

Longitudinal Study 2001 - 2011

Completeness of census linkage

Series LS No. 11

Kevin Lynch, Shayla Leib, James Warren, Nicky Rogers, Julian Buxton

Contents

Management Summary	6
1. Introduction	7
2. Background	8
3. The 2011 Census link project	9
Stage 1: LS census extract	9
Stage 2: Tracing of census records at HSCIC.....	10
Automatch process.....	10
Non-matches.....	11
Multiple matches	11
Stage 3: Identification of traced and untraced census records	12
Stage 4: Resolution of multiple matches and other matching anomalies	12
Stage 5: Final census extract.....	12
Stage 6: LS derived variables	13
Stage 7- Quality checking the LS database.....	13
Stage 8: Launch of the linked 2011 Census data	14
4. The 2011 Census	16
Population base for enumeration	16
Usual residents.....	17
Short-term UK resident.....	17
Visitors	17
2011 Census questions.....	17
Other differences	18
Edit and imputation methodology	18
Classifications	18
Census geography	19
Census quality	19
Response rates	19
Census Quality Survey	19
Data capture, coding and cleaning	20
Edit and item imputation	20
Census corrections.....	20
5. Tracing rates	21

Tracing rates definition.....	21
Tracing rate tables	21
Tracing rates summary	22
Age and sex	23
Marital and civil partnership status.....	23
Economic position	23
National Statistics Socio-economic Classification (NS-SEC)	23
Country or region.....	24
Country of birth.....	24
Year of arrival.....	24
Ethnic group	24
Communal establishments	24
6. Sampling fractions.....	26
Sampling fractions definition	26
Sampling fractions tables	27
Sampling fractions summary	27
Overall.....	27
Age and sex	28
Ethnic group	31
Country and region.....	33
7. Linkage rates.....	34
Forward linkage rates definition	34
Forward linkage rate tables.....	34
Forward linkage summary – 2001 to 2011 Census	36
Overall forward linkage rates for by census.....	36
Overall forward linkage rates for 2001 to 2011 Censuses.....	36
Age and sex	37
Marital status.....	37
Household position.....	37
Country of birth.....	38
Ethnic group	38
Housing tenure.....	38
Highest level of qualification	38
Economic position	39
National Statistics Socio-economic Classification (NS-SEC)	39
Country, region and local authority of enumeration.....	39

Forward linkage summary – LS members born between the 2001 and 2011 Censuses	39
Overall forward linkage rates for by census, 1971 to 2011	39
Overall forward linkage rates for LS members born between 2001 and 2011 Censuses	40
Parents’ age at birth	41
Registration type	41
Birth parity	41
Year of birth.....	41
Country and region of birth	41
Parent’s country of birth.....	41
National Statistics Socio-economic Classification (NS-SEC)	42
Forward linkage summary – LS members entering as immigrants between 2001 and 2011 Censuses.....	42
Overall forward linkage rates for LS members entering as immigrants between the 2001 and 2011 Censuses	42
Age and sex	43
Year of entry.....	43
Country and region of first NHS registration	43
Backward linkage rates	44
Overall backward linkage rates.....	44
8. Conclusion	46
Appendices	49
Appendix A: Census data - summary of variables by topic area, 1971 to 2011 Censuses.....	49
References	52

List of Tables

Table 1: Summary of the 2011 Census data in the LS research database	14
Table 2: LS Tracing rates, 1971-2011	22
Table 3: LS tracing rates by resident type in 2011.....	23
Table 4: Observed sampling fractions by census	27
Table 5: 2011 LS Sampling fractions for traced LS members by sex and age - Males	29
Table 6: 2011 LS Sampling fractions for traced LS members by sex and age - Females	30
Table 7: 2011 LS Sampling fractions by ethnic group	32
Table 8: 2011 LS sampling fractions by country or region.....	33
Table 9: Forward linkage rates for traced LS members, 1971 to 2011 Censuses.....	36
Table 10: Overall forward linkage rates.....	37
Table 11: Forward linkage for traced LS members born between censuses, 1971 to 2011	40
Table 12: Forward linkage for LS members born between 2001 and 2011 Censuses	40
Table 13: LS immigrants forward linkage, 2001 to 2011	43
Table 14: Backward linkage rate for the traced LS census sample, 2011 to 2001	44
Table 15: Backward linkage rates for traced LS members, 1971 to 2011 Censuses	45

List of Figures

Figure 1: Overview of the 2011 Census link project	9
Figure 2: Structure of the Longitudinal Study research database	15
Figure 3: Sample sizes for traced LS members, 1971 to 2011	47

Management Summary

The ONS Longitudinal Study (LS) links census, life event and NHS registration information for 1% of the population of England and Wales. The first sample was taken from the 1971 Census by selecting all people born on one of four birth dates. The sample has been replenished continuously ever since from birth records, NHS registrations and later censuses where people have met the same sample selection criteria.

Longitudinal resources appreciate in value over time. The LS is now more than 40 years old and contains linked data for more than 1.1 million study members. Information from five successive Censuses (1971 to 2011) is available for more than 200,000 people. In addition, the study has been refreshed with more than 110,000 new entrants between the 2001 and 2011 Censuses.

The LS experiences low levels of attrition in comparison to other longitudinal datasets such as cohort studies and longitudinal surveys and provides unrivalled levels of richness and coverage. Because of this, the LS has traditionally attracted users from a wide range of research fields including public health, demography, social mobility, human geography and deprivation. It also underpins some of the National Statistics relating to health inequality published by ONS's Health Analysis Team.

This report is the fifth in a series reporting on the quality of linkage between the LS and the England and Wales census, as well as changes in individual circumstances over the intercensal periods. The report provides an assessment of the linkage of 2011 Census records to existing study records, and compares success rates in tracing and linking 2011 Census records to those achieved at the 1981, 1991 and 2001 Censuses. Data on the extent to which individuals are enumerated, traced and linked correctly are vital for accurate use of the LS, as are data on the demographic characteristics of those previously in the LS without a valid exit event and not found at the 2011 Census. The report also usefully provides LS users with a description of how the 2011 Census data were linked to the LS database, and a summary of the main differences between the 2001 and 2011 Censuses.

Tracing of LS records in 2011 was extremely successful with 98.8% of records identified from the 2011 Census traced to an NHS record. Tracing rates have increased due to a combination of more use of automated matching methodologies, clerical matching and the availability of more detailed information. 87.7% of those enumerated and traced at the 2001 Census, after excluding those who were known to have died or emigrated, were linked to the 2011 Census. Linkage failures are often due to unreported emigration, census non-response and discrepancies in the date of birth recorded on different data sources.

Further analysis of unlinked records has revealed some common socio-demographic characteristics that make non-linkage to the LS in 2011 more likely, including:

- being a young adult
- being born outside of the UK
- living in a communal establishment
- living in London, particularly inner London

The LS gives unique insight into the society in which we live, and is of fundamental importance to policy, government and academia. The quality of the 2011 Census and the successful linkage to the LS database gives a stamp of assurance to users that the LS research database is maintained to high standards, and also reassurance that the LS meets their research needs.

1. Introduction

The ONS Longitudinal Study (LS) links census and life event information for 1% of the population of England and Wales. Following the first successful linkage between the 1971 and 1981 Censuses, a detailed report was produced on the quality of linkage and changes in individual circumstances over the intercensal period (OPCS 1988). LS 11 is the fifth report in this series and provides an assessment of the linkage of 2011 Census records to the existing study records. It compares success rates in tracing and linking 2011 Census records to those achieved in 1981, 1991 and 2001.

Data on the extent to which individuals are enumerated, traced and linked correctly are vital for accurate use of the LS, as are data on the demographic characteristics of those previously in the LS without a valid exit event and not found at the 2011 Census, or those individuals with multiple enumerations at the 2011 Census. The report attempts to also explain loss to follow up and multiple enumerations of individuals.

This report will provide users of the LS with:

- a brief background and history of the LS
- a description of how the 2011 Census data were linked to the LS database
- a summary of the main differences between the 2001 and 2011 censuses
- the quality of the LS sample and how representative it is of the England and Wales population through tracing rates, sampling fractions and linkage rates

2. Background

The ONS Longitudinal Study (LS) links census and life event information for a 1% sample of the population of England and Wales. The original sample was selected from the 1971 Census, and incorporated data on individuals born on one of four selected dates of birth. The sample has been updated at each successive census by taking individuals with the same four dates of birth in each year and linking them to existing data.

Life event information has been added to the LS since census day in 1971, including birth and immigration (entry events) and death and emigration (exit events) of individuals with the four dates of birth. The LS now holds data on more than 1 million sample members and, at each census, data on more than 500,000 sample members. Census information is also included for all people enumerated in the same household as an LS member (referred to as LS non-members), but only information on LS members is linked over time. In 2011 there were approximately 1.2 million LS non-members linked to LS members.

Linkage of data over time is achieved by tracing against NHS patient records. Prior to 2011 tracing involved matching individuals from the census who were born on one of the four LS birth dates to the National Health Service Central Register (NHSCR). In 2011, census data were traced against the new Medical Research Information Service Integrated Database Administrative System (MIDAS) which takes patient data from the Personal Demographics Service. MIDAS will be used for all future tracing activity on the LS.

For both the 2001 and 2011 Censuses, tracing was a combination of an automated process with clerical follow up. For censuses prior to 2001, tracing was a clerical process. This work is carried out for ONS by the Medical Research Information Service team, part of the Health and Social Care Information Centre¹ (HSCIC) based at Southport. This team were previously part of the ONS, and were transferred to HSCIC on 1 April 2008 as a consequence of the 2007 Statistics and Registration Service Act.

