setting out the proposed approach to Census 2021 in England and Wales. The paper recommended that the census should include a voluntary question on gender identity for people aged 16 years and over.

While the white paper recommended the inclusion of a question on this topic, it did not propose a particular wording for the question. As with all census questions, the gender identity question went through a detailed process of development and testing. This evaluated three core designs as described in detail in our [Sex and gender identity question development for Census 2021 report](#). Table 3 in Annex 2 of that report lists the testing activities conducted for this topic. This included:

- qualitative research involving both trans participants and those whose gender identity is the same as their sex registered at birth
- quantitative research through five online and multi-modal surveys with a range of respondents
- inclusion in the 2019 Census Rehearsal, which covered all households in four local authority areas (Carlisle, Ceredigion, Hackney and Tower Hamlets) and which obtained responses from more than 100,000 households

As part of our assessment of the questions for inclusion in the census, potential questions were evaluated for:

- potential (negative) impact on data quality
- public acceptability (avoiding asking intrusive questions that might affect the census response rate)
- respondent burden
- risk of "mode effect" (where people's responses are affected by whether they respond using the online questionnaire or the paper version)

This was done using criteria published in our [Question and questionnaire development overview for Census 2021](#). These evaluations concluded that a question on gender identity should be included in Census 2021 and the development and testing we have described reduced the potential impact that was initially identified. The points where a potential for impact remained are described in detail in Annex 3 of our [Sex and gender identity question development for Census 2021 report](#) and are summarised as follows:

- this question had not been asked on a census before (which increases the risk of unexpected issues in data collection or processing)
- the question included the option to write in a response (identified during stakeholder engagement as needed for trans people to be able to answer the question accurately), which increases the complexity of responding to the question and processing the data
- this question asks information that cannot be observed and so can be difficult to answer on behalf of another person
- the responses to this question could be at risk of social desirability bias, when a response is given that is considered socially desirable rather than accurate
- the question asks for sensitive information that a respondent may not want answered by a proxy on their behalf; this was mitigated by respondents being able to request an individual access code or paper form if they wanted to respond separately to the rest of their household

While our quantitative research and the Census Rehearsal were not designed to estimate the size of the trans population, the testing did not provide any evidence of people systematically misunderstanding the tested questions by incorrectly recording themselves as trans.

The final English language version of the question was agreed by the UK and Welsh Parliaments through [The Census \(England\) Regulations 2020](#) and [The Census \(Wales\) Regulations 2020](#). It appeared in the English Household questionnaire as:

Is the gender you identify with the same as your sex registered at birth? This question is voluntary.

- Yes
- No, enter gender identity

Section 7 of our [Sex and gender identity question development for Census 2021 report](#) provides the Welsh language version of the question together with an illustration of how the questions were presented on the paper and on the online questionnaires.

On each version of the questionnaire a note between the question and the response options stated that the question was voluntary. The routing of the question was designed so that it was only asked of those aged 16 years and over.

On the online version, no radio button was selected as the default. Entering text into the write-in box forced the radio button to show "no" selected.

The census questionnaires for completion (both paper and online) themselves were only provided in English and, in Wales, English and Welsh. However, respondents could access copies of translations of the questionnaire in one of 50 different languages as listed in the archived version of the [Language support section of the Census 2021 website](#). These translation booklets were also made available through our community engagement staff, who worked with communities across England and Wales.

As with all census questions, guidance was made available to respondents to help them answer the question. This guidance is provided in the archived version of the [Census 2021 website](#). While the [guidance for the question on sex was changed in March 2021](#) as a result of a judicial review, the guidance for the gender identity question remained unchanged.

While translations of the guidance for each individual question were not published, respondents could get help in understanding the questions by calling our freephone Census 2021 contact centre. The centre provided translators to support those for whom English was not their first language.

The Office for National Statistics (ONS) is responsible for running the census in England and Wales, while separate censuses for the other parts of the UK are run by National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA). While these three organisations work closely together to try to harmonise census questions where possible, we recognise that each country has its own user and respondent needs. A result of that is that while a question on gender identity was also asked on the 2022 census in Scotland, that question used a different wording. The 2021 census in Northern Ireland did not ask a question on gender identity.

