Longitudinal Study 2011 to 2021: Completeness of census linkage

The quality of the linkage of Census 2021 to the Office for National Statistics Longitudinal Study, including tracing rates, sampling fractions, and forwards and backwards linkage rates.

Contact:
Longitudinal Study Development team
pop.info@ons.gov.uk
+44 1329 444661

Release date:
7 July 2026

Next release:
To be announced

Table of contents

1 . Overview

From Census 2021 of England and Wales, there were 650,317 retained records with one of the four Longitudinal Study (LS) sample birth dates. Of these records, 647,804 (99.6%) were traced to an NHS patient record via the Personal Demographics Service (PDS). This compares with a tracing rate of 98.8% in the 2011 Census.

We can attribute this high tracing rate in 2021 in part to more complete and higher quality NHS registrations since the coronavirus (COVID-19) pandemic and associated vaccine deployment. Additionally, increased online census completion (89% of households) reduced the scope for data entry and scanning error.

Sampling fractions for both male and female LS members in Census 2021 are 1.11%, which is an improvement on the last two censuses. This is close to the target of 1.1%, showing that we have approximately the expected number of records for LS members, given the number of people who responded to the census.

There were 526,289 LS members traced at the 2011 Census who were not known to have died or emigrated from England and Wales before the census day and who were therefore eligible to be in Census 2021. Of these, 473,815 (90.0%) were enumerated at Census 2021. Forward linkage rates are slightly lower for those born since 2011 (88.3%) and are significantly lower for intercensal immigrants since 2011 (45.2%). Overall forward linkage rates and linkage rates for intercensal births and immigrants for 2021 are higher than for 2011 and 2001.

Linkage failure occurs when LS members' emigration is not recorded in NHS systems. Census non-response and errors in date of birth or other corroborating information in the census data also contribute to linkage failure. The characteristics associated with higher rates of linkage failure are:

- being aged 15 to 29 years or over 80 years
- never being married or civil partnered
- being non-UK born
- being in the "Other White" ethnic group
- living in a communal establishment
- being a student or unemployed
- living in London, especially Inner London

There are now six census points available for analysis in the LS. Figure 1 shows the sample sizes for traced LS members across censuses from 1971 to 2021.

Figure 1: There are 1.3 million traced Longitudinal Study members since 1971

Sample sizes for traced Longitudinal Study (LS) members, 1971 to 2021

Downloac

2 . Tracing rates

Tracing rates show the proportion of the Census 2021 sample who reported one of the four Longitudinal Study (LS) sample dates of birth and were successfully traced to their health record on the NHS Personal Demographics Service (PDS). These people would either have been an existing LS member or were identified as a new LS member at Census 2021.

The tracing rate is expressed as:

$$\frac{\text{Number traced in the sample or subgroup}}{\text{Number of LS members in the sample or subgroup}} * 100$$

The untraced rate is expressed as:

$$\frac{\text{Number untraced in the sample or subgroup}}{\text{Number of LS members in the sample or subgroup}} * 100$$

Tracing rates were calculated for all individuals enumerated in Census 2021 that were:

- a usual resident
- a short-term resident (in Census 2021)
- a student only enumerated at their family address
- a visitor (not included in demographic breakdowns)

Tracing rates are calculated based on the number of LS members traced at the time of Census 2021.

Table 1 shows the LS tracing rates, broken down by resident type in 2021. Usual residents have the highest tracing rate, at 99.7%, compared with a 99.0% tracing rate achieved in 2011. The tracing rate for short-term residents improved from 50.4% in 2011 to 88.3% in 2021.

Table 1: The overall tracing rate is 99.6% for Longitudinal Study members enumerated at Census 2021
Longitudinal Study (LS) tracing rates by resident type in 2021

Resident type	Number traced	Number untraced	Total	Tracing rate (%)
All LS members enumerated at Census 2021	647,804	2,513	650,317	99.6
of which:				
usual residents	642,837	1,898	644,735	99.7
short-term residents	1,382	182	1,564	88.3
student only enumerated at their term time address	2,029	38	2,067	98.0
visitors	1,556	395	1,951	79.8

Source: Longitudinal Study from the Office for National Statistics

Table 2 shows that tracing rates have been above 98% since 1981. Tracing rates in 2021 are the highest to date. There is a small difference between rates for males and females of 0.2 percentage points, compared with 0.6 percentage points in 2011.

Table 2: Tracing rates have been high across all censuses since 1981
Longitudinal Study (LS) tracing rates by sex, censuses between 1971 and 2021 (%)

Census	Males	Females	All
1971	97.2	96.5	96.8
1981	98.8	98.9	98.8
1991	98.3	98.5	98.4
2001	99.1	99.5	99.3
2011	98.5	99.1	98.8
2021	99.6	99.8	99.7

Source: Longitudinal Study from the Office for National Statistics

Notes

1. Data for 2021 do not include visitor-only enumerations. The overall tracing rate is 99.6% for 2021 including visitor enumerations.

2. Visitor enumerations were included in this breakdown in 2011.

We can partially attribute higher tracing rates to more complete and higher quality GP registration data, as people updated their details in the context of the coronavirus (COVID-19) pandemic and associated vaccine rollout. A high online response rate (89% of households) also contributed to higher tracing rates. While online response can be prone to some types of respondent error, such as transposing characters or numbers, the quality of online data is higher than data derived from scanning handwritten forms.

Age and sex

Tracing rates by age and sex are all higher for Census 2021 than in 2011. The lowest tracing rate is for males aged 25 to 29 years (99.0%) in 2021, which is a notable improvement from 94.7% for this group in 2011. The lowest tracing rate for females in 2021 is for those aged 20 to 24 years (99.3%), which is an improvement from 96.4% for this group in 2011.

The highest tracing rates for both males and females are for those aged 75 years and over, at over 99.9%. This age group also had the highest tracing rate (99.8%) in 2011.

Marital and civil partnership status

Tracing rates do not vary much by legal partnership status and are higher in 2021 for all marital and civil partnership groups than in 2011. Tracing rates are highest for those who are widowed or are a surviving civil partner (99.9%), compared with 99.6% in 2011. The lowest tracing rate by legal partnership status is for those who have never been married or in a civil partnership (99.4%), compared with 97.4% in 2011.

Economic position

The tracing rates by economic position for 2021 exceed the rates achieved for all groups in 2011. Tracing rates are highest for those who were retired (99.9%), compared with 2011 (99.8%). Excluding the "Other" category, students have the highest untraced rates in 2021, at 1.0% for males and 0.7% for females. This is a large improvement on the untraced rate for male and female students in 2011 (5.2% and 3.0%, respectively).