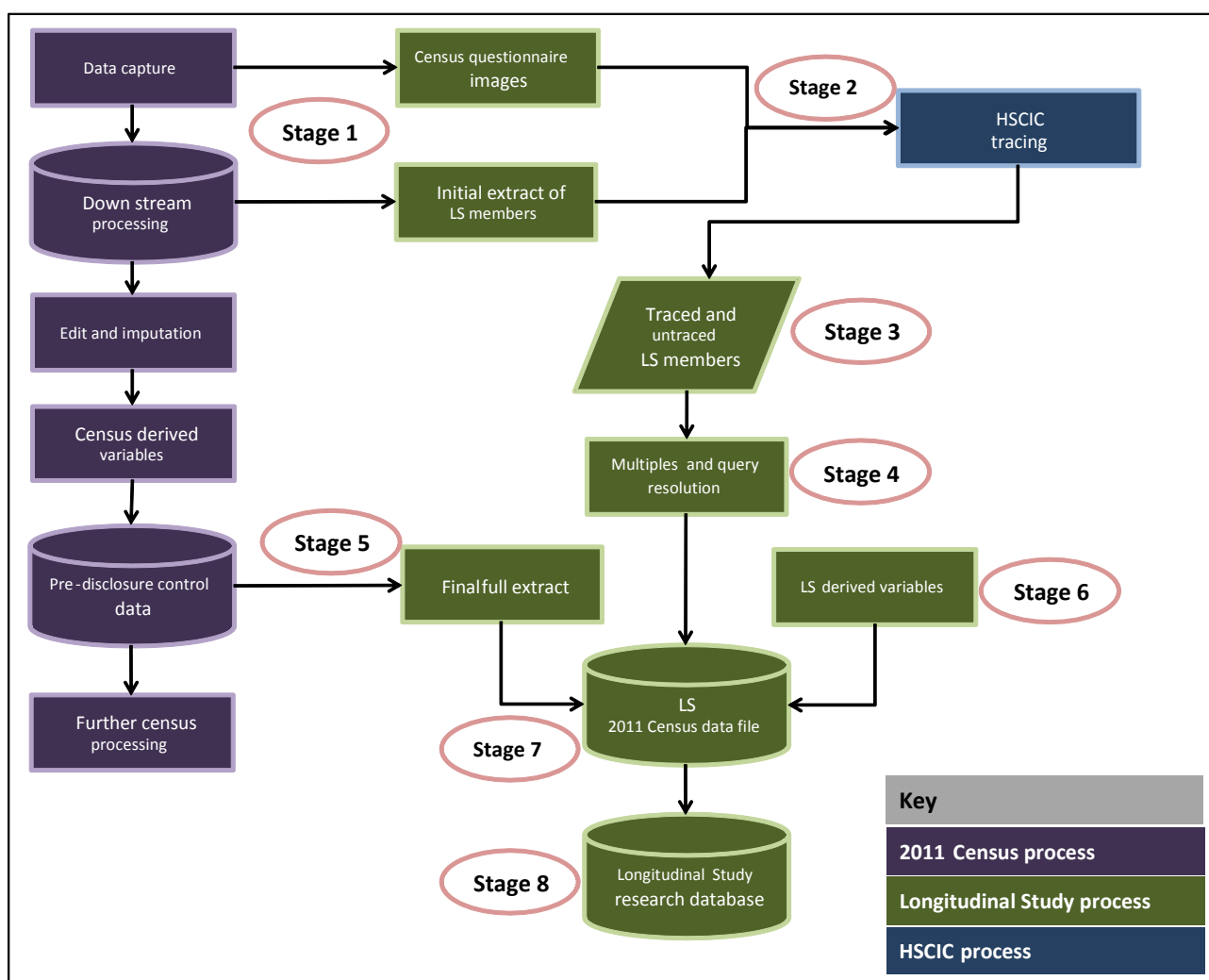
Tracing of records from the 2011 Census was completed in 2012. This was achieved using an initial tracing extract of census data created prior to the removal of multiple responses in census processing. Data sets were provided to HSCIC in batches of delivery groups. The returned traced data sets contained indicators of whether the individual was traced and whether they had been previously identified as an LS member. Where this was the case, a unique identifier was returned allowing these individuals to be linked to their previous LS history. Individuals who had been enumerated more than once in the 2011 Census were identified by cases where more than one record was returned with the same unique identifier.

More detail on the Longitudinal Study is available on the [ONS website](#).

3. The 2011 Census link project

The 2011 Census link project was the process to link the Longitudinal Study (LS) sample of 2011 Census data to the LS database. There were a number of different stages to the project and this section gives an overview of that process. Figure 1 below gives an overview of the process and each stage in the project is described more fully below.

Figure 1: Overview of the 2011 Census link project



Stage 1: LS census extract

Completed paper questionnaires were scanned and the data were captured along with data from questionnaires completed via the internet (2011 Census process - Data capture). The captured data entered the 2011 Census downstream processing environment (2011 Census process - Downstream processing). The ONS LS team took images of Census questionnaires (Longitudinal Study process – Census questionnaire images) and an initial extract of census data (Longitudinal

Study process – initial extract of LS members) for all LS members – that is all people enumerated at the 2011 Census who recorded their date of birth as one of the four LS sample dates of birth. This initial extract of data contained a limited number of census variables that were required for tracing against NHS records (see stage 2 below).

Stage 2: Tracing of census records at HSCIC

The initial extract of the census data along with images of the census questionnaires from ‘Stage 1’ were sent for tracing to the NHS Health and Social Care Information Centre (HSCIC) ([HSCIC process – HSCIC tracing](#)).

Tracing involves matching people from the census with LS dates of birth with their NHS registration on MIDAS. Tracing was carried out using NHS registration data as of July 2011. Tracing was carried out in stages through matching routines using the following variables:

Census variables available for matching	NHS variables available for matching
Channel ID (paper / internet questionnaire)	Surname
Questionnaire ID	Forename
Person number	Other name
Person forename	Address
Person surname	Postcode
Full name	Date of birth
Date of birth	Sex
Sex	Postings
Postcode of enumeration	Historic records of patients GP/home address and dates of births
Visitor postcode of usual residence	
Postcode of residence one year ago	
Address one year ago indicator	
Person type (resident / visitor)	
Month/year of arrival	
Intention to stay	
Student indicator	
Term-time address indicator	
Second address postcode	
Second address type	
Corrected postcode of enumeration	

Automatch process

Automatch was the automated computer process of matching the date of birth, surname, full name and sex from the census record with that of the NHS registration system. If a single match was made on these variables the match was said to be successful (the LS member is traced). If there was more than one match on these variables but the first two characters of the census enumeration postcode (either the usual residence postcode on a usually resident record, or postcode of usual residence on a visitor record) matches the address on the NHS registration system, a match was also said to be successful.

As well as identifying LS members that are traced, the automatch process also identified non-matches (untraced LS members) and multiple possible matches. These are described below.

Non-matches

Records that were not matched (unmatched) by the automatch procedure were sent for matching by an operator who would visually examine 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 discrepancy between the matched census and NHS records for sex or date of birth of the LS member was 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
- scanned incorrectly when reading the census record

These discrepancies only occur in a minority of cases

Multiple matches

During the tracing process, if a NHS record matched to more than one census record it was sent to an operator for clerical resolution. The operator visually examined the data to see if a match could be made. Common reasons for multiple matches include:

- students enumerated at their term-time and their home address
- enumerated multiple times on the same census questionnaire
- internet and paper return for the same address
- enumerated on more than one census questionnaires for the same address
- 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 can share the same date of birth, surname, and have very similar first names.

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

Overall around two-thirds of matched records came from automatic matching. The remaining third of matched records were completed by an operator visually looking at the census and NHS record.

The final tracing rates are shown in chapter 5 'Tracing rates'.

Stage 3: Identification of traced and untraced census records

On completion of Stage 2, a file was sent back to the LS team from HSCIC indicating whether the LS member was traced (matched to NHS records), untraced (not matched to NHS records), whether there were multiple matches to NHS records or there were further anomalies identified by the matching process. Multiple matches and anomalies were then resolved at 'Stage 4' (see below).

Failure to trace LS members can occur for a number of reasons.

- An individual may not have registered for NHS health care (they have private health care or, have recently arrived in England and Wales and have not registered with a doctor, or they 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 able to be determined
- Their date of birth was incorrect on NHS records and no match to the census has been made

Untraced LS members can subsequently at a later date 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 indicating when a match was made as part of routine updates.

Stage 4: Resolution of multiple matches and other matching anomalies

The multiple matches and other matching anomalies identified by HSCIC during the tracing process were resolved by the LS team. The decision which census record to treat as the primary census record for the LS member was made by viewing both the full census record and the census questionnaire image. The other matched census record(s) were then recorded as multiple enumeration records.

Results from the tracing and multiple resolution process were fed back into the census quality assurance process². A 'Longitudinal Study 2011 Census Linkage Report'³ has been published detailing how the Longitudinal Study was used to quality assure 2011 Census data.

Stage 5: Final census extract

A final extract of census data for the LS research database was taken for all the LS members and LS non-members present in a LS member's household. This was done after edit and imputation and the process to derive census variables, (2011 Census processes). The data include:

- Census response variables
- Census derived variables (including geographies)
- Imputation flags

Stage 6: LS derived variables

At stage 6 additional variables were derived by the LS team 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 LS database

The results from stages 4, 5 and 6 were combined to create a series of 2011 Census data sets ready for use by researchers. The LS Research Database was then tested for accuracy and completeness through Alpha and Beta Testing. Each of these test phases is discussed below.

Alpha testing

This comprised a number of data integrity checks to ensure that the data were consistent, complete and representative of a 1% sample of 2011 Census records. The alpha test was successfully completed at the end of April 2013.

Beta testing

10 research projects were selected to test the new LS database containing the 1% sample of 2011 Census data. The research projects, undertaken by experienced LS users, ran from May 2013 to October 2013.

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 2011 Census 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.

No significant errors were identified during this testing and the beta test was successfully concluded at the end of October 2013.

Stage 8: Launch of the linked 2011 Census data

The LS database containing the 2011 Census data was launched and made available for research on 1 December 2013. The final database contains 589,966 LS members and 1,264,869 non-LS members enumerated in the 2011 Census. The sample of LS members is classified in Table 1 according to how they were enumerated in the 2011 Census. This means, LS members were either usual resident, short-term residents or visitors. More explanation of these types of residents is given in the section 'Population base for enumeration' in chapter 5, 'The 2011 Census'.