3 . Response to the gender identity question

Of people aged 16 years and over who responded to Census 2021, 94% chose to answer the voluntary gender identity question. Response was high across a range of demographic groups (Table 1), with somewhat lower rates for respondents who did not have English, or English or Welsh in Wales, as their first language. "Missing" responses for this question were not imputed, as explained in [Section 5: Adjusting for non-response](#). This means higher question non-response rates will result in higher numbers in the "Not answered" category in published outputs and lower numbers in the trans categories and in the category of those with a gender identity the same as their sex registered at birth.

A small majority of respondents who ticked "No" also wrote in a response, though this proportion is substantially lower for respondents who could not speak English well or at all. This does not affect the total estimate of the trans population but will affect which of the trans categories the response is counted in as explained in [Section 4: Coding responses](#).

Table 1: Responses to gender identity questions, Census 2021
All people aged 16 years and over, England and Wales

	Percentage of census respondents who answered gender identity question	Percentage of census respondents answering 'no' to the gender identity question who wrote in a gender identity
All people	94	56
Sex		
Female	94	60
Male	94	52
Age		
16 to 25 years	93	74
26 to 64 years	95	50
65 years and over	93	50
Ethnic group		
White	95	62
Mixed or Multiple	93	63
Asian, Asian British, Asian Welsh	91	42
Black, Black British, Black Welsh, Caribbean or African	92	51
Other ethnic group	90	36
How well respondent speaks English		
Has English (English or Welsh in Wales) as main language	95	63
Can speak English very well	92	50
Can speak English well	88	34
Cannot speak English well	88	30
Cannot speak English	87	29

Source: Census 2021 from the Office for National Statistics

Notes

1. Figures in this table will differ to standard published results as they relate to responses rather than final estimates. Responses to the gender identity question coded as "missing" are not counted as responses to that question in this table.

4 . Coding responses

The collected response data for the tick box and write-in parts of this question went through coding processes to assign each response to the categories used in published outputs. There were several steps to achieving this.

As the paper questionnaire could not reflect the automated routing used for online responses, we applied processing rules to resolve anomalies in these paper responses. These were of three types.

First, it was possible for someone aged under 16 years responding on a paper questionnaire to have provided an answer to the gender identity question. Since the question was only intended for people aged 16 years and over, these responses were set to the correct code of "Not required" unless there were other inconsistencies suggesting that the stated age was wrong. If that was the case, the anomaly was resolved through edit and imputation processes which could result in gender identity for that particular record being set to "Not required" or the gender identity response being kept and an age of over 15 years imputed.

Second, it was possible (again, on paper questionnaires only) for someone to tick "Yes" (that is, that their gender identity was the same as sex registered at birth) but then to write in a response in the box under the "No" option. In these instances, we followed the standard rule across census coding that a write-in answer is given precedence over a tick box. This means that the conflict was resolved by changing the tick box selection from "Yes" to "No". There were approximately 1,300 responses where this rule was applied.

Third, it was possible (again, on paper questionnaires only) for someone to tick both boxes. In these instances, the response was initially set to "missing" though, again, this might change if a write-in response had also been provided.

While one set of processes resolved these anomalies, another process coded the write-in responses (from both paper and online responses). The method of coding is described in our [Automated text coding: Census 2021 methodology](#). In summary, the write-in responses were "cleaned" to remove extraneous characters, then compared (using exact and fuzzy matching methods) against an index of anticipated potential responses. This index was updated during census processing to provide rules for additional types of response observed among received census forms.

Coding of the write-in responses to this question was complicated by the "free" nature of the response, which meant that there were a very large number of unique write-ins. This is in contrast to the write-in for country of birth, for example, where the great majority of written-in responses would be known to relate to countries listed on a standard index. To add to this, as the question had not been asked previously on the census, we did not have such a large dataset of previous answers to help in the development of the coding index. We will use the information from the census in the development of coding indexes for future collection of data on this topic.

Sometimes, write-in text responses could arguably belong in multiple categories, depending on how they were interpreted. In these cases, a judgment call was made when processing the data. Usually, the decision was between two categories within the wider group of trans responses so the decision will not have affected the overall count of transgender people, but only which subgroup they were coded to.

