Opportunities for new research using the post-2001 ONS Longitudinal Study

Louisa Blackwell, Bola Akinwale, Angela Antonatos, John Haskey
Office for National Statistics

INTRODUCTION

In September 2004 the Office for National Statistics (ONS) launched the post-2001 Census Longitudinal Study (LS) database for research use. The addition of 2001 Census data to the LS extends the scope for longitudinal research using this unique data source. The addition of the 2001 Census to the 1971, 1981 and 1991 information already linked allows analysis of individual change through a three-decade span of the life course. There are new possibilities for analysing inter-generational transmission and change, in addition to new topics such as religion and self-reported health. This article highlights the information available for addressing specific social policy issues. It describes the implications of the new census methodologies in 2001 for LS users. It also describes the quality of the LS, considering the completeness of the linkage of LS members through time and the representativeness of the LS data to the census. Highlighting the potential of the LS for investigating a wide range of social policy concerns, we also describe data access arrangements and the provisions for user support.

DESCRIPTION OF THE LONGITUDINAL STUDY

The LS is a record linkage study that includes census and life event information for one per cent of the population of England and Wales. The original LS sample was drawn from the 1971 Census, using four dates of birth. The same dates were used to select the sample from subsequent censuses and to add events such as births, deaths, cancer registrations, widow(er)hoods and armed forces enlistments to the study from vital events registrations. New LS members with the same four dates of birth enter the study at birth and on immigration to England and Wales. LS members leave through death and emigration; exit events are recorded in the LS and the LS members’ records retained for analysis. LS members’
Figure 1: Structure of the ONS Longitudinal Study

Entries 1971–2001:
Births 228,000
Immigrations 122,000

Exits 1971–2001:
Deaths 201,000
Emigrations 32,000

1971
530,000 LS members found in the 1971 Census

1981
534,000 LS members found in the 1981 Census

1991
543,000 LS members found in the 1991 Census

2001
540,000 LS members found in the 2001 Census

Birth and death events 1971–2002/04 for sample members:
Births to sample mothers 215,000
Births to sample men 49,500
Infant deaths 2,500
Widow(er)hoods 70,000

Other events
Cancer registrations 78,000
re-entry to England and Wales is also recorded. The quality of such data is discussed in a separate section later in this article. Information on other people living in LS members’ households is also collected at each census, and is available for analysis.

**Structure of the LS database**

Figure 1 describes the structure of the LS. Around 500,000 LS members have been selected at each census. The LS sample is updated between censuses through entry (birth, immigration) and exit (death, embarkation) events. Since the study began, 944,000 LS members have been included. There have been changes over time in the data that are linked for each member, beyond those arising through classificatory change. For example, questions asked at each census have varied slightly. While key characteristics such as age, sex, legal marital status, occupation and economic activity were collected at all four censuses, items such as limiting long-standing illness and ethnicity were only collected in 1991 and 2001, and religion and self-rated health were only collected in 2001. In terms of events, births to fathers in the LS sample were discontinued and 2001, and religion and self-rated health were only collected in 2001. In terms of events, births to fathers in the LS sample were discontinued in 1979 because of data quality shortfalls. A comprehensive account of LS events processing of the pre-2001 dataset can be found in Hattersley and Creeser.1

The LS uses routinely collected administrative data and thus it does not impose any respondent burden. Although the collection of census data does impose a burden, there is no additional burden as a result of the LS. Data linkage takes place largely through ‘tracing’, which involves finding LS members’ records on the National Health Service Central Register (NHSCR). Census records are traced at NHSCR to enable linkage with event data and previous census records. Birth, death and cancer registrations are traced at NHSCR while migration and enlistment records are supplied to the LS from NHSCR. Cross-referencing of NHSCR and registration files for cancers and deaths helps to maintain high quality for these events. Tracing rates are an important indicator of quality for the LS since the NHSCR enables records for LS sample members to be linked to various life events for these individuals. It also facilitates matching of records collected at different points. Tracing rates achieved in the 2001 Census-LS link were the highest so far, with only 0.7 per cent of 2001 Census records untraced, compared with 1.6 per cent in 1991 (Table 1). This improvement can be attributed to new methodologies such as form scanning introduced in the 2001 Census process, as well as the skill of the tracing team based at the NHSCR in utilising the new technologies.2