Table 1: Summary of the 2011 Census data in the LS research database

	Number of records
LS members enumerated at the 2011 Census	589,996
<i>of which:</i>	
- enumerated and a usual resident	581,691
- enumerated and a short-term resident	1,848
- student only enumerated at their family address (not at their term-time usual residence) ¹	2,232
- enumerated but unable to determine if a usual resident or a short-term resident ¹	129
- enumerated only as a visitor ¹	4,096
Non-members enumerated in a LS member's household	1,264,869
LS members - multiple enumerations	12,950
Non-members - multiple enumerations²	53,808

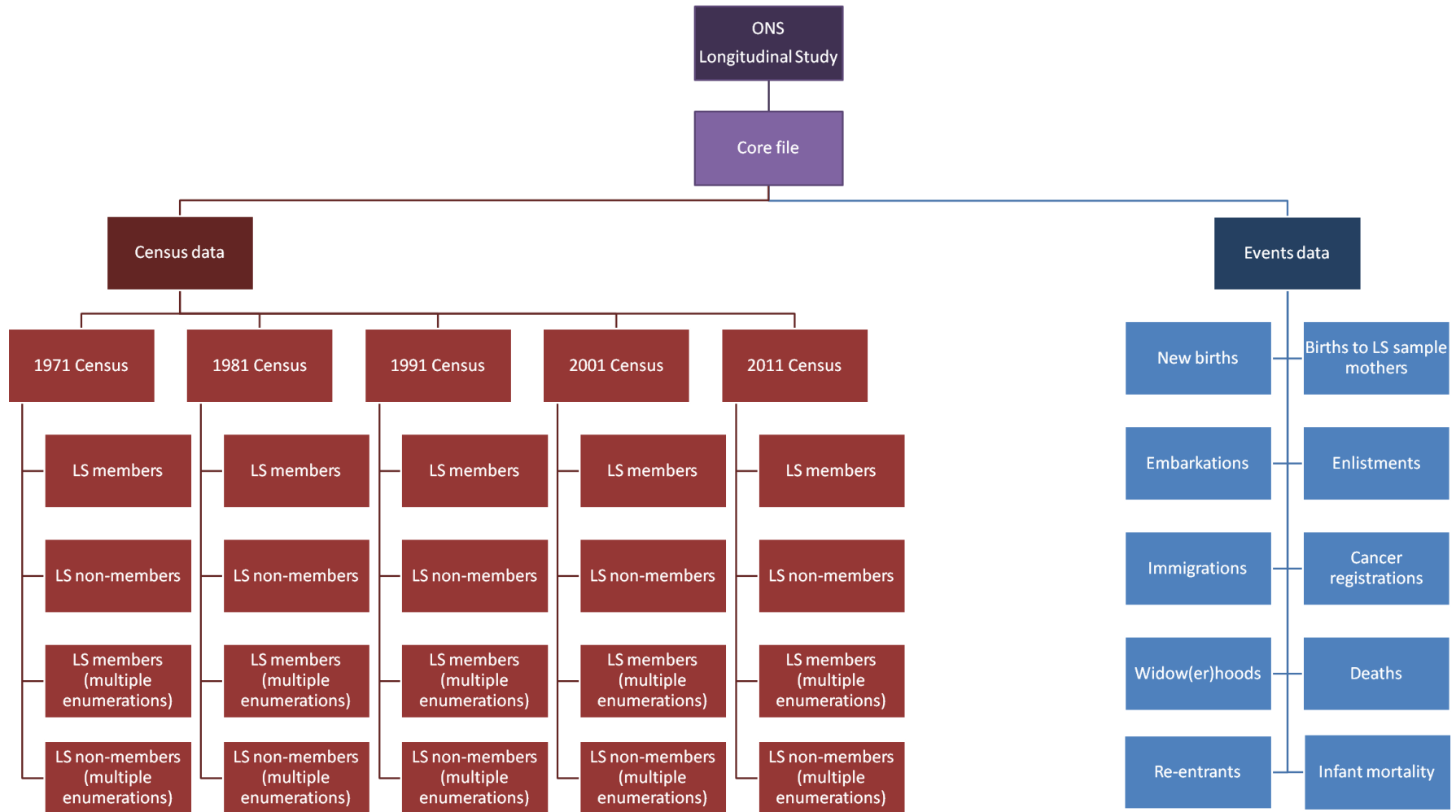
¹ Neither students enumerated at their family address nor visitors completed a full census questionnaire. In addition a small number of LS members' questionnaires were insufficiently completed to distinguish between usual and short-term residents.

² Non-member multiple enumerations are only identified where the LS member themselves has a multiple enumeration (non-member information is only held where they are resident in the same household as the LS member).

The 2011 Census data along with all the other LS data sets available for research are held within the secure Virtual Microdata Laboratory (VML)⁴ at ONS.

A diagram describing all the available data in the LS is shown in Figure 2.

Figure 2: Structure of the Longitudinal Study research database



4. The 2011 Census

Every census is slightly different from the last. A new census gives the opportunity to update the way a census is conducted to reflect changes in society, user requirements, statistical methodology and technology.

The 2011 Census was no different and a number of innovations and changes were made to the way the census was conducted compared to 2001.

To help understand the census data in the LS, this next section highlights some of the main features of the 2011 Census and some of the differences between the 2001 and 2011 Censuses. More information on the 2011 Census is available on the [ONS website](#).

To help with some of the terms in this section a glossary⁵ of terms used in the 2011 Census is available.

Census enumeration

The 2011 Census for England and Wales was the first to post out all household questionnaires⁶ based on a specially developed national address register⁷. For other types of accommodation, including communal establishments, special enumerators delivered questionnaires by hand. Household and communal establishment residents could either complete their census questionnaire on paper or for the first time, or by internet. Overall 16 % of questionnaire returns were made by internet in 2011⁸.

All questionnaires (online or paper) were tracked using a questionnaire tracking system. This allowed targeted field follow-up by census collectors to identify and follow up households which had not returned a questionnaire.

Population base for enumeration

The enumeration base for the 2011 Census was usual residents, census short-term UK residents plus visitors⁹. For the purposes of enumeration, the agreed definitions stated that a full census return should be completed by:

- anyone who has stayed or intends to stay in the UK for three months or more
- anyone outside the UK for less than 12 months

Usual residents

For the purposes of defining individuals for census outputs and population estimates, a usual resident of the UK was anyone who, on 27 March 2011:

- was in the UK and had stayed or intended to stay in the UK for a period of 12 months or more
- had a permanent UK address and was outside the UK and intended to be outside the UK for less than 12 months

Short-term UK resident

A census short-term UK resident was anyone born outside the UK who had stayed or intended to stay in the UK for a period of three months or more but less than 12 months.

Visitors

The 2011 Census defined two types of visitor.

- A domestic visitor - a person staying overnight at an address on 27 March 2011 at which they were not usually resident. They were usually resident in the UK
- An international visitor - a person who intended to stay in the UK for less than three months. They were usually resident outside the UK

Domestic visitors were asked to complete a census questionnaire for their usual residence when they returned to that address.

2011 Census questions

The 2011 Census paper and internet questionnaires were identical except where the internet questionnaire automatically filtered out questions which were not applicable to individuals.

Similar to the 2001 Census, the 2011 Census questionnaire asked questions related to the household and the individual. In 2011, there were 15 questions related to the household.

- Seven were unchanged since 2001
- Three were new for 2011 (number of bedrooms; type of central heating; and visitor details). Three household questions from 2001 had been dropped (whether or not have central heating; whether or not have bath/toilet access; and lowest floor level lived on)
- Two were guidance questions (to help individuals work out whom to include in the questionnaire). These were equivalent to the guidance notes from 2001. However, two additional questions were simple counts of the number of usual residents and visitors)
- One (the household relationships question) was a modification of the question asked in 2001.

In 2011, there were 43 questions related to the individual

- 20 were unchanged since 2001
- Nine were modifications of the questions asked in 2001
- Eight were new for 2011 (second residence and type of second residence; main language and English language proficiency; month/year of entry into UK; intended length of stay in UK; passports held (as a proxy for citizenship); and national identity)
- Six were routing questions (to help individuals omit the questions which did not apply to them)
- Two questions from 2001 were dropped (number of employees at workplace, and professional qualifications)
- ONS published a report in December 2012 on the comparability of the 2001 and 2011 Census questions entitled '2011-2001 Census in England and Wales Questionnaire Comparability'¹⁰. More detail on the comparability of the census questions is given in this report

A summary of the topics included as questions on the census since 1971 and therefore included in the LS database is given in Appendix A

Other differences

Edit and imputation methodology

The purpose of edit and imputation is to resolve non-responses and inconsistencies within the data obtained from completed or partially completed questionnaires. It did not address complete non-response where no questionnaire was returned. There were two main reasons for applying item editing and imputation. Firstly, it was desirable to provide a complete and consistent database for future research and analysis. Secondly, it provided an opportunity to adjust for non-response and inconsistencies, which can lead to bias or inconsistency in resulting estimates.

As in 2001, users of the LS are still able to see when a variable has been imputed for the variables that are collected on the 2011 census questionnaire. However the methodology underpinning this has changed since 2001.

More details on the process used for edit and imputation in the 2011 Census is available in the report '2011 Census Item Edit and Imputation Process'¹¹.

Classifications

The classifications, that is the full set of values that a variable can have, used in the 2011 Census have been reviewed and updated to reflect changes that have taken place since 2001.

Details of the classifications used in the 2011 Census are available on the [ONS website](#).

Census geography

The main geographies directly associated with a census are output areas (OA) and super output areas (SOA). Output areas are the base unit for census data releases. There have been some changes to census geography for 2011 compared to 2001.

All 2011 Census geographies are contained in the LS database. However, geography lower than local authority level is restricted and only available to researchers under certain conditions where these lower levels of geography can be used to derive or link other data. There is a greater risk of disclosure of individual records at low levels of geography and restricting access to these variables helps to reduce this risk.

More information on census geography is available on the [ONS website](#).

Over time geographical classifications are updated and changed. The LS research database contains geography variables with classifications from 1971 to the present day. As result when comparing geography over time there may be differences in the classifications. The best resource for looking up geography classifications can be found in the [ONS Geography Portal](#).

Census quality

The overall quality of the census information contained in the LS database is largely determined by the quality of the 2011 Census itself. The 2011 Census was highly successful and the quality of the statistics produced was of a high level.

The main measures of quality which have an impact on the research potential of the LS are: census response rates, the accuracy of individual responses to the census questions, data capture accuracy, and editing and item imputation rates. Each of these is discussed below.