National Statistics Socio-economic Classification

Tracing rates are above 99% for all National Statistics Socio-economic Classification (NS-SEC) groups in 2021. The highest tracing rates are for those in "Higher professional and higher managerial and administrative occupations", at 99.8%.

The highest untraced rate for males in 2021 is for those who had "never worked" (0.8%), compared with 2.7% in 2011. For female LS members, the highest untraced rate is for students, at 0.6%, compared with 3.1% in 2011.

Region

The highest untraced rates are for those living in London (0.7% for males and 0.5% for females in 2021). This is a major improvement to the untraced rates of 4.4% for males and 2.9% for females in 2011.

Wales had the lowest untraced rate (0.5%) for all members in 2011. All regions except London have a lower untraced rate in 2021 than Wales in 2011. The South West has the lowest overall, at 0.2%, for all members.

Country of birth

People born in the US have the lowest tracing rate, of 92.8%, in 2021. This can be partly explained by the presence of US Air Force personnel and support staff living in the UK. These individuals are unlikely to register with the NHS and would therefore not have been traced on the PDS. The tracing rate becomes 98.4% if these 120 military records (out of 151 total untraced records) are removed from the calculation, which is in line with other non-UK countries of birth.

There were substantial improvements to the tracing rates of EU pre-accession states, EU accession states, and Eastern Asia. Rates increased by approximately 8.5, 15.4 and 10.6 percentage points, respectively, between 2011 and 2021.

Religion

For all religions, the tracing rate is over 99% for both males and females. The highest tracing rate is for people identifying as "Christian" (99.7%) and the lowest is for those identifying as "Buddhist" (99.1%).

Ethnic group

The "White British" ethnic group has the highest tracing rate (99.9%) for 2021, which is consistent with the tracing rate of 99.8% in 2011.

The lowest tracing rate for 2021 is for the "Gypsy/Irish Traveller/Roma" group, at 97.3%, which is an improvement on the 2011 rate (for "Gypsy or Irish Traveller") of 95.5%.

The ethnic group with the next lowest tracing rate is "Chinese" for both males (97.6%) and females (98.6%). A similar pattern was found in 2011, when the corresponding tracing rates were 89.1% for males and 91.5% for females.

Year of arrival

For a census record to be traced, the individual must have a record on the PDS. The longer someone has been in the country, the more likely they are to have engaged with NHS services. Consequently, the more recently non-UK-born people had arrived in the UK, the lower their tracing rate.

Tracing rates for 2021 are:

- 97.7% for those who arrived in 2018
- 97.1% for those who arrived in 2019
- 95.3% for those who arrived in 2020
- 92.5% for those who arrived in 2021

These rates are an improvement from equivalent tracing rates in 2011, which were:

- 88.9% for those who arrived in 2008
- 83.7% for those who arrived in 2009
- 67.8% for those who arrived in 2010
- 42.9% for those who arrived in 2011

Communal establishments

Overall, the 98.6% tracing rate for LS members enumerated in 2021 in communal establishments (CE) is lower than the tracing rate for those in private households (99.7%). This is an improvement from 2011, when the tracing rate was 94.0% for CE and 98.9% for private households.

Those living in defence CE have the lowest tracing rate of 88.1% for 2021 and 84.5% in 2011. Residents of defence CE are likely to be serving armed forces personnel who have their general medical needs met by the Defence Medical Services. They are less likely to be present on the PDS and are therefore less likely to be traced.

3 . Sampling fractions

The Longitudinal Study (LS) was designed as a 1.10% sample of the population of England and Wales by selecting individuals who are born on one of four selected dates in a calendar year. When a new LS sample is taken at a census, we measure how representative the sample is, to determine if any biases exist within it.

Sampling fractions measure whether the LS population (both as a whole and in subgroups) represents approximately 1.10% of the census population (or specific subgroups of that population). Variations in tracing rates and in the availability of accurate birth date data mean that the sample proportions may vary.

The observed net sampling fraction $_{i}$ (where "i" identifies the subgroup of interest) is defined as:

$$\alpha_i = \frac{\text{Number in subgroup i (traced LS population)}}{\text{Number in subgroup i in census population}} * 100$$

The expected net sampling fraction gives an estimate of the proportions of the LS sample that would have been expected in the census population, based on overall sampling fractions and the relative levels of tracing in a population subgroup. Ideally, observed and expected net sampling fractions should be the same. However, biases in tracing rates and in the sampling of certain population subgroups can distort that relationship.

Expected net sampling fractions are defined as:

$$\alpha_i(expected) = \frac{\alpha(1-\beta_{\text{i}})}{1-\beta}$$

Where “i” indicates the subgroup of interest, “” is the overall sampling fraction (the sampling fraction for either all LS males, for all females, or for the total LS population), and ““ is the overall untraced rate (the untraced rate for either all LS males, for all LS females, or for the total LS population).

Sampling fractions are calculated for usual residents only.

Summaries in the following subsections provide observed sampling fractions based on census counts. Sampling fractions based on census estimates, which include imputation and adjustment, are available in our accompanying datasets.

Table 3 shows the observed sampling fractions for each census since 1971. Sampling fractions are 1.11% for both male and female LS members in Census 2021. Overall, we have approximately the expected number of records for LS members given the number of people who responded to the census.

Table 3: Sampling fractions of 1.11% in 2021 are close to expected levels
Observed sampling fractions by sex, censuses between 1971 and 2021 (%)

Census	Males	Females	Overall
1971	1.06	1.05	1.05
1981	1.09	1.09	1.09
1991	1.07	1.07	1.07
2001	1.09	1.10	1.09
2011	1.09	1.10	1.09
2021	1.11	1.11	1.11

Source: Longitudinal Study from the Office for National Statistics

Age and sex

There is some variability in the observed sampling rate for LS members by age. They range from 1.16% for those aged 45 to 49 years to 1.06% for those aged 20 to 24 years. Young adults, especially males, are often more difficult to trace on NHS records because they do not always interact with services where they are usually resident. The greatest sex difference is for those aged 30 to 34 years (1.14% for females and 1.11% for males).

Marital and civil partnership status

The highest observed sampling rate is for those who were "Married, Civil Partnered or Separated" (1.14%). The expected sampling rate of 1.12% implies oversampling for this group. The sampling fraction is 1.08% for LS members who had never married or registered a civil partnership, suggesting undersampling relative to the expected 1.11%.