<table>
<thead>
<tr>
<th>Census</th>
<th>Males (per cent)</th>
<th>Females (per cent)</th>
<th>All LS members (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>97.2</td>
<td>96.5</td>
<td>96.8</td>
</tr>
<tr>
<td>1981</td>
<td>98.8</td>
<td>98.9</td>
<td>98.8</td>
</tr>
<tr>
<td>1991</td>
<td>98.3</td>
<td>98.5</td>
<td>98.4</td>
</tr>
<tr>
<td>2001</td>
<td>99.1</td>
<td>99.5</td>
<td>99.3</td>
</tr>
</tbody>
</table>

Source: ONS Longitudinal Study

**Research using linked 2001 Census information**

When the LS was first established, the report announcing the study explained that while it would inform a range of research and policy topics, including fertility, its primary objective was to permit more accurate and sophisticated analyses of mortality:

‘An important advance in the quality and scope of medical statistics will be achieved by this study of a nationwide sample of individuals. By this method characteristics of persons in the sample, as recorded at Censuses of Population, can be related to their health experience so far as recorded in the study - in particular, deaths and diseases recorded at death registration. No other kind of study could achieve comparable gains in our knowledge about the relation between, on the one hand, mortality and, on the other hand, migration, education, occupation, housing conditions, family structure and ownership of a car. The resulting statistics will contribute directly to a number of areas of medical research.3

Linkage of census and death records has allowed the LS to produce more reliable occupational mortality and life expectancy statistics.4 At death registration, the socio-demographic information collected is limited and occupational data from an informant is likely to differ from that provided at census when the person was still alive. Beyond enhanced mortality analysis, the LS has increasingly supported a diverse range of analyses that variously use the information provided at census, through events registration or both.

Social policy issues that LS analysis has informed include:

- equal opportunities: for women; racial and ethnic groups; the long-term sick
- inequalities: in health; employment; education; geography
- social exclusion, including the long-term outcomes of education and employment status2
- economic integration of migrant groups13
- housing and geographic mobility14
- family policy, including late/early parenthood, different childbearing patterns of advantaged and less advantaged groups, cohabitation, complex households, 15, 16

The addition of 2001 Census records to the LS extends the research potential of the study, to include life course transitions spanning the 30 years between 1971 and 2001. It is possible to analyse the associations between childhood circumstances and various outcomes. For example,
for those aged 15 years and under in 1971, it is possible to explore their childhood circumstances, such as household characteristics, and their outcomes in 2001 at ages 30–45 such as self-reported health, marital status, employment and educational outcomes. In addition, inter-generational analyses, of social or geographic mobility and educational attainment, are possible for children at each census prior to 2001. Similarly, transitions between middle age and retirement can be analysed.

New questions on religion and the provision of care in 2001 mean that these can be analysed in terms of the information supplied in previous censuses. Figure 2 is an example using new information on caring. It shows the 1991/2001 employment histories of males aged 26–74 for analysis, carers were more likely than non-carers to have left employment over the 1990s (17.5 compared to 11.1 per cent).

The repetition in 2001 of a census question on long-term illness and employment over the 1990s (17.5 compared to 11.1 per cent).

The size of the LS, which includes around 500,000 LS members at any one census, makes it a particularly valuable resource for studies of ethnic identity change. The large sample size and length of follow-up make the LS a unique source for statistically robust studies of particular groups, for example, selected occupational and ethnic groups. Figure 3 describes the longitudinal census-to-census sample sizes that are available from the LS. This shows that adding more census points reduces the size of the continuously captured LS sample, but there are still more than a quarter of a million LS members available at all four censuses.

In addition, on a very practical level, the LS has provided a valuable tool in exploring the full potential of the information collected on relationships. It has helped in formulating additional census analyses and derived variables. Because analysis of relationships within a household is essentially multi-dimensional, it is helpful, as a first step, to examine relationships to a specific person who can sometimes be the household reference person and sometimes another person. Analyses of LS members’ relationships has provided important new insights into both the structure of families and households and also the extent to which imputation puts a limit on such interpretations.