Response rates

The overall response rate for the 2011 Census in England and Wales was 94.0%. This is comparable with the 2001 Census which also had a 94.0% response rate. Overall, the quality of response rates in 2011 were higher than in 2001; response rates for local authorities varied between 64.0% and 99.0% in 2001 compared with a smaller variation of 82.0% to 98.0% in 2011.

Achieving a response rate of 80.0% in all local authorities was one of the Key Performance Indicators for the 2011 Census.

Detailed response rates are available in the report 'Response Rates in the 2011 Census'¹².

Census Quality Survey

The Census Quality Survey (CQS) ran soon after the census to assess the quality of responses to the census questionnaire. It was a voluntary survey, independent of the census.

It aimed to measure the accuracy of answers given to census questions by asking a sample of households the census questions in a face-to-face interview. By comparing the responses given in

the CQS to those given in the census, agreement rates were calculated which provided an indication of how accurately the 2011 Census questionnaire had been completed by the general public.

The CQS found that agreement rates ranged from 55.0% for the question which asked the 'year that you last worked' to 99.8% for the question 'last week, were you waiting to start a job already obtained?' The extent of agreement varied depending on the type of question: whether it was objective or subjective, whether it was answered by tick box or free text, and how many response categories it had. Of the questions asked in the CQS, almost three-fifths achieved agreement rates of 90.0% and over, and almost three-quarters achieved over 85.0%.

More detail can found in the full 'Census Quality Survey Report'¹³.

Data capture, coding and cleaning

Data capture, coding and cleaning is all about turning census returns into data that can be used to produce statistics.

To ensure the quality of census statistics it is important that the entries individuals have made on their census questionnaires are accurately captured when they are automatically scanned. All of the targets set for 2011 Census data capture and coding accuracy were exceeded.

The report 'Data capture, coding and cleaning'¹⁴ contains the full accuracy results.

Edit and item imputation

For the household questions on the 2011 Census, the imputation rates (which include imputation for non-response and edits) ranged from 2.3% of responses for question H14 on the questionnaire relating to the number of cars or vans, to 3.6% for question H11 for central heating.

For the individual questions, the imputation rates ranged from 0.5% for question 2 relating to sex to 14.3% for question 40 relating to the postcode in the workplace address question.

Question 20 on religion had 0% imputation because it was a voluntary question but it had a non-response rate of 7.2%.

The full edit and imputation rates for census questions are available in 'Item non-response, editing and imputation rates - Updated 26 July 2013'¹⁵ and an evaluation of the edit and imputation process is available in the report 'Evaluation - Item edit and imputation'¹⁶.

Census corrections

The census is however, a complex data set and occasionally errors are identified as the data is published and used.

An [issues and correction page](#) is provided for the 2011 Census.

5. Tracing rates

Tracing rates definition

Tracing of the 1% LS sample from the 2011 Census extract took place at HSCIC in the year following census. An LS member is deemed to be traced if a record for them can be found on NHS registers. Linkage of life event data can only take place if the LS member has been traced. Records will be untraced if there are inconsistencies in the data recorded at census (for example, an incorrect date of birth) or if they have not registered with the NHS.

The tracing rate is expressed as:

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

Alternatively we can examine the untraced rate, which 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 are calculated for all individuals enumerated in the 2011 Census. This means whether they were enumerated as:

- a usual resident
- a short-term resident (2011 Census)
- a student only enumerated at their family address
- only as a visitor (2011 Census)

This is slightly different from the tracing rates published for the 2001 Census as short-term residents and visitors were not enumerated in 2001.

Tracing rates are calculated based on the number of LS members traced at the time of the 2011 Census. Those that are untraced at census can become traced at a later date.

Tracing rate tables

Full tables for tracing rates are available to download in the [‘Quality of tracing at the 2011 Census’](#)¹⁷ and show tracing rates for:

- age and sex
- marital and civil partnership status
- economic activity
- National Statistics Socio-economic Classification (NS-SEC)

- country and region
- country of birth
- year of arrival
- ethnic group
- communal establishments

Tracing rates summary

The next section gives a summary of the findings from the traced and untraced rate tables for the 2011 Census.

Table 2 shows that overall tracing rates have improved at each census since 1971. At earlier censuses the tracing exercise was purely a clerical process but since 2001 tracing has been carried out using a combination of automated and clerical matching.

Compared to the tracing rate for the 2001 Census, there has been a slight decrease in the tracing rate in 2011. This can be explained by the inclusion of short-term residents and visitors in the 2011 Census, and taking the LS extract of data from census at an earlier point in census processing (compared to 2001), before any poor quality records had been filtered out.

Table 2: LS Tracing rates, 1971-2011

Census	Tracing Rates (%)		
	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

¹⁸ The quality of tracing from previous censuses is available on the ONS website.

In Table 3, the tracing rate has been split out for the different resident types in 2011. Here you can see that the tracing rate is highest for usual residents at 99.0%. Short-term residents have a much lower tracing rate of 50.4%. Many of these recent migrants to England and Wales would not have registered with a GP before the census which contributes to the lower tracing rate for this group.

Table 3: LS tracing rates by resident type in 2011

Resident type	Number traced	Number untraced	Total	Tracing rate (%)
All LS members enumerated at the 2011 Census	581,362	8,634	589,996	98.5
<i>of which:</i>				
- usual residents	575,678	6,013	581,691	99.0
- short-term residents	931	917	1,848	50.4
- student only enumerated at their family address (not at their term-time usual residence)	2,125	107	2,232	95.2
- enumerated but unable to determine if a usual resident or a short-term resident	87	42	129	67.4
- visitors	2,541	1,555	4,096	62.0

In the majority of cases LS members are traced at HSCIC. As result, the next section focuses on those that are untraced and describes the untraced rate for different demographics. These are calculated for all LS members regardless of their resident type in 2011.

Age and sex

Young adults aged from 20 to 39 years of both sexes were the most likely to be untraced with 5.3% of males aged from 25 to 29 years untraced, and 3.6% of females aged from 20 to 24 years untraced. Highest tracing rates were found in older age groups with only 0.2% of both males and females aged 75 and over untraced. Overall, females were more likely to be traced (0.9% untraced) than males (1.5% untraced).

Marital and civil partnership status

For both males and females, never married individuals aged 16 and over were the least likely to be traced with 1.9% of females and 3.2% of males untraced. For both sexes, widows and surviving partners in civil partnerships were most likely to be traced with only 0.4% of females and 0.5% of males untraced.

Economic position

Retirees of both sexes were most likely to be traced with only 0.2% of both males and females untraced. The highest untraced rates were found among students where 3.1% of females and 5.1% of males were untraced. Among the economically active, those in employment were most likely to be traced compared to the unemployed (1% of employed females were untraced compared to 1.5% of unemployed females, while 1.7% of employed males were untraced compared to 1.8% of unemployed males).

National Statistics Socio-economic Classification (NS-SEC)

Again those classified as students had the highest tracing failure rates with 3.0% of females and 5.1% of males untraced. Those classified as long term unemployed had the next highest rates with

1.8% of females and 2.7% of males untraced. Among those with a classification based on occupation, 0.7% of females and 1.4% of males were untraced. For both sexes, those in routine occupations had the highest untraced rates (1.2% for females and 1.7% for males). The lowest untraced rate for females was for those in intermediate occupations (0.4%) and for males those in lower supervisory roles (1.2%).

Country or region

LS members enumerated in London were most likely to be untraced with failure rates of 2.9% for females and 4.4% for males. This region alone accounted for 44.0% of all untraced cases but only 15.0% of the total enumerated. Outside of London, the East of England and the South East had higher untraced rates for both genders with lowest rates found in the North East (0.4% female and 0.8% male) and Wales (0.4% female and 0.6% male).

Country of birth

Overall those born in England and Wales were most likely to be traced with untraced rates of 0.1% for females and 0.2% for males. Of those untraced 84.6% were non-UK born. Highest untraced rates were for those born in the USA (14.6% for females and 21.6% for males) and in the EU accession states (11.8% for females and 23.3% for males). Migrants from the accession states made up only 1.7% of all females and 1.5% of all males, but accounted for 21.7% of untraced females and 22.8% of untraced males. For all countries the untraced rate for females was lower than for males.

Year of arrival

For both male and female non-UK born LS members there is a clear trend showing recent arrivals were less likely to be traced. For females who arrived before 2001 the untraced rate was 0.8% rising to 28.7% for those arriving in 2010 and 57.8% for those arriving in 2011. For males 1.3% of those arriving pre-2001 were untraced with 35.9% for 2010 and 56.4% for 2011.

Ethnic group

Highest untraced rates were found for both sexes in the Other White ethnic group (15.2% for males and 8.6% for females) reflecting the impact of new migrants from the EU accession states and those born in the United States. This group made up 43.0% of all females and 45.0% of males untraced. Untraced rates were also high among the Chinese ethnic group with 8.5% of females and 10.9% of males untraced. Across all ethnic groups females were more likely to be traced than males.

Communal establishments

Overall LS members enumerated in communal establishments were less likely to be traced than those enumerated in private households (6.1% were untraced, compared to 1.1%). The untraced rate for males in communal establishments (8.2%) was nearly twice that for females (4.2%). Highest untraced rates were found for those in enumerated in defence establishments (15.7% for males and 13.8% for females). However it should be noted that this group was relatively small (54 untraced cases) and that 77.0% of these untraced cases were born outside of England and Wales. Individuals enumerated in educational establishments (boarding schools and university halls of

residence) accounted for 63.0% of those untraced in communal establishments with 10.2% of males and 8.3% of females untraced.

6. Sampling fractions

Sampling fractions definition

The LS was designed as a 1% sample of the enumerated census population of England and Wales. When a new LS sample is taken at a census, it is essential to measure how representative the sample is and what biases exist in it.

Sampling fractions are used to measure whether the LS population (both as a whole and in subgroups) represents approximately 1% of the originating census population (or specific subgroups of that population). Because of variations in tracing rates and in the availability of accurate birth date data (the sampling criterion) the sample proportions may vary. In most cases, only the traced LS population is used in analysis, therefore the net sampling fractions rather than the gross sampling fractions are used as these exclude the untraced population.

The **observed** net sampling fraction α_i (where i identifies the subgroup of interest) is defined as:

$$\alpha_i = \frac{\text{number in subgroup } i \text{ (traced LS population)}}{\text{number in subgroup in census population}} \times 100\%$$

The **expected** net sampling fraction gives an estimate of the proportions of the LS sample that would have been expected in the census population and allows comparison with what was actually observed. Ideally, observed and expected net sampling fractions should be the same, but biases, not only in tracing rates but also in the sampling of certain population subgroups can distort that relationship.