Some write-ins, such as "not sure" or "I don't know", were ambiguous between expressing a nonbinary, fluid, or "questioning" gender identity, and expressing confusion at the question itself. The most common terms along these lines sum to around 1,000 responses. These were initially classified as "uncodeable" and then coded to "missing".

The coding of write-in answers did not make use of information from the sex question. It was recognised that some trans people might record the same response for both sex and gender identity. Write-in responses of "man" or "woman" were therefore coded to "trans man" or "trans woman".

When this coding process was finally complete, every adult responding to the census was coded to one of the following groups.

Gender identity the same as sex registered at birth:

93.6% of all responses. The overwhelming majority of this category are responses where the respondent ticked "yes", but it also contains some records (less than 0.01% of the responses in this category) where "no" was ticked and write-in response such as "yes" was provided.

Gender identity different from sex registered at birth (not otherwise specified):

0.23% of all responses. The overwhelming majority of this category are responses where the respondent ticked "no" and did not provide a write-in answer, but it also contains a small number of responses (less than 1.0% of the responses in this category) where a write-in response such as "no" was provided.

Trans gender identity, woman:

0.10% of responses. This category includes write-in responses of variants of female, woman, trans woman and similar terms. The single term "female" accounts for 84% of responses in this category.

Trans gender identity, man:

0.10% of responses. This category includes write-in responses of variants of male, man, trans man and similar terms. The single term "male" accounts for 85% of responses in this category.

Non-binary gender identity (not otherwise specified):

0.06% of all responses. This category contains write-in variants of "non-binary".

All other identities:

0.04% of all responses. This category includes write-ins such as "transgender", where no further detail was provided, "genderqueer", "genderfluid" and "agender".

Not answered:

5.9% of all responses. The great majority of this category consists of responses where the respondent has not ticked either box nor provided a write-in answer. Less than 1.0% of the responses in the category contained a write-in that was treated as equivalent to leaving the question blank. These included responses of "n/a", "human", "not known" and facetious answers. We examined the responses in this category to look for write-ins appearing more than once that should clearly have belonged to substantive codes, either trans or not trans. The most common of these were compound phrases like "trans non-binary" or "gender non-binary". We detected over 2,000 such responses, indicating a slight undercount of trans respondents from this issue.

Note that figures in this section may differ to standard published results as they relate to responses rather than final estimates.

In summary, of all responses indicating a trans identity, 87% either contained no write-in or one of the terms "male", "female", or "non-binary". Coding accuracy is very high within the trans group. We have identified some responses coded as "Not answered" which should have been coded to a trans group, which would have resulted in a very slightly higher final estimate of the trans population. We have no evidence of errors or inconsistencies in the application of coding which would affect the likely use of published statistics.

5 . Adjusting for non-response

The census results cover the entire usually-resident population of England and Wales. We use statistical methods to estimate and adjust for non-response to the census. There are three aspects of this to highlight for the gender identity data.

Question non-response

Most questions on the census are mandatory - there is a legal requirement to answer them. Where a census response is received but a mandatory question has been missed, we use a statistical method to impute a response as described in our [Item editing and imputation process for Census 2021, England and Wales methodology](#).

Gender identity was not a mandatory question and respondents could legitimately decide to leave the question unanswered. These responses were coded as "Not answered" and appear as such in the published tables.

Person non-response (undercoverage)

While Census 2021 achieved a high estimated response rate of 97%, it is inevitable that some people will not be included in a census response. We use a combination of survey data, statistical models, and administrative data to estimate the number and main characteristics of people not covered on a response. We then add records to the census database to reflect the estimated missing people and complete the remaining characteristics for each person by "imputing" values seen on similar records.

All published census estimates include this adjustment for undercoverage. This provides a more reliable picture of the whole population than only including people covered on a census response but means that census estimates will contain an element of uncertainty through sampling error.

Gender identity was not one of the "main characteristics" used when estimating non-response, but it is correlated with main characteristics such as age and geographical area. We would therefore expect the distributions of final estimates for gender identity to be slightly different from the unadjusted response data.

In particular, as the trans population is relatively concentrated in those aged 18 to 25 years and in urban areas, both of which are associated with higher undercoverage rates than average, we would expect the adjustment for undercoverage to result in a higher proportional increase in the trans population than the population as a whole. Table 2 shows this. We have found no evidence that the methods used to adjust for undercoverage have not worked as designed in the final estimates of the trans population.