National Statistics Socio-economic Classification

The highest observed sampling fraction by National Statistics Socio-economic Classification (NS-SEC) is 1.21% for the "never worked" group, compared with an expected rate of 1.11%. The sampling fraction is particularly high for females (1.25%) in this group, compared with males (1.14%). The sampling rate is low for both males and females for the "Lower Managerial and Professional" group, at 0.97%.

Country of birth

The observed sampling fraction for UK-born LS members is 1.09%, compared with 1.08% in 2011. Sampling fractions for LS members born across the UK are:

- 1.09% for England
- 1.08% for Wales
- 1.07% for Northern Ireland
- 1.05% for Scotland

Sampling fractions by country of birth for LS members born outside of the UK vary considerably. For example, the sampling fraction is 1.96% for those born in Bangladesh, compared with the expected value of 1.12%. This indicates significant oversampling, which is consistent with results from 2011 and 2001. Observed sampling fractions suggest that people born in Europe, North America and Central America are generally undersampled, while people born in Africa, the Middle East and Asia are generally oversampled.

Ethnic group

Sampling fractions by ethnic group are generally around 1.10% to 1.20%. However, the sampling rate is 1.30% for the Asian/Asian British ethnic group (consisting of Indian, Pakistani, Bangladeshi, and Chinese ethnic groups). This indicates oversampling in these groups, which is in line with previous censuses.

The lowest sampling rate is 1.06% for the Gypsy or Irish Traveller ethnic group. There is a clear difference between the sampling rate for male and female Gypsy/Irish Travellers, at 0.96% and 1.15%, respectively. The sample size and number traced for the Gypsy/Irish Traveller ethnic group are small when compared with other ethnic groups.

Religion

There are higher sampling fractions for those describing themselves as Hindu (1.27%), Muslim (1.32%), or Sikh (1.54%). This mirrors the oversampling of the Asian/British Asian ethnic group.

Country and region

England has a higher sampling fraction (1.11%) than Wales (1.09%). London has the highest overall sampling fraction (1.16%), which is slightly higher than 2011 (1.12%). Outside of London, the West Midlands has the highest sampling fraction (1.13%), while the North East has the lowest (1.09%). The higher sampling fractions for London and the West Midlands may reflect their ethnically diverse populations and oversampling of some ethnic groups.

Short-term residents

The sampling fractions for short-term residents are generally lower than 1.10%.

When broken down by age and sex, the lowest figure is 0.79% for males under 14 years of age.

4 . Forward linkage rates

Forward linkage rates for Longitudinal Study members traced in 2011 Census

Forward linkage rates measure the percentage of Longitudinal Study (LS) members who were present at a census and also were present at the following census, excluding those who died or "embarked" (when an LS member notifies their GP surgery that they are emigrating from England and Wales) in the intercensal period.

Forward linkage rates are expressed as:

$$\frac{\mathrm{N} - \mathrm{S_i} - \mathrm{r_i}}{\mathrm{N} - \mathrm{S_i}} * 100$$

Where "i" indicates the subgroup of interest, “N” is the number of traced LS members found at the first census, "S_i" is the number of traced LS members who have died or embarked before the second census, and “r_i” is the number of surviving traced LS members not found at the second census.

Linkage failure occurs because of unreported embarkations, census non-response, and discrepancies in date of birth or other corroborating information at either census.

Linkage rates are calculated for all individuals enumerated and traced in the census. This includes those enumerated as:

- a usual resident
- a short-term resident (in Census 2021)
- a student only enumerated at their family address
- a visitor only (not included in demographic breakdowns)
- unretained by census (not included in demographic breakdowns)

Table 4 summarises the 10-year forward linkage rates from 1971 to Census 2021.

Table 4: Approximately 90% of Longitudinal Study members traced at the 2011 Census were linked to their census record in 2021
Forward linkage rates for traced Longitudinal Study (LS) members, by decade, 1971 to 2021

Censuses	Traced at prior census	Died or embarked	Eligible	Not found at following census	Enumerated at following census	Linkage rate (%)
1971 to 1981	512,970	66,098	446,872	39,009	407,863	91.3
1981 to 1991	527,956	61,341	466,615	46,143	420,472	90.1
1991 to 2001	535,015	59,637	475,378	57,174	418,204	88.0
2001 to 2011	536,565	55,533	481,032	59,237	421,795	87.7
2011 to 2021	585,068	58,779	526,289	52,474	473,815	90.0

Source: Longitudinal Study from the Office for National Statistics

There were 585,068 LS members who were enumerated and traced in England and Wales at the 2011 Census. During the following decade, 58,779 of these members died or embarked and did not re-enter England and Wales. This left 526,289 expected members at Census 2021. Of these, 473,815 (90.0%) were enumerated and traced in 2021.

Analysis of forward linkage between 2011 and 2021 is based on demographic characteristics of individuals at the 2011 Census.

Age and sex

Linkage rates vary noticeably by age and sex. The lowest rates are among young adults aged 15 to 29 years and older adults aged 80 to 85 years and over. The lowest linkage rates for males are:

- 80.1% for those aged 20 to 24 years
- 81.5% for those aged 25 to 29 years
- 79.4% for those aged 85 years and over

There are slightly higher corresponding rates for females:

- 86.2% for those aged 20 to 24 years
- 86.7% for those aged 25 to 29 years
- 87.1% for those aged 85 years and over

Lower linkage rates among younger adults may reflect increased mobility because of early career changes, renting patterns, relationship transitions, and international migration.

Lower linkage among those aged 85 years and over may reflect moves into care homes or between health and care settings and an increased likelihood of census forms being completed by carers or staff. These are all factors that may have introduced inconsistencies or errors. Reduced linkage may also be influenced by delays in death registrations during the coronavirus (COVID-19) pandemic, particularly for this age group. There were 2.8% of death registrations that had a delay of six months or longer between occurrence and registration in 2021.

Marital and civil partnership status

People who were married or in a civil partnership in 2011 have the highest forward linkage rates at Census 2021, at 92.8% for males and 93.9% for females. Individuals who had never been married or civil partnered in 2011 have the lowest forward linkage rate, at 87.5%. This is up from 84.6% in the 2001 to 2011 forward linkage rates. This group is also the lowest among females, at 89.3%, which is up from 86.6% in 2011. Forward linkage rates are lowest for males who were separated, at 85.0%, which is slightly higher than 82.1% in 2011.

Household position

Household position is based on "minimal household units". Individuals living in communal establishments (CE) in 2011 have the lowest linkage rates overall; there are lower rates for males (69.2%) than for females (74.6%). These rates are slightly higher than the equivalent 2011 census rates, which were 66.8% for males and 72.7% for females. This group represents only 0.9% of the overall sample.