Expected net sampling fractions are defined as:

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

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

We would expect the observed net sampling fraction to equal 1.1% because the LS sample is taken from four dates of birth. This is shown in the calculation below which takes account of a leap year every four years:

$$\frac{(4+4+4+4)}{(365+365+365+366)} \times 100\% = 1.1\% \text{ of the population (rounded to one decimal place).}$$

Sampling fractions are calculated for usual residents only.

Sampling fractions tables

Full tables for sampling fractions are available to download in the '[Longitudinal Study sampling fractions, 2011](#)'¹⁹ and show sampling fractions for:

- sex and age
- marital and civil partnership status
- National Statistics – Socio-economic Classification
- country of birth
- religion
- ethnic group
- region

Sampling fractions summary

The next section gives a summary of the observed and expected sampling fractions for a number of different categories. It also shows sampling fractions based upon the census count and the census estimate. The census count is simply the number of people who were enumerated on a 2011 census questionnaire. The census estimate is the census count plus an estimate of the number of people who did not complete a questionnaire plus an adjustment for over enumeration. It is the census estimate that is published as census statistics.

Overall

Table 4 below shows the observed sampling fractions for each census since 1971. It shows that sampling fractions for the 2011 Census are similar to 2001 for males, females and overall.

Table 4: Observed sampling fractions by census

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

²⁰ Sampling quality for the 1971 to 2001 Censuses is available on the ONS website.

Age and sex

Sampling fractions for males and females by age can be seen in Tables 5 and 6. The sampling fractions for both sexes are broadly comparable at all ages except for males from ages 15 to 34 years, which show a small dip compared to females of the same age, although the sampling fraction is still representative of a 1% sample of census counts. These age groups are typically the hardest to enumerate at census.

Table 5: 2011 LS Sampling fractions for traced LS members by sex and age - Males

Age group	Number of usual residents			Sampling fraction (%) based on			
	2011 traced LS sample	2011 Census count	2011 Census estimate	2011 Census count		2011 Census estimate	
				Observed	Expected	Observed	Expected
0-4	17,432	1,618,119	1,789,744	1.08	1.11	0.97	1.03
5-9	16,138	1,477,675	1,604,914	1.09	1.11	1.01	1.03
10-14	16,655	1,557,333	1,668,191	1.07	1.11	1.00	1.03
15-19	17,750	1,669,399	1,808,009	1.06	1.11	0.98	1.03
20-24	17,387	1,673,775	1,919,169	1.04	1.07	0.91	1.00
25-29	17,424	1,637,558	1,914,298	1.06	1.07	0.91	0.99
30-34	17,555	1,616,514	1,842,456	1.09	1.08	0.95	1.00
35-39	18,903	1,689,079	1,855,920	1.12	1.10	1.02	1.02
40-44	20,398	1,889,564	2,027,932	1.08	1.10	1.01	1.02
45-49	21,192	1,923,089	2,028,286	1.10	1.11	1.04	1.03
50-54	19,081	1,716,171	1,786,690	1.11	1.11	1.07	1.03
55-59	17,012	1,519,691	1,573,599	1.12	1.11	1.08	1.03
60-64	17,487	1,614,611	1,658,007	1.08	1.12	1.05	1.04
65-69	14,041	1,266,439	1,299,081	1.11	1.12	1.08	1.04
70-74	11,283	1,007,446	1,032,457	1.12	1.12	1.09	1.04
75 and over	18,973	1,717,122	1,764,623	1.10	1.12	1.08	1.04
All people	278,711	25,593,585	27,573,376	1.09	1.10	1.01	1.02

Table 6: 2011 LS Sampling fractions for traced LS members by sex and age - Females

Age group	Number of usual residents			Sampling fraction (%) based on			
	2011 traced LS sample	2011 Census count	2011 Census estimate	2011 Census count		2011 Census estimate	
				Observed	Expected	Observed	Expected
0-4	16,712	1,543,864	1,707,006	1.08	1.11	0.98	1.05
5-9	15,607	1,408,553	1,530,797	1.11	1.11	1.02	1.05
10-14	15,923	1,484,819	1,590,486	1.07	1.11	1.00	1.05
15-19	17,359	1,603,239	1,731,376	1.08	1.11	1.00	1.05
20-24	18,577	1,719,628	1,888,076	1.08	1.08	0.98	1.03
25-29	19,343	1,774,698	1,922,311	1.09	1.09	1.01	1.03
30-34	19,344	1,729,905	1,841,459	1.12	1.10	1.05	1.04
35-39	20,157	1,780,857	1,876,241	1.13	1.11	1.07	1.05
40-44	21,835	1,987,048	2,071,157	1.10	1.11	1.05	1.05
45-49	22,076	2,001,961	2,072,240	1.10	1.11	1.07	1.05
50-54	19,684	1,765,562	1,815,004	1.11	1.11	1.08	1.05
55-59	17,397	1,573,641	1,610,316	1.11	1.11	1.08	1.05
60-64	18,508	1,682,401	1,719,155	1.10	1.11	1.08	1.06
65-69	14,812	1,344,872	1,375,080	1.10	1.11	1.08	1.06
70-74	12,165	1,118,062	1,146,215	1.09	1.11	1.06	1.06
75 and over	27,468	2,525,013	2,605,617	1.09	1.11	1.05	1.06
All people	296,967	27,044,123	28,502,536	1.10	1.11	1.04	1.05

Ethnic group

Table 7 shows the sampling fractions by ethnic group. The observed sampling fractions based on the census count are in line with what would be expected (1%) for all ethnic groups except for the Indian, Pakistani and Bangladeshi ethnic groups. Here the sampling fractions are higher at 1.4%, 1.3% and 1.5% respectively. This indicates that there appears to be some slight over sampling in these groups. Similar results to this have been observed in LS samples from previous census years. Users should consider this finding when analysing ethnic groups.

Table 7: 2011 LS Sampling fractions by ethnic group

Ethnic group	Usual residents			Sampling fraction (%) based on			
	2011 traced LS sample	2011 Census count	2011 Census estimate	2011 Census count		2011 Census estimate	
				Observed	Expected	Observed	Expected
White	490,909	45,732,848	48,209,395	1.07	1.11	1.02	1.04
English/Welsh/Scottish/Northern Irish/British	462,246	42,938,894	45,134,686	1.08	1.11	1.02	1.05
Irish	5,214	499,301	531,087	1.04	1.09	0.98	1.02
Gypsy or Irish Traveller	528	51,948	57,680	1.02	1.07	0.92	1.01
Other White	22,921	2,242,705	2,485,942	1.02	1.00	0.92	0.94
Mixed/multiple ethnic groups	11,085	1,025,828	1,224,400	1.08	1.10	0.91	1.03
White and Black Caribbean	3,805	355,790	426,715	1.07	1.11	0.89	1.04
White and Black African	1,490	137,447	165,974	1.08	1.09	0.90	1.02
White and Asian	3,144	291,278	341,727	1.08	1.10	0.92	1.03
Other Mixed	2,646	241,313	289,984	1.10	1.09	0.91	1.02
Asian/Asian British	50,028	3,845,490	4,213,531	1.30	1.09	1.19	1.02
Indian	18,035	1,331,757	1,412,958	1.35	1.09	1.28	1.02
Pakistani	13,885	1,051,325	1,124,511	1.32	1.10	1.23	1.03
Bangladeshi	6,129	413,958	447,201	1.48	1.10	1.37	1.03
Chinese	3,623	332,477	393,141	1.09	1.03	0.92	0.96
Other Asian	8,356	715,973	835,720	1.17	1.07	1.00	1.01
Black/African/Caribbean/Black British	18,631	1,613,860	1,864,890	1.15	1.09	1.00	1.02
African	10,366	872,579	989,628	1.19	1.08	1.05	1.02
Caribbean	5,972	547,071	594,825	1.09	1.11	1.00	1.04
Other Black	2,293	194,210	280,437	1.18	1.09	0.82	1.03
Other ethnic group	5,025	419,682	563,696	1.20	1.07	0.89	1.01
Arab	1,969	166,967	230,600	1.18	1.07	0.85	1.00
Any other ethnic group	3,056	252,715	333,096	1.21	1.07	0.92	1.01
All people	575,678	52,637,708	56,075,912	1.09	1.11	1.03	1.04

Country and region

Table 8 shows sampling fractions for England, Wales and the English regions. Overall the sampling fractions are as expected and show no bias in the proportion selected from either country or the English regions.

Table 8: 2011 LS sampling fractions by country or region

Country or region	Usual residents			Sampling fraction (%) based on			
	2011 traced LS sample	2011 Census count	2011 Census estimate	2011 Census count		2011 Census estimate	
				Observed	Expected	Observed	Expected
England	544,920	49,781,971	53,012,456	1.09	1.10	1.03	1.04
North East	26,248	2,448,504	2,596,886	1.07	1.11	1.01	1.04
North West	72,000	6,612,432	7,052,177	1.09	1.11	1.02	1.04
Yorkshire and The Humber	54,435	4,947,933	5,283,733	1.10	1.11	1.03	1.04
East Midlands	46,685	4,309,190	4,533,222	1.08	1.11	1.03	1.04
West Midlands	58,378	5,261,955	5,601,847	1.11	1.11	1.04	1.04
East of England	60,270	5,560,757	5,846,965	1.08	1.11	1.03	1.04
London	82,772	7,391,858	8,173,941	1.12	1.08	1.01	1.02
South East	89,394	8,190,285	8,634,750	1.09	1.11	1.04	1.04
South West	54,738	5,059,057	5,288,935	1.08	1.11	1.03	1.04
Wales	30,758	2,855,737	3,063,456	1.08	1.11	1.00	1.04
All people	575,678	52,637,708	56,075,912	1.09	1.11	1.03	1.04

7. Linkage rates

Forward linkage rates definition

Forward linkage rates measure the percentage of LS members present at a census who were also present at the following census, excluding those who have died or embarked (embarked means they have notified their GP surgery that they are moving out of England and Wales) in the intercensal period. It is expressed as:

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

Where '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 prior to the second census 'r_i' is the number of surviving traced LS members not found at the second census.