Table 2: Adjustment for undercoverage, Census 2021
All people aged 16 years and over, England and Wales

	Percentage adjustment to total population	Percentage adjustment to trans population
All people	3	6
Sex		
Female	3	5
Male	4	8
Age		
16 to 25 years	6	6
26 to 64 years	3	7
65 years and over	2	2
Ethnic group		
White	2	5
Mixed or Multiple	5	7
Asian, Asian British, Asian Welsh	5	8
Black, Black British, Black Welsh, Caribbean or African	8	10
Other ethnic group	6	10
How well respondent speaks English (English or Welsh in Wales)		
Has English as main language	3	5
Can speak English very well	6	9
Can speak English well	7	9
Cannot speak English well	7	10
Cannot speak English	7	11

Source: Census 2021 from the Office for National Statistics

Notes

1. This table shows the undercoverage adjustment expressed as a percentage of the final population estimate for that group.

Adjustment for students recorded only at their home address

The standard population definition used in the census counts students as usually resident at their term-time address. Students were asked to complete the full census questionnaire at their term-time address but only provide some basic information, including their term-time address, if they were separately included on a census response from a non-term-time address such as their parents'.

If a student provided that basic information from a non-term-time address but did not respond from the term-time address they stated on the questionnaire, we copied their data from the out-of-term-time to the term-time address. This meant we could correctly record them as part of the population for that area. This process was applied to some 96,000 records.

This improved the reliability of the population estimates for areas with high numbers of students. However, since only basic demographic information was collected about these students in the questionnaire completed for their out-of-term-time address, the remaining questions were completed using the standard methods for question non-response. Furthermore, since these methods did not impute responses for the voluntary gender identity question, this meant that students recorded at a non-term-time address but not at their term-time address will be recorded as having no response to the gender identity question.

The impact of this on the final estimates of gender identity cannot be calculated exactly. However, if it is assumed that the responses to the gender identity question from this subset of students were to follow the same distribution as those of full-time students responding from their term-time address, then the estimated number of people with a gender identity different to sex registered at birth would increase by around 1,000 people.

6 . Glossary

Gender identity

Gender identity refers to a person's sense of their own gender, whether male, female or another category such as non-binary. This may or may not be the same as their sex registered at birth.

Trans

The term "trans" is used in this methodology to describe anyone who stated in the census that their gender identity was different to their sex registered at birth. This includes people who identify as a trans man, trans woman, non-binary or with another minority gender identity.

Non-binary

Someone who is non-binary does not identify with the binary categories of man and woman. In these results the category includes people who identified with the specific term "non-binary" or variants thereon. However, those who used other terms to describe an identity which was neither specifically man nor woman have been classed in "All other gender identities".

Trans man

A trans man is someone who was registered female at birth, but now identifies as a man.

Trans woman

A trans woman is someone who was registered male at birth, but now identifies as a woman.

7 . Data sources and quality

Quality considerations along with the strengths and limitations of Census 2021 more generally can be found in the [Quality and Methodology Information \(QMI\) for Census 2021](#). Read more about the [Sexual orientation and gender identity quality information for Census 2021](#).

Further information on our quality assurance processes is provided in our [Maximising the quality of Census 2021 population estimates methodology](#).

8 . Related links

[Gender identity, England and Wales: Census 2021](#)

Bulletin | Released 6 January 2023

The gender identity of usual residents aged 16 years and over in England and Wales, Census 2021 data.

[Gender identity: age and sex, England and Wales: Census 2021](#)

Bulletin | Released 25 January 2023

Gender identity by age and sex of usual residents aged 16 years and over in England and Wales, Census 2021 data.

[Automated text coding: Census 2021](#)

Methodology | Last revised 3 April 2023

Automated coding methods for write-in text responses from the Census 2021 online questionnaire and their results.

[Sex and gender identity question development for Census 2021](#)

Article | Released 26 June 2020

How we researched, developed and tested the Census 2021 questions on sex and gender identity.

[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.

9 . Cite this methodology

Office for National Statistics (ONS), released 19 June 2023, ONS website, methodology, [Collecting and processing data on gender identity, England and Wales: Census 2021](#)