Over half of those living in CE (56.9%) were full-time students in 2011. Lower forward linkage reflects high mobility and changing living arrangements, which are often associated with student populations, and is consistent with patterns observed between 2001 and 2011.

Married couples in private households have the highest forward linkage rates (94.0% for males and 94.3% for females). There is a larger difference between unmarried males and females (83.7% for males and 89.4% for females). Cohabiting couples also have higher linkage rates for females (91.1%), compared with males (88.8%). The largest sex difference is for lone parents, with a linkage rate of 91.3% for females, compared with 84.8% for males. Linkage patterns for dependent children reflect those of their household type, so children living with couples have higher linkage rates (91.0%) than those living with lone parents (86.0%).

Country of birth

Forward linkage rates are consistently higher for individuals born in the UK. Males were linked at 90.9% and females were linked at 93.5%, which is higher than 87.4% for males and 90.3% for females in 2011. Those born in England have the highest linkage rates, compared with other UK countries. This is broken down as:

- 92.3% for those born in England
- 92.1% for those born in Wales
- 90.8% for those born in Scotland
- 89.8% for those born in Northern Ireland

This results in an overall UK-born linkage rate of 92.2%.

The non-UK-born forward linkage rate is 79.0%, comprising 76.3% for males and 81.3% for females. This also closely reflects the pattern seen between 2001 and 2011. The highest linkage rates for non-UK-born LS members are for those born in:

- Bangladesh (88.7%)
- Pakistan (85.6%)
- Ireland (84.3%)

The lowest linkage rates are for those born in:

- the EU (72.0%)
- the Americas (71.0%)
- Antarctica and Oceania (67.3%)

Non-UK-born females (except those born in Canada) have higher linkage rates than non-UK-born males.

Ethnic group

Forward linkage rates are highest for individuals identifying as:

- English, Welsh, Scottish, Northern Irish, or British (92.6%)
- Bangladeshi (89.3%)
- Indian (87.3%)
- Pakistani (86.8%)

This follows the same pattern seen in 2011.

Linkage rates are lower for those identifying as:

- Gypsy or Irish Traveller (63.5%)
- Chinese (68.6%)
- Other White (72.0%)

Females showed consistently higher linkage rates than males across all ethnic groups.

Housing tenure

Forward linkage rates reflect the stability of living arrangements. The highest rates are for those owning outright (93.7%) and living in mortgaged homes (93.1%). The lowest rates are for those in more temporary arrangements, including private renting (81.7%) and living in CE (67.1%). This follows the same pattern of linkage between 2001 and 2011.

Females consistently have higher linkage than males across nearly all tenure types.

Highest level of qualification

The "highest level of qualification" was broken down into seven categories:

- no qualifications
- Level 1 – one to four GCSEs or equivalent
- Level 2 – five or more GCSEs or equivalent
- apprenticeship
- Level 3 – two or more A Levels or equivalent
- Level 4 plus – undergraduate degree or higher
- other, which includes vocational or unclassified foreign qualifications

Linkage rates are higher for those whose highest qualification was Level 2 (91.9%), apprenticeship (93.1%), or Level 3 (91.9%). Linkage rates for graduates (Level 4 plus) are slightly lower, at 90.9%. Across all qualification levels, linkage rates are higher for females than for males.

Economic position

Linkage rates differed notably across economic activity. The highest rates are among retired individuals (93.5%), those in part-time employment (93.3%), and those in full-time employment (91.6%). This broadly reflects patterns seen in 2011. The lowest linkage rates in 2021 are for economically inactive students, at 79.9%, which is comparable to 79.7% in 2011. After students, those in the "other" economic activity category (81.0%) and those who are unemployed (83.2%) have the lowest linkage rates.

Forward linkage is higher for females across every economic activity type. For example, part-time employed females linked at 94.6%, compared with 88.5% for males, and unemployed females linked at 88.3%, compared with 79.5% for males.

National Statistics Socio-economic Classification

Forward linkage rates are highest among those in intermediate occupations (94.0%), and in higher and lower managerial/professional groups (both at 92.6%). This is an increase from 2011, where intermediate occupations had the highest linkage rate of 91.2%.

Linkage rates generally decline across National Statistics Socio-economic Classification (NS-SEC), with employer /self-employed individuals linking at 90.2% and those in routine occupations linking at 89.0%. The lowest linkage rates are for full-time students (81.5%) and individuals who had never worked or were long-term unemployed (83.2%), which reflects 2011 rates.

Females consistently have higher linkage than males. Lower linkage among students and those who are long-term unemployed may reflect higher mobility, irregular engagement with formal systems, and the additional disruptions of the coronavirus (COVID-19) pandemic.

Region and local authority

Linkage rates are broadly consistent across most regions of England and Wales, with rates of 90.0% or higher in nearly all areas. The highest linkage is observed in the East of England and the South West (both 91.6%). The lowest rate is in London (84.6%), which reflects a more mobile and diverse population. London also has the largest variation by sex, with linkage of females at 86.3%, compared with 82.9% for males.

There is more variation in linkage rates at local authority level. The highest forward linkage rates are in rural or less densely populated districts, including:

- Ryedale (95.6%)
- North West Leicestershire (95.5%)
- Melton (95.3%)
- Amber Valley (94.7%)
- Cotswold (94.5%)
- West Dorset (94.5%)

The population of these areas tends to be older and more residentially stable.

The lowest linkage rates are in large urban areas, particularly parts of London and cities, including:

- Kensington and Chelsea (72.3%)
- Westminster (73.4%)
- Camden (80.0%)
- Oxford (82.4%)
- Manchester (84.2%)

These areas tend to have younger populations with higher levels of mobility and diverse migrant flows, especially in inner London, which may have both contributed to lower linkage.

5 . Birth forward linkage rates

Forward linkage rates were also calculated for Longitudinal Study (LS) members born between the 2011 Census and Census 2021. This allows us to quantify the proportion of these people born during this period who were enumerated at Census 2021.

Table 5 summarises the forward linkage rates for traced LS members born between each census since 1971. It shows that the forward linkage of intercensal births declined steadily from 1981 onwards but partially recovered to 88.3% in 2021.