A 'linkage failure' is said to have occurred if a LS member was not enumerated at the second census, having been enumerated in the previous census, and they do not have a recorded embarkation or death. Linkage failure can occur because of:

- unreported embarkations
- census under enumerations (people don't respond to the census)
- discrepancies in date of birth quoted at either census

It is difficult to estimate the relative contribution of each of these factors to linkage failure. However a number of linkage rate tables have been produced to show the association of linkage rate and socio-demographic factors at census.

Linkage rates are calculated for all individuals enumerated in the census. This means whether they are enumerated as:

- a usual resident
- a short-term resident (2011 Census)
- a student only enumerated at their family address
- only as a visitor (2011 Census)

Forward linkage rate tables

Full tables for linkage rates are available to download in '[Quality of linkage 2001 to 2011](#)'²¹ and show linkage rates for the 2001 and 2011 Censuses by:

- age and sex

- marital status
- household position
- country of birth
- tenure
- highest qualification
- economic activity
- National Statistics Socio-economic Classification (NS-SEC)
- country and region
- local authority

Linkage rates are also available for people born since the 2001 Census and linked to the 2011 Census by:

- mother's and father's age
- type of registration
- birth parity
- year of birth
- country and region of mother's residence
- mother's and father's country of birth
- mother's and father's NS-SEC

As well as, people immigrating to England and Wales since the 2001 Census and linked to the 2011 Census by:

- age and sex
- year of entry
- country and region at entry

Forward linkage summary – 2001 to 2011 Census

The next section gives a summary of the findings from the forward linkage rate tables for the 2001 to 2011 Censuses.

Overall forward linkage rates for by census

Detailed forward linkage rates for previous censuses are available on the Longitudinal Study section of the ONS website²². Table 9 summarises the forward linkage rates for the 1971 to 2011 Censuses.

Table 9: Forward linkage rates for traced LS members, 1971 to 2011 Censuses

Censuses	Linkage rate (%)
1971 to 1981	91.3
1981 to 1991	90.1
1991 to 2001	88.0
2001 to 2011	87.7

Overall forward linkage rates for 2001 to 2011 Censuses

As shown in Table 10, at the 2001 Census, 540,068 LS members were enumerated in England and Wales, of whom 536,565 (99.3%) were successfully traced at the NHS Central Register (NHSCR). During the decade 52,500 traced members died and a further 3,033 members embarked leaving 481,032 expected at the 2011 Census of whom 421,795 were enumerated.

The 'not found at 2011' represent traced LS members who either did not respond to the 2011 Census, or have embarked (emigrated from England and Wales) but not cancelled their GP registration. Therefore there is no record of their embarkation in the LS. The overall linkage rate for 2001 to 2011 is 87.7%.

Table 10: Overall forward linkage rates

	Number
2001 Census – Traced LS members	536,565
Less deaths prior to 2011 Census	52,500
Less embarkations prior to 2011 Census	3,033
Eligible to be in 2011 Census	481,032
Not found at 2011 Census	59,237
Enumerated at 2011 Census	421,795
Linkage rate (%)	87.7

Age and sex

The lowest linkage rates occur in young adults and at the oldest ages for both sexes. The lowest linkage rates for males were in the age groups (age as at 2001 Census) 25 to 29, 30 to 34 and 95 and over (these were 77.6%, 78.5% and 78.3% respectively). The corresponding rates for females were 84.6%, 85.2% and 80.6% respectively.

Those in the young adult age ranges tend to be more mobile and therefore thus linkage failure may be due to unreported embarkations or failure to complete a census form. Linkage failure at older age groups can be associated with late registered mortality and possible inaccurate date of birth data for those enumerated in communal establishments.

Marital status

Never married individuals at 2001 were the least likely to be linked at 2011 (84.6%) and this was also the lowest rate among females (86.6%). However for males those separated were the least likely to be linked (82.1%). Married individuals were the most likely (90.6% males and 91.5% females) to be linked with a slightly lower rate for those who were remarried.

Household position

Household position was derived using minimal household units. Minimal household units were derived from household composition and were described as the smallest unit or group within a household expected to share a similar lifestyle (See LS User Guide 20²³). Dependent children are grouped with their parents and non-dependent children are regarded as adults not in a family.

Within private households those enumerated as dependent children in 2001 were least likely to be linked at 2011. As dependent children are those aged under 16 or 18 years if in full time education they fall in to the 10 to 28 year age groups in 2011. Overall dependent children with a single parent

in 2001 were less likely to be linked than those with a couple (79.8% compared to 87.8%). Married couples were most likely to be linked with a slightly higher rate for those without dependent children (91.3% as opposed to 91.0%). The linkage rate for adults not living in a family was 84.4%, dropping to 81.8% for males.

For those enumerated in communal establishments the linkage rate falls to 69.3% (72.7% for females and 66.8% for males). However this group accounts for only 1.8% of the overall sample and therefore has minimal impact on the overall linkage rate. Of those enumerated in communal establishments, 31.0% were recorded as students in full time education in 2001.

Country of birth

Linkage rates were highest among the UK born for both sexes with 90.3% of females and 87.4% of males successfully linked. There was little difference between rates for England or Wales. However rates for those born in Northern Ireland were lower at 88.0% for females and 83.3% for males and also lower for Scotland at 83.0% for females and 81.2% for males.

For those born outside the UK, the rate is lower again at 78.5% for females and 75.7% for males. For those born in the Indian sub continent rates were relatively high at 85.0% for India, 82.0% for Pakistan and 83.4% for Bangladesh. In contrast to all other groups the rate for Pakistani born females was lower than that for males (81.0% compared to 82.8%). The lowest rates were for those born in Oceania or the Americas (62.2% and 69.5% respectively). Within the EU countries (as at 2001) the overall rate was 72.2%. Those born in Germany accounted for 36.0% of the EU born and had the highest linkage rate at 80.2%.

Ethnic group

The highest linkage rates were achieved within the White British group (90.5% for females and 87.7% for males). Lowest linkage rates were among those classed as Other White or Other. Both of these groups tended to include younger adults (33.0% of Other White and Other were aged from 25 to 34 years), and a higher proportion of students (10.0% for Other White and 14.0% for Other). Across all ethnic groups females were more likely to be linked than males.

Housing tenure

Linkage rates were lowest for those enumerated in a communal establishment in 2001, at 69.2%. Among individuals in private households, the highest linkage rate was for those who own their home outright with 90.2% found at 2011. For non-owner occupiers, those in the home rented from the local authority and those renting from a housing association had linkage rates of 83.6% and 83.3% respectively. This was higher than those in privately rented accommodation which had a linkage rate of 78.4%.

Highest level of qualification

There was little variation in linkage rates by qualification. However for both sexes those with Level 3 qualifications (A level and equivalent) had the lowest rates (87.9% for females and 85.0% for

males). This can be seen as age dependent as 49.2% of those with Level 3 qualifications were aged from 15 to 29 years in 2001.

Economic position

The lowest linkage rates for both males and females classified as economically activeⁱ, were the unemployed (84.3% females and 75.9% males). Employees, both full-time and part-time, were slightly more likely to be linked than the self employed. Males employed full-time were more likely to be linked than part-time employees (88.3% compared to 85.6%). However this position was reversed for females, with 90.3% of those in full-time employment linked and 92.6% of those in part-time employment.

Economically inactive students were the least likely to be linked (82.8% of females and 76.3% of males) and those retired most likely. However there is a slight reversal among retirees with 90.9% of females linked and 91.4% of males. This could be influenced by lower linkage rates at very old ages.

National Statistics Socio-economic Classification (NS-SEC)

Excluding the full time student group, the lowest linkage rate for both males and females was for those who had never worked or were long term unemployed (74.8% for males and 83.1% for females). For those with a classification based on their current or previous employment, the lowest rates were among males in the routine and semi routine groups (85.9% and 85.7% respectively). In contrast for females, the lowest rates were recorded in the higher managerial and professional group (89.1%).

Country, region and local authority of enumeration

For both males and females linkage rates for those enumerated in London were the lowest (84.4% for females and 81.2% for males). The highest rates were recorded in the East Midlands and the South West and Wales. At a local authority level, 17 of the 20 areas with lowest rates were in London with Kensington & Chelsea lowest at 71.5%. Outer London boroughs tended to have higher linkage rates with Bexley at 88.7%. Outside London the lowest rates were recorded in Manchester (80.8%), Cambridge (81.2%) and Oxford (82.9%). The highest rate was recorded in Kettering (92.6%).

Forward linkage summary – LS members born between the 2001 and 2011 Censuses

Forward linkage rates are also calculated for LS members born between the 2001 and 2011 censuses. These are used to quantify the proportion of those born during this period who were also enumerated at the 2011 Census.

Overall forward linkage rates for by census, 1971 to 2011

ⁱ The economically active are defined as those aged 16 or over, who are either employed or unemployed in the survey reference week.

The economically inactive are defined as those aged 16 or over who are neither in employment nor unemployed. This group includes, for example, all those who are looking after a home or family, have a long-term illness or disability which prevents them working, or are retired.

Detailed forward linkage rates for previous censuses are available on the Longitudinal Study section of the ONS website²⁴. Table 11 summarises the forward linkage rates for traced LS members born between each census since 1971. It shows a continued trend of decreasing linkage success between censuses since 1971.

Table 11: Forward linkage for traced LS members born between censuses, 1971 to 2011

Traced LS members born between censuses	Linkage rate to next census after birth (%)
1971 to 1981	92.8
1981 to 1991	91.3
1991 to 2001	85.7
2001 to 2011	85.1

Overall forward linkage rates for LS members born between 2001 and 2011 Censuses

Table 12 below breaks down the linkage rate at 2011 for those LS members born since the 2001 Census. Overall, the rate at 85.1% was slightly lower than that for LS members enumerated at 2001.

Table 12: Forward linkage for LS members born between 2001 and 2011 Censuses

	Number
Births between 2001 and 2011 censuses – Traced LS Members	72,845
Died prior to 2011 Census	415
Embarked prior to 2011 Census	601
Eligible to be in 2011 Census	71,829
Not found at 2011 Census	10,680
Enumerated at 2011 Census	61,149
Forward linkage rate (%)	85.1

This next section gives a summary of the findings for this group and describes the factors associated with successful linkage at the 2011 Census.

Parents' age at birth

Both mothers' and fathers' age had an association with linkage failure with lowest rates recorded for those with parents aged 19 years and younger (77.1% for mothers and 76.5% for fathers). Highest linkage rates were found for mothers and fathers in their 30s with a slight fall off for older parents.

Registration type

Children born to lone mothers were least likely to be linked with only 74.7% of sole registered births and 78.7% of non-cohabiting joint registration births linked at 2011. Births to married couples had the highest rate at 87.3%, whilst the rate for births to cohabiting couples was 85.2%.

Birth parity

Birth parity is recorded for married mothers only. Highest linkage rates were found for second births (parity 1 at 89.0%). For higher order births (birth parity of 4 and 5 and above), linkage rates were lower (78.4% at parity 4 and 74.3% at parity 5 and above. However higher order births only account for 3.2% of births to sample mothers.

This decrease is likely to be caused partly by a failure to enumerate younger children in large households. The census paper questionnaire had space for six people to be recorded. If there were more than six people in a household, the householder needed to request and complete a household continuation questionnaire for the additional household members. The internet questionnaire had space for 30 people.

An examination of higher order births to LS sample mothers enumerated at the 2011 Census shows that 19.7% of these children were not found in the LS member's household (based on matching dates of birth). Paper returns (as opposed to internet returns) accounted for 51.0% of these enumerations but 81.0% of the cases where the child was not found. For paper enumerations 39.0% had a household of six persons (maximum on household form) but 64.0% of cases where the child was not enumerated had a household size of six.

Year of birth

Linkage rates by year of birth fell across the decade, with 86.8% of 2001 births linked, falling to 84.4% of births in 2010. Given the association between parents' age and census enumeration this is unsurprising.

Country and region of birth

Linkage rates by region of usual address of mother at birth follow the same pattern as census to census enumeration, with children born in London least likely to be linked (79.3%) and those born in the East Midlands most likely (88.0%).

Parent's country of birth

Children born to parents who were also born in England and Wales were most likely to be linked (87.5% for mothers and 88.2% for fathers). Those born to parents from the Irish Republic were least likely to be linked (75.2% for mothers and 71.4% for fathers). There were 15,488 children

born to mothers whose place of birth was outside the UK and Ireland (21.2% of all births) with a linkage rate of 77.0%. These children accounted for 32.0% of all unlinked births.

National Statistics Socio-economic Classification (NS-SEC)

Classification by NS-SEC is based on occupation and employment status information supplied at birth registration. Occupation data was available for mothers in 63.0% of births and for fathers in 88.0% of births. For both mothers and fathers, the highest linkage rates were for those in professional and managerial categories, while for mothers the lowest rate was for the routine category (83.7%). For fathers, the lowest linkage rate was in the small employers/ self employed category (83.8%).

The linkage rate was lower (79.3%) for unclassified mothers. However there is a relationship between these cases and age and type of registration. 38.1% of all mothers were aged under 25 years but 55.5% of the unclassified were aged under 25 years. 15.5% of births were sole registrations or non-resident joint registrations rising to 24.3% for those unclassified. For unclassified fathers, the linkage rate was 75.0%. In sole registrations no data was available for father's occupation.

Forward linkage summary – LS members entering as immigrants between 2001 and 2011 Censuses.

Forward linkage rates are also calculated for LS members immigrating to England and Wales for the first time between the 2001 and 2011 Censuses to quantify the proportion of this group that are also enumerated at the 2011 Census. This next section gives a summary of the findings for this group.

Overall forward linkage rates for LS members entering as immigrants between the 2001 and 2011 Censuses

Table 13 shows the linkage rate at 2011 for those LS members who entered the study as immigrants since the 2001 Census. The overall linkage rate at 44.6% was lower than for those enumerated at 2001 or for new births between the two censuses. However there was an improvement to the linkage rate for immigrants in 2011 compared to 2001 (44.6% in 2011 compared to 33.6% in 2001).

Table 13: LS immigrants forward linkage, 2001 to 2011

	Number
2001 Census – Traced LS members	81,134
Died prior to 2011 Census	245
Embarked prior to 2011 Census	2,735
Eligible to be in 2011 Census	78,154
Not found at 2011 Census	43,270
Enumerated at 2011 Census	34,884
Forward linkage rate (%)	44.6

Age and sex

Female immigrants were more likely to be linked than male immigrants (47.5% compared to 41.8% respectively). For both males and females, linkage rates declined with age with 59.1% of those aged 19 years or younger linked, compared to 28.9% of those aged 70 and over. However, these age groups were in the minority, with most immigrants aged from 20 to 39 years (66.6%). Of these, 45.7% of females and 38.2% of males were linked.

Year of entry

Year of entry to the study is recorded at first registration with the NHS. Linkage rates were higher for those entering in the years immediately before the 2011 Census than for earlier years. For those entering in 2010 the rate was 52.8% compared to 38.7% for 2002. A proportion of those entering the study as immigrants were also likely to leave within a few years, for example, international students. Exit from the study is dependent on the NHS being informed of an embarkation. Therefore low linkage rates for immigrants are typically achieved for immigrants due to unreported embarkations.

Country and region of first NHS registration

For immigrants, region is defined by where they first registered with the NHS. As with linkage from the 2001 Census and new births, the highest linkage rates for both female and male immigrants were recorded in the East Midlands region (52.7% for females and 46.6% for males). The region with the lowest linkage rates was the North East (43.2% for females and 36.9% for males). However, this region accounted for only 2.6% of total immigrations expected at the 2011 Census. In contrast, London accounted for 39.7% of all expected immigrants and the South East for 13.6%. London had a linkage rate of 45.8% for females and 40.0% for males, whilst the rates for the South East were 46.6% for females and 43.2% for males. Together, London and the South East accounted for 54.4% of immigrants not linked at the 2011 Census.

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 intercensal period as new births or immigrants. It is expressed as:

$$\frac{N - t_i - p_i}{N - t_i} * 100\%$$

Where 'N' is the number of traced LS members found at the current census; 't_i' is the number of new entrants (births and immigrants) in the decade preceding the census; and 'p_i' is the number of traced LS members found at the current census but not enumerated at the previous census.

Overall backward linkage rates

As shown in Table 14, at the 2011 Census, 578,821 LS members were enumerated in England and Wales and successfully traced at HSCIC. During the decade 61,149 traced members were born since the 2001 Census and a further 34,884 members were immigrants to England and Wales since the 2001 Census. A further 15,304 were new entrants at the 2011 Census. These are 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 immigration at HSCIC.

A total of 467,484 LS members were therefore expected at the 2011 Census, of whom 421,856 were enumerated at the 2001 Census (45,628 were LS members who were not enumerated ('not found') at the 2001 Census). The overall backward linkage rate for 2011 to 2001 was 90.2%.

Table 14: Backward linkage rate for the traced LS census sample, 2011 to 2001

	Number
2011 Census – Traced LS members	578,821
Born since the 2001 Census	61,149
Immigrant since the 2001 Census	34,884
<i>New Entrants at 2011 Census</i>	15,304
Should have been present at 2001	467,484
Not found at 2001 Census	45,628
Enumerated at 2001 Census	421,856
Backward linkage rate (%)	90.2

Detailed backward linkage rates for previous censuses are available on the Longitudinal Study section of the ONS website²². Table 15 summarises the backward linkage rates for the 1971 to 2011 Censuses. The backward linkage rate for the 2011 to 2001 Censuses is slightly lower than previous censuses.

Table 15: Backward linkage rates for traced LS members, 1971 to 2011 Censuses

Censuses	Linkage rate (%)
1981 to 1971	92.6
1991 to 1981	91.4
2001 to 1991	90.7
2011 to 2001	90.2

8. Conclusion

The ONS Longitudinal Study is a unique data set which draws from census and other administrative data sources to provide a richness and coverage unrivalled by other longitudinal studies. The ONS LS now holds linked data from five successive censuses for England and Wales and has been used in studies on ageing, life course, migration, health, mortality, fertility, deprivation, family formation and inter-generational change.

Tracing of LS records in 2011 was extremely successful with 98.8% of records identified from the 2011 Census traced to an NHS patient record. This was a slight decrease compared to an overall tracing rate of 99.3% at the 2001 Census, but can be explained by: the inclusion of short-term residents and visitors; taking the extract from census earlier than 2001, before they had filtered out any poor quality records; and an increase in immigration in the period prior to the 2011 Census. Many of these recent migrants to England and Wales would not have registered with a GP before the census. Tracing rates have, however, increased since 1991 due to a combination of the use of automated matching methodologies, clerical matching and the availability of more detailed information, including access to census images to aid query resolution.

87.7% of those enumerated and traced at the 2001 Census, after excluding those who were known to have died or emigrated, were linked to the 2011 Census. The proportion not found at the 2011 Census was 12.3%. Linkage failures between the 2001 and 2011 Censuses were often due to unreported embarkations, census non-response and discrepancies in the date of birth quoted at either census. Further analysis of linkage rates by age and sex, marital status, country of birth, tenure, qualifications, economic activity and geographical area has revealed common socio-demographic characteristics affecting non-linkage to the LS in 2011. These factors included:

- being a young adult
- never married or never separated
- being born outside the UK
- belonging to the Other White ethnic group
- living in a communal establishment
- being unemployed or a student
- living in London, particularly Inner London