Table 5: Linkage rates in 2021 for people born since the last census are the highest in 30 years
Forward linkage for Longitudinal Study (LS) members who entered as a birth in the intercensal period, by decade, 1971 to 2021

Traced LS members born between censuses	Births between prior and following census	Died	Embarked	Eligible to be on following census	Not found at following census	Enumerated at following census	Linkage rate to following census (%)
1971 to 1981	71,887	1,189	696	70,002	5,075	64,927	92.8
1981 to 1991	73,005	703	313	71,989	6,288	65,701	91.3
1991 to 2001	70,548	461	301	69,786	9,953	59,833	85.7
2001 to 2011	72,845	415	601	71,829	10,680	61,149	85.1
2011 to 2021	75,375	330	720	74,325	7,785	66,540	88.3

Source: Longitudinal Study from the Office for National Statistics

Parents' age at birth

Linkage rates for those born between censuses increase with parents' age at birth. The lowest linkage rates are for those with a mother aged under 18 years (79.7%) or with a father aged under 18 years (80.9%). The highest rates are among parents aged 39 to 45 years (90.8% for mothers and 91.1% for fathers). Rates then decline again for parents aged over 45 years.

Registration type

Linkage rates where the baby was jointly registered by both parents are highest (90.6% for cohabiting parents and 90.5% for parents who were married or in a civil partnership). Linkage rates are lower for joint registrations where parents were not cohabiting (86.0%) and for sole registrations (80.6%).

Birth parity

Linkage rates decline as the number of previous births increases. Forward linkage of first-born and second-born children are the highest, at 89.9% and 90.8%, respectively. The rate is 74.1% for children of parity six or more.

Year of birth

Linkage rates vary minimally, between 88.9% and 90.0%, by year of birth of the child.

Region of birth

Linkage rates by region of birth are largely consistent across England and Wales. The North East, South East and South West have linkage rates of 91.3%. Most other English regions and Wales are close to 90.0%. London has a noticeably lower linkage rate of 84.4%.

Parents' country of birth

Linkage rates for those born between censuses are highest for parents born in England (92.3% for mothers and 92.9% for fathers). They are lowest for parents born outside of the UK (82.4% for mothers and 82.6% for fathers).

Within UK nations, there is a gradient in linkage rates based on mother's country of birth:

- England (92.3%)
- Wales (91.0%)
- Scotland (89.1%)
- Northern Ireland (85.6%)

For father's country of birth, forward linkage rates are consistent across all UK nations.

6 . Immigrant forward linkage rates

Forward linkage rates were also calculated for Longitudinal Study (LS) members immigrating to England and Wales for the first time between the 2011 Census and Census 2021. This allows us to quantify the proportion of this group that were enumerated at the Census 2021.

Table 6 shows the linkage rate in 2021 for LS members who entered the study as immigrants since the 2011 Census. The forward linkage rate is higher for immigrants who arrived between 2011 and 2021 (45.2%), compared with those who arrived between 2001 and 2011 (44.6%).

Table 6: Linkage rates for intercensal migrants are lower than for those who were present at the last census or for intercensal births
Forward linkage for Longitudinal Study (LS) members who entered as immigrants in the intercensal period, by decade, 1971 to 2021

Censuses	Immigrant new entrants between prior and later census	Died	Embarked	Eligible to be on following census	Not found at following census	Enumerated at following census	Linkage rate to next census after immigration (%)
1971 to 1981	31,255	198	2,734	28,323	13,466	14,857	52.5
1981 to 1991	22,957	122	1,261	21,574	13,331	8,243	38.2
1991 to 2001	45,908	212	1,835	43,861	28,966	14,895	34.0
2001 to 2011	81,134	245	2,735	78,154	43,270	34,884	44.6
2011 to 2021	102,492	417	4,327	97,748	53,553	44,195	45.2

Source: Longitudinal Study from the Office for National Statistics

Age and sex

Forward linkage rates in 2021 for intercensal immigrants (45.2%) are approximately half the equivalent rate for UK-born LS members who were present at the 2011 Census (92.2%). Overall, 44,195 of 97,748 immigrants were linked.

Children have the highest linkage rates for intercensal migrants. Rates are 63.8% for those aged 5 to 9 years and 60.5% for those aged 10 to 14 years. This is consistent with the relatively high rates observed in 2011 for those aged under 19 years (59.1%). Rates fall sharply for young adults aged 20 to 24 years (34.1%), before rising to around 50.0% for those in their 30s and 40s. Linkage declines steadily to 33.9% for those aged 60 to 64 years. The lowest levels (9.0%) are seen among older entrants aged 85 years and over. This age gradient was also seen in 2011, where linkage fell to 28.9% among those aged 70 years and older.

Sex differences are consistent across decades. In 2021, linkage rates for females are generally higher than for males. The largest difference by sex is for those aged in their 30s (for example, for those aged 35 to 39 years, rates are 56.4% for females and 44.7% for males). Similarly, 47.5% of female intercensal immigrants were linked in 2011, compared with 41.8% of males. In 2021, male linkage rates are higher than female linkage rates for those aged 5 to 9 years (64.6%, compared with 62.9%), and for those aged 85 years and over (10.3%, compared with 8.0%).

Year of entry

Year of entry to the LS is recorded as the first registration with the NHS. Forward linkage rates for immigrants trend upwards across the decade:

- 42.5% for those arriving in 2011
- 43.9% for those arriving in 2015
- 49.6% for those arriving in 2020

Sex differences are consistent across decades. In 2021, linkage rates for females are generally higher than for males. The largest difference by sex is for those aged in their 30s (for example, for those aged 35 to 39 years, rates are 56.4% for females and 44.7% for males). Similarly, 47.5% of female intercensal immigrants were linked in 2011, compared with 41.8% of males. In 2021, male linkage rates are higher than female linkage rates for those aged 5 to 9 years (64.6%, compared with 62.9%), and for those aged 85 years and over (10.3%, compared with 8.0%).

This reflects the pattern observed in 2011, where linkage rates were higher for more recent arrivals (52.8% for 2010, compared with 38.7% for 2002). This may be because immigrants who arrived earlier in the decade have had more time to subsequently leave the country. Particularly among short-term migrants, such as international students, it is possible that their embarkation has not been flagged on the NHS Personal Demographics Service. The larger cohorts in later years (for example, 15,382 arrivals in 2020) also contribute to overall linkage outcomes, which reinforces the upward trend over the decade.

7 . Backward linkage rates

Backward linkage rates measure the percentage of LS members present at a census who were also present at the previous census, excluding those who have entered the study during the intercensal period as new births or immigrants. It is expressed as:

$$\frac{(\mathrm{N}-\mathrm{t}-\mathrm{p})}{\mathrm{N}-\mathrm{t}} * 100$$

Where “N” is the number of traced LS members found at the current census; “t” is the number of new entrants (births and immigrants) in the decade preceding the census and other new entrants that were first captured in Census 2021 and; “p” is the number of traced LS members found at the current census but not enumerated at the previous census.

Table 7 shows 648,111 LS members were enumerated in England and Wales and successfully traced by NHS England Digital at Census 2021. Between the 2011 Census and Census 2021:

- 66,539 traced members were born
- 43,194 members were immigrants to England and Wales
- 13,978 were new entrants at Census 2021 (people with a LS date of birth who have never been enumerated at a previous census and have not been traced at birth or as an immigrant in the Personal Demographics Service)

Of the 524,400 LS members expected at the 2011 Census, 473,815 were enumerated and 50,585 were "not found". The overall backward linkage rate was therefore 90.4% for 2021 to 2011.

Table 7: The backward linkage rate for 2021 to 2011 is higher than the previous decade, but lower than rates achieved in earlier censuses

Backward linkage rate for the traced Longitudinal Study (LS) census sample, by decade, 1971 to 2021

Censuses	Traced members at later census	Born since prior census	Immigrated since prior census	New entrants at later census	Eligible at prior census	Not found at prior census	Enumerated at prior census	Linkage rate (%)
1981 to 1971	527,956	64,669	14,799	33,125	448,488	33,161	415,327	92.6
1991 to 1981	535,015	65,700	8,259	16,901	461,056	39,693	421,363	91.4
2001 to 1991	536,377	59,820	14,943	14,954	461,614	43,052	418,562	90.7
2011 to 2001	578,821	61,149	34,884	15,304	467,484	45,628	421,856	90.2
2021 to 2011	648,117	66,539	43,194	13,978	524,406	50,590	473,816	90.4

Source: Longitudinal Study from the Office for National Statistics

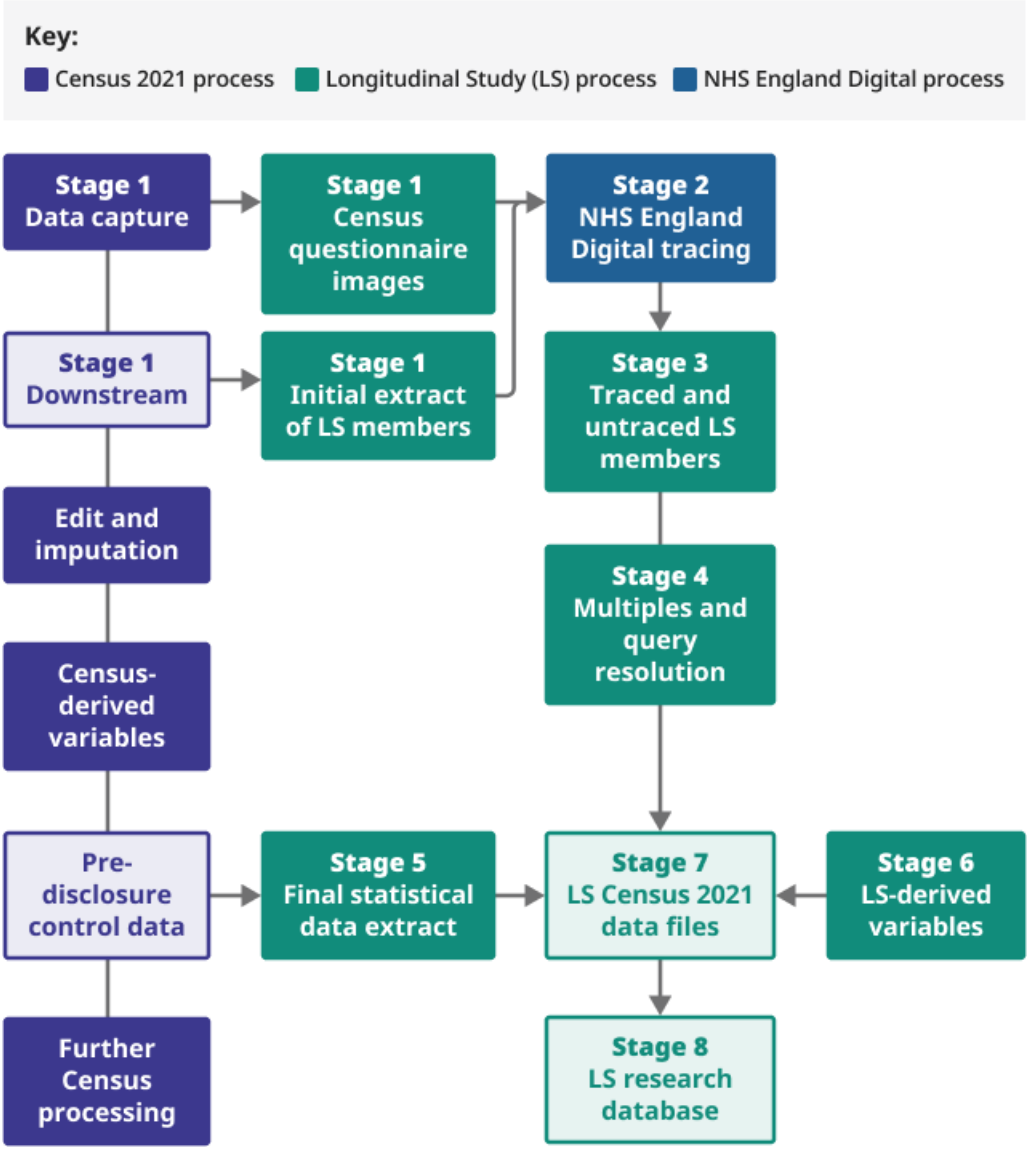
Notes

1. New entrants at later censuses were included in the eligible population for 1981 to 1971, 1991 to 1981, and 2001 to 1991 in our published tables. Some members may enter the LS as an intercensal birth or intercensal immigrant but can be traced back to the previous census. These individuals appear as eligible in this table, as they were present at the previous census.

8 . The Census 2021 link project

The Census 2021 link project linked the Longitudinal Study (LS) sample of Census 2021 data to the LS database. Figure 2 gives an overview of the process, and each stage is described in the following subsections.

Figure 2: Overview of the Census 2021 linkage process



Source: Longitudinal Study from the Office for National Statistics

Stage 1: Longitudinal Study census tracing extract

We captured data from questionnaires that were completed online and scanned completed paper questionnaires. The captured data entered the Census 2021 downstream processing pipelines.

Our LS team took extracts of census data and images of paper census questionnaires for all people who were enumerated at the census who recorded their date of birth as one of the four LS sample dates of birth, and for non-members residing within the same Unique Property Reference Number (UPRN). These initial extracts contained only the variables that were required for tracing against NHS records (listed in the following subsection).