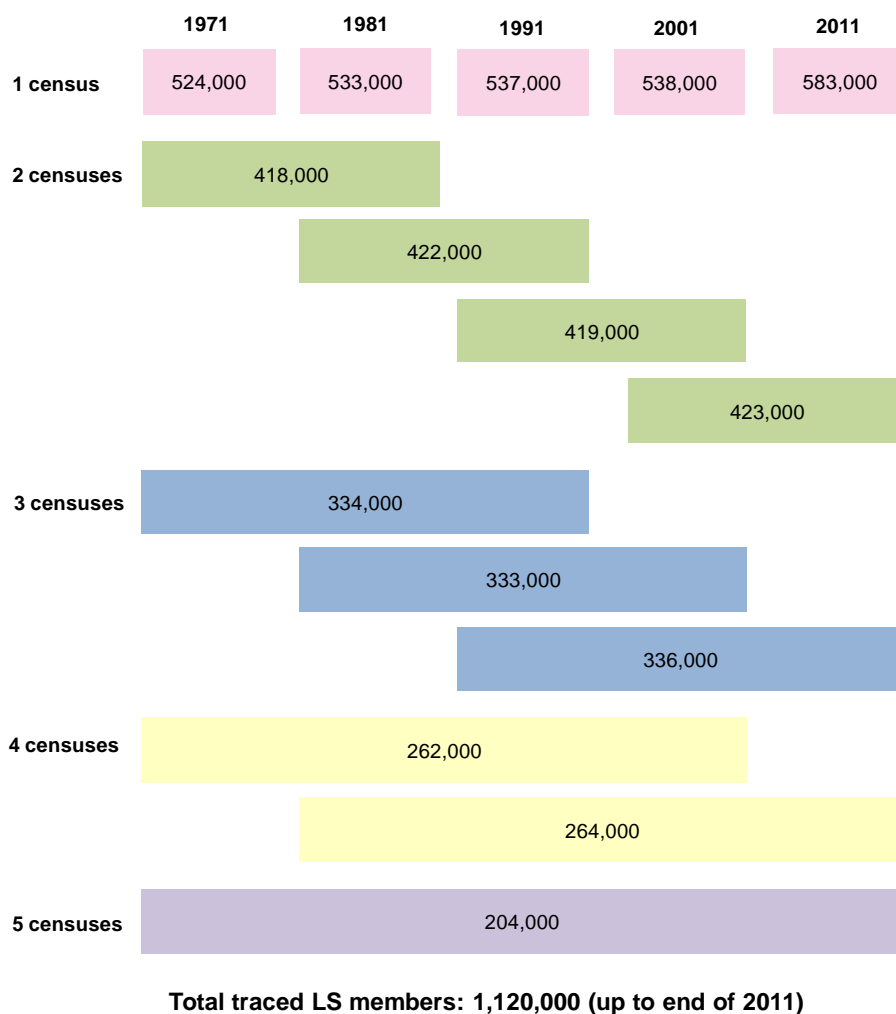
These factors were very similar to those found at the 2001 Census link. Very often these factors are associated with each other rather than acting independently. For example, a non-UK born young adult male living in London is more likely not to be linked. However, it is difficult to measure the relative contribution of each of the above factors to linkage failure. Interestingly these characteristics were associated with lower tracing rates.

There were approximately 111,337 new entrants to the LS between 2001 and 2011, including 61,149 members who were born since the 2001 Census, 34,844 members who migrated to England and Wales since 2001 and 15,304 new entrants at the 2011 Census as their birth date fell on one of the four LS dates of birth.

With the addition of the 2011 Census data, there are now five census points available for analysis in the LS. This means there are sample sizes of over:

- 400,000 traced LS members when looking across two censuses
- 330,000 traced LS members when looking across three censuses
- 250,000 traced LS members when looking over four censuses
- 200,000 traced LS members over five censuses

Figure 3: Sample sizes for traced LS members, 1971 to 2011ⁱⁱ



ⁱⁱ Up to census day, 27 March 2011

Along with linked events data on births, deaths, cancer registrations, embarkations and immigrations, the ONS Longitudinal Study continues to offer a rich data set with large sample sizes that enables social research into various areas, including transitions in socio-economic and demographic characteristics over time, and rates of fertility, mortality and cancer survivorship which can all be analysed alongside census characteristics.

The LS continues to provide useful resource for research from government, academia and the public.

Appendices

Appendix A: Census data - summary of variables by topic area, 1971 to 2011 Censuses

Topics covered	1971	1981	1991	2001	2011
Demography					
Sex	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓
Date of birth	✓	✓	✓	✓	✓
Marital Status	✓	✓	✓	✓	✓
Civil partnership status					✓
Usual address	✓	✓	✓	✓	✓
Whereabouts on census night	✓	✓	✓		
Household composition	✓	✓	✓	✓	✓
Household relationships	✓	✓	✓	✓	✓
Second address					✓
Term time address			✓	✓	✓
Education					
Student or school child				✓	✓
School level qualifications	✓			✓	✓
Scientific and technical qualifications			✓	✓	✓
Qualifications	✓	✓	✓	✓	✓
Ethnicity, identity, language and religion					
Ethnic group			✓	✓	✓
National Identity					✓
Passports held (citizenship)					✓
Main language spoken					✓
English language proficiency					✓

Topics covered	1971	1981	1991	2001	2011
Welsh language	✓	✓	✓	✓	✓
Religion				✓	✓
Health & caring					
General health				✓	✓
Long term limiting illness			✓	✓	✓
Provision of care				✓	✓
Housing					
Accommodation type				✓	✓
Central heating			✓	✓	✓
Tenure	✓	✓	✓	✓	✓
Landlord				✓	✓
Number of rooms	✓	✓	✓	✓	✓
Number of bedrooms					✓
Number of cars or vans	✓	✓	✓	✓	✓
Shared accommodation	✓	✓	✓	✓	✓
Labour Market					
Economic activity (employed unemployed, retired)	✓	✓	✓	✓	✓
Employee status (employee, self-employed)	✓	✓	✓	✓	✓
Supervisor of employees				✓	✓
Occupation	✓	✓	✓	✓	✓
Occupation 1 year ago	✓	✓	✓	✓	✓
Industry of employer	✓	✓	✓	✓	✓
Workplace address			✓	✓	✓
Distance travelled to work			✓	✓	✓
Transport to work	✓	✓	✓	✓	✓
Hours worked	✓		✓	✓	✓
Time since last worked				✓	✓
Migration					

Topics covered	1971	1981	1991	2001	2011
Address one year ago	✓	✓	✓	✓	✓
Address five years ago	✓				
Address ten years ago	✓	✓	✓	✓	✓
Distance moved	✓	✓	✓	✓	✓
Country of birth	✓	✓	✓	✓	✓
Parents country of birth	✓				
Intended length of stay in UK					✓
Year of entry to UK	✓				✓
Socio-economic					
Social class (NS-SEC, RG Social Classification)	✓	✓	✓	✓	✓

+ Some variables such as date of birth are restricted due to their more disclosive nature.

References

¹ The NHS Health and Social Care Information Centre,
<http://www.hscic.gov.uk/>

² Details of the 2011 Census quality assurance process are available at:
<http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-we-took-the-2011-census/how-we-processed-the-information/data-quality-assurance/index.html>

³ Office for National Statistics (2014), 'Longitudinal Study 2011 Census Linkage Report', available at:
<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/quality/quality-assurance/index.html>

⁴ The VML (Virtual Microdata Laboratory) is ONS's facility for providing secure access to sensitive detailed data for statistical research purposes that serve the public good. More information is available at:
<http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/virtual-microdata-laboratory--vml-/index.html>

⁵ Office for National Statistics (2014), '2011 Census Glossary of Terms' available at:
<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/glossary/index.html>

⁶ Copies of the Household questionnaires for England and Wales are available at:
<http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-we-took-the-2011-census/how-we-collected-the-information/questionnaires--delivery--completion-and-return/2011-census-questions/index.html>

⁷ Office for National Statistics (2012), 'Address Register', available at:
<http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-did-we-do-in-2011-/index.html>

⁸ Office for National Statistics (2012), 'Providing the online census', available at:
<http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-did-we-do-in-2011-/index.html>

⁹ Office for National Statistics (2010), 'Final recommended questions 2011 - Enumeration base', available at:
<http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-we-planned-the-2011-census/questionnaire-development/finalising-the-2011-questionnaire/index.html>

¹⁰ Office for National Statistics (2012), '*2011-2001 Census in England and Wales Questionnaire Comparability*', available at:

<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/comparability-over-time/index.html>

¹¹ Office for National Statistics (2012), '*2011 Census Item Edit and Imputation Process*', available at:

<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/methods/index.html>

¹² Office for National Statistics (2012), '*Response Rates in the 2011 Census*', available at:

<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/quality/quality-measures/response-and-imputation-rates/index.html>

¹³ Office for National Statistics (2014), '*2011 Census Quality Survey Report*', available at:

<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/quality/quality-measures/assessing-accuracy-of-answers/index.html>

¹⁴ Office for National Statistics (2012), '*Data capture, coding and cleaning*', available at:

<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/quality/quality-measures/data-capture--coding-and-cleaning/index.html>

¹⁵ Office for National Statistics (2013), '*Item non-response, editing and imputation rates*', available at:

<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/quality/quality-measures/response-and-imputation-rates/index.html>

¹⁶ Office for National Statistics (2012), '*Evaluation – Item edit and imputation*', available at:

<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/quality/quality-measures/response-and-imputation-rates/index.html>

¹⁷ Office for National Statistics (2014), '*Quality of tracing at the 2011 Census*, at:

<http://www.ons.gov.uk/ons/guide-method/user-guidance/longitudinal-study/what-s-new/index.html>

¹⁸ Office for National Statistics, '*Quality of tracing in the Longitudinal Study*'; available at:

<http://www.ons.gov.uk/ons/guide-method/user-guidance/longitudinal-study/data-quality/quality-of-tracing/index.html>

¹⁹ Office for National Statistics (2014), '*2011 Census, Longitudinal Study sampling fractions*', available at:

<http://www.ons.gov.uk/ons/guide-method/user-guidance/longitudinal-study/what-s-new/index.html>

²⁰ Office for National Statistics, 'Sampling quality; available at:
<http://www.ons.gov.uk/ons/guide-method/user-guidance/longitudinal-study/data-quality/sampling-quality/index.html>

²¹ Office for National Statistics, 'Quality of linkage 2001 to 2011', at:
<http://www.ons.gov.uk/ons/guide-method/user-guidance/longitudinal-study/what-s-new/index.html>

²² Office for National Statistics, 'Census to census linkage', at:
<http://www.ons.gov.uk/ons/guide-method/user-guidance/longitudinal-study/data-quality/quality-of-census-to-census-linkage/index.html>

²³ Office for National Statistics (2003), 'LS user guide 20; Researching Households and Families using the ONS', at:
<https://www.ucl.ac.uk/celsius/other-online-resources-selected/ls-user-guides>

²⁴ Office for National Statistics, 'Census to census linkage', at:
<http://www.ons.gov.uk/ons/guide-method/user-guidance/longitudinal-study/data-quality/quality-of-census-to-census-linkage/index.html>