Stage 2: Tracing of census records at NHS England Digital

Initial census data extracts and questionnaire images (where required to assist with tracing) were sent to the NHS England Digital (NHSD) tracing team.

Tracing involves matching people from the census with LS dates of birth with their NHS registration on the Personal Demographic Service (PDS). Tracing used NHS registration data as of July 2021. Tracing was carried out in stages through matching routines, using the following variables.

Census variables available for matching

- questionnaire type (internet/paper questionnaire)
- questionnaire ID
- family name
- given name
- other given name
- date of birth
- sex
- address of enumeration
- postcode of enumeration
- visitor postcode of usual residence
- address one year ago indicator
- postcode of residence one year ago
- address country one year ago
- person type (resident or visitor)
- month/year of arrival
- intention to stay
- student indicator
- term-time address indicator
- alternative address postcode
- alternative address type
- alternative address country
- corrected postcode of enumeration
- UPRN

NHS variables available for matching

- surname
- forename
- other name
- address
- postcode
- date of birth
- sex
- postings
- historic records of patient's GP/home addressand dates of births

The tracing service provided by NHSD used the Master Person Service (MPS). The matching strategy was based on a combination of automatic, probabilistic (algorithmic), and clerical matching methods.

Automatic matching

Alphanumeric tracing was the first stage for census tracing. This relies on exact matching of similar information between the census and PDS, and required family name, date of birth and sex, at a minimum.

Other information, such as other given names, Postcode Address File (PAF), address key and postcode, were used where required.

Algorithmic tracing

This stage produces a probability score that records are a match based on different blocks of information. This relies on information for family name, given name, sex, date of birth, and postcode. This stage uses Soundex to account for differences in spelling between similar sounding names, for example, "Smith" and "Smyth".

Clerical matching

An NHSD team carried out clerical matching for records that were not linked by the previous two stages. This team also had access to census paper questionnaires to aid their decision-making.

Clerical matching accounts for 9.2% of all traced records. However, some subgroups of LS members are more likely to be traced clerically than others. Males were less likely to be traced automatically, particularly young adult males. Males aged 25 to 29 years were least likely to be automatically traced, at 72.8%; this is 18% lower than the overall 90.8% rate for automatic tracing. Males aged 20 to 24 years were traced at 76.1% and those aged 30 to 34 years were traced at 76.8%. Females aged 20 to 24 years were also less likely to be automatically traced, at 80.0%.

The White Roma ethnic group were least likely to be traced automatically, at 76.1%, followed by Asian British Chinese, at 79.1%, and White Gypsy/Irish Traveller, at 83.2%. Considering economic activity, students were least likely to be automatically traced; economically active students were automatically traced 80.6% of the time, compared with 81.7% for other students.

The LS would have traced fewer of these groups without additional clerical matching. This would have introduced bias into the LS sample and made it less representative of the England and Wales population.

No matches

Records that were not matched by the automatic matching process were sent for matching by an operator who examined the data to see if a match could be made. Through this clerical matching process, these records either became matched and therefore traced, or remained as unmatched (untraced).

Any discrepancies between the matched census and NHS records for sex or date of birth of the LS member were recorded on the final LS database. These discrepancies can occur because the sex or date of birth has been entered incorrectly on the census or the NHS record or scanned incorrectly when reading the census record.

Census 2021 traced records show a date of birth discrepancy in 1.7% of cases.

Multiple matches

Multiple matches occur where an NHS record matches to more than one census record.

Common reasons for multiple matches include:

- students who were enumerated at their term-time and their home address
- people enumerated multiple times on the same census questionnaire
- an internet and a paper return for the same address
- people enumerated on more than one census questionnaire for the same address
- people enumerated with parents at two different addresses
- other complex enumerations involving two or more census questionnaires
- incorrectly matched records

Incorrectly matched records occur when information for two or more individuals on a census form is the same or very similar. This happens most often in the case of twins who share the same date of birth and surname and may also have similar first names.

Multiple matched records that were resolved by the operator were either matched to an NHS record and marked as traced, or remained unmatched and therefore untraced.

Overall, 90.8% of matched records came from automatic matching. The remaining 9.2% of matched records were completed by an operator visually looking at the census and NHS record. Final tracing rates are shown in Section 2: Tracing rates.

Stage 3: Identification of traced and untraced census records

On completion of Stage 2, a file was returned to our LS team from NHSD that indicated whether the LS member was traced (matched to NHS records), untraced (not matched to NHS records), or if there were multiple matches to NHS records or further anomalies identified by the matching process. Multiple matches and anomalies were then resolved at Stage 4.

Failure to trace LS members can occur for several reasons, including that:

- an individual may not have registered for NHS health care; they may have private health care, may have recently arrived in England and Wales and have not registered with a doctor, or they may have left England and Wales without ever registering with a doctor
- their date of birth was entered incorrectly (or scanned incorrectly) on the census questionnaire, and no match to NHS records has been determined
- their date of birth was incorrect on NHS records, and no match to the census has been made

Untraced LS members can subsequently become traced if sufficient additional information is received, such as a correction to a date of birth on a LS member's NHS record or the LS member registers with a doctor, that allows a match to be made. This is recorded in the LS research database through a variable that indicates when a match was made.

Stage 4: Resolution of multiple matches and other matching anomalies

The multiple matches and other matching anomalies identified by NHSD during the tracing process were then resolved by our LS team. We then decide which census record to treat as the primary census record for the LS member by viewing both the full census record and the census questionnaire image, where required. The other matched census record(s) were then recorded as secondary, multiple enumeration records.

Stage 5: Final census extract

A final extract of census data for the LS research database was taken for all LS members and for all non-members present in a LS member's household. This was completed after editing and imputation, and after the process to derive census variables.

These data include:

- census response variables
- census-derived variables (including geographies)
- imputation flags

Stage 6: Longitudinal Study-derived variables

At Stage 6, we derived additional variables, including:

- flags to show whether someone had moved between censuses
- distance moved between censuses
- relationships between the LS member and LS non-members
- resident type
- residence type

Some census variables were renamed for consistency with previous census variables in the LS.

Stage 7: Quality checking the Longitudinal Study database

The results from Stages 4, 5 and 6 were combined to create a series of Census 2021 datasets ready for use by researchers. The LS Research Database was then tested for accuracy and completeness through alpha and beta testing.

Alpha testing

Alpha testing comprised several data integrity checks to ensure that the data were consistent, complete and representative of a 1.1% sample of Census 2021 records.

The alpha test was successfully completed in October 2025.

Beta testing

Nine research projects were selected to test the new LS database containing the 1.1% sample of Census 2021 data. The research projects were undertaken by experienced LS users and ran from March 2025 to October 2025.

The beta tests aimed to test the data in their final form to ensure they were correct and fit for purpose. The beta tests used the Census 2021 data and covered a broad range of topics. Any errors in the data were highlighted and addressed before the launch of the final version of the LS research database.

Stage 8: Launch of the linked Census 2021 data

The LS database containing the Census 2021 data was launched and made available for research in October 2025. The final research database contains 648,366 LS members (excluding visitors and unretained records) and 1,397,084 non-LS members enumerated in Census 2021.

The sample of LS members is classified in Table 8 according to how they were enumerated in Census 2021. This means that LS members were either usual residents, short-term residents, or visitors.

Further explanation of census processing and its stages can be found in our Census 2021 for England and Wales general report.

Table 8: Summary of Census 2021 data in the Longitudinal Study research database
Longitudinal Study (LS) members and non-members enumerated in an LS member's household at Census 2021

		Traced	Untraced	Number of records
LS members enumerated at Census 2021		648,117	2,564	650,681
of which:				
	- enumerated as a usual resident	642,837	1,898	644,735
	- enumerated as a short-term resident	1,382	182	1,564
	- student only enumerated at their family address (not at their term-time usual residence) (Note 1)	2,029	38	2,067
	- visitors (Note 1)	1,556	395	1,951
	- unretained2 (Note 2)	313	51	364
Non-members enumerated in an LS member's household				1,397,084
LS members - multiple enumerations (Note 3)				10,646
Non-members - multiple enumerations (Note 4)				27,979

Source: Longitudinal Study from the Office for National Statistics

Notes

1. Students enumerated at their family address and visitors did not complete a full census questionnaire.
2. Unretained records are partial records that are present at Stage 1 but dropped by census processing by Stage 5.
3. Includes all secondary records of LS members who also have a primary record enumerated at Census 2021.
4. Non-member multiple enumerations are only identified where the LS member has a multiple enumeration themselves; non-member information is only held where they are resident in the same household as the LS member.

Events that affected the Census 2021 link project

We required the tracing activity at NHSD to take place within a secure, controlled office environment. As a result of restrictions imposed during the coronavirus (COVID-19) pandemic, NHSD prepared new office space that was sufficiently secure and also allowed staff to socially distance while working.

The Medical Research Information Service Integrated Database Administration System (MIDAS) had been used by NHSD since the 2011 Census for three main purposes:

- to carry out LS tracing (automated and clerical) that underpinned all LS data linkage
- to provide data on NHS registration events that happened to flagged LS members
- to manage the LS cohort, adding members to the cohort from birth, immigrant and census records

The MIDAS system was used for the last time to process 2019 life events data and to create a 2020 file of NHS registration data. The cohort at the end of this round of processing was then extracted from MIDAS and used for tracing Census 2021 records, alongside "live" PDS data, where new LS members were identified. NHSD used the MPS system for automated tracing and recruited a team to clerically trace records that were not traced automatically.

We have developed new business processes and system functionality to replace previous arrangements since MIDAS was decommissioned. This included transferring the cohort file from NHSD to the Office for National Statistics (ONS), within data security and governance requirements. New processes were developed for births, deaths and immigrants. All other LS topics remain unchanged from the last run under MIDAS, and new processes are in development to update the LS cohort in the future.

9 . Definitions

Household

One person living alone, or a group of people (not necessarily related) living at the same address, who share cooking facilities and share a living room or sitting room or dining area.

Communal establishment

An establishment that provides "managed" residential accommodation, which means providing full- or part-time supervision of the accommodation.

Usual resident

Anyone who, on 21 March 2021:

- was in the UK and has stayed for 12 months
- was in the UK and intended to stay for 12 months
- had a permanent UK address, but was outside the UK and intended to be outside the UK for less than 12 months

Short-term resident

Anyone born outside the UK who has stayed, or intends to stay, in the UK for a period of three months or more but less than 12 months.

Visitor

A person staying overnight at an address at which they are not usually resident on 21 March 2021. These people should be counted as visitors where they are staying overnight and were separately counted at their usual UK address, or have provided their country of usual residence if they are a non-UK resident.

Minimal household unit

The smallest unit or group within a household expected to share a similar lifestyle. More information on households is available in the Guide to using the Office for National Statistics Longitudinal Study: Households and families thematic guide (PDF, 494KB).

Dependent children are grouped with their parents and non-dependent children are regarded as adults not in a family.

Dependent children

Those aged under 16 years, or those who are aged 16 to 18 years and in full-time education.

10 . Data on Longitudinal Study 2011 to 2021: Completeness of census linkage

Longitudinal Study, England and Wales: tracing rates
Dataset | Released 7 July 2026
Tracing rates by different variables for members of the Office for National Statistics Longitudinal Study in Census 2021 for England and Wales.

Longitudinal Study, England and Wales: sampling fractions
Dataset | Released 7 July 2026
Sampling fractions by different variables for members of the Office for National Statistics Longitudinal Study that were in Census 2021 for England and Wales.

Longitudinal Study, England and Wales: forward linkage rates
Dataset | Released 7 July 2026
Forward linkage rates by different variables between the 2011 Census and Census 2021 for the Office for National Statistics Longitudinal Study for England and Wales.

Longitudinal Study, England and Wales: forward linkage rates for intercensal births
Dataset | Released 7 July 2026
Forward linkage rates for members of the Office for National Statistics Longitudinal Study for England and Wales who were born between the 2011 Census and Census 2021.

Longitudinal Study, England and Wales: forward linkage rates for intercensal immigrants
Dataset | Released 7 July 2026
Forward linkage rates for immigrants who entered the Office for National Statistics Longitudinal Study for England and Wales between the 2011 Census and Census 2021.

Longitudinal Study, England and Wales: tracing method
Dataset | Released 7 July 2026
Automatic and clerical tracing rates by different variables for members of the Office for National Statistics Longitudinal Study in Census 2021 for England and Wales.

11 . Related Links

Office for National Statistics (ONS) Longitudinal Study
Article
The ONS Longitudinal Study (LS) contains linked census and life events data for a 1% sample of the population of England and Wales.

Longitudinal Study 2001 to 2011: completeness of census linkage (PDF, 1.5MB)
Article
Assessment of the linkage of 2011 Census records to the existing study records, comparing success rates in tracing and linking 2011 Census records to rates achieved in 1981, 1991 and 2001.

12 . Cite this page

Office for National Statistics (ONS), released 07 July 2026, ONS website, supporting methodology article, Longitudinal Study 2011 to 2021: Completeness of census linkage