### New methodologies in the 2001 Census and their implications for LS analysis

#### Optical character recognition and 100 per cent coding

The 2001 Census was the first at which ONS was allowed to hold names and addresses electronically. Because of the nature of data held in the LS, confidentiality is of paramount importance. Information used to trace LS members at NHSCR is used only to identify the correct LS member and link their data: names and addresses are not retained on the working database. For the first time in 2001, census forms were scanned electronically and from these files, optical character recognition and automatic coding enabled 100 per cent coding of all items; at previous censuses, only 10 per cent of ‘hard-to-code’ items such as occupation and industry were coded. Census form images were available to the LS tracing team, which helped to improve the tracing rate achieved. As an example, information on co-residents helped to verify that accurate matches had been found.

Where data items were missing or inadequately described on census forms, values were imputed to produce 100 per cent coding.18 The inclusion of imputed values could generate spurious results in longitudinal analysis, so in the LS imputed values are flagged, allowing users to exclude them if appropriate.

#### Implications of item imputation for longitudinal analysis

Prior to the release of the updated database, the LS Development Team commissioned a series of projects to test the 2001 LS/Census data for accuracy, comparability, consistency and accessibility. Six substantive analysis projects covered the topics of families and the 2001 relationship matrix, ethnicity and identity 1991–2001, occupational change and family change 1981–2001, health and social mobility 1981–2001, return migration within England and Wales and the social mobility of minority ethnic groups. A common theme emerging from these projects was that the impact of imputation on longitudinal analysis was uneven.

The most sophisticated set of LS imputation flags is for data derived from the Census relationship grid. When up to five people live in a household, the relationship grid asks them to report their relationship to all other members of the household. If there are more than five people in a household, subsequent persons only report their relationship to the first person on the form and to the two persons preceding them. These flags allow LS users to identify and, where appropriate, exclude from their analysis records where sample members either: i) have relationship(s)

---

**Figure 3**

<table>
<thead>
<tr>
<th>Traced LS members found at census</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1971</strong></td>
</tr>
<tr>
<td>513,000</td>
</tr>
</tbody>
</table>

408,000
420,000
418,000
327,000
331,000
256,000

**Total traced LS members:** 944,000

**Source:** ONS Longitudinal Study
Table 2: 2001 highest and lowest relationship imputation rates measured for individuals, families and households

<table>
<thead>
<tr>
<th>Level of analysis</th>
<th>Characteristic</th>
<th>Imputation Category (per cent imputed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Households</td>
</tr>
<tr>
<td></td>
<td>Highest</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Highest</td>
<td>Male (11.6)</td>
</tr>
<tr>
<td></td>
<td>Lowest</td>
<td>Female (10.8)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>Highest</td>
<td>12 (17.4)</td>
</tr>
<tr>
<td></td>
<td>Lowest</td>
<td>93 (3.1)</td>
</tr>
<tr>
<td>Ethnic group</td>
<td>Highest</td>
<td>Bangladeshi (42.4)</td>
</tr>
<tr>
<td></td>
<td>Lowest</td>
<td>White British (9.5)</td>
</tr>
<tr>
<td>Marital status</td>
<td>Highest</td>
<td>Single (13.8)</td>
</tr>
<tr>
<td></td>
<td>Lowest</td>
<td>Widowed (4.2)</td>
</tr>
<tr>
<td>County District</td>
<td>Highest</td>
<td>LB Tower Hamlets (28.3)</td>
</tr>
<tr>
<td></td>
<td>Lowest</td>
<td>City of London (3.9)</td>
</tr>
</tbody>
</table>

Source: ONS Longitudinal Study

between themselves and other household members imputed, or ii) live in families with imputed relationships between family members, or iii) live in households with imputed relationships between household members.

This was illustrated by Haskey, Antonatos and Duke-Williams, who found that imputation levels were high in larger households, rising from 3 per cent in households of two people to 37 per cent for those of six or more people. The increase from smaller to larger households was greater than might simply be expected from the number of relationships that applies to each individual in the household.

Researchers need to be aware of these effects and also take into account the fact that imputed values can generate inconsistencies over time, or implausible results, for example because of inconsistencies between variables within the same census. One example of unlikely combinations of characteristics are medical practitioners without professional qualifications. It would be wise to assess the impact of imputation on the selected sample and for particular variables by running separate analyses, with and without imputed data before deciding on how to proceed.

However, the decision of whether to include or exclude records with imputed items involves a balance between the effect of longitudinal inconsistencies, and the potential bias introduced by dropping large numbers of imputed records within a subsample of interest.

Table 2 shows, for example, that 8.5 per cent of males had a relationship between them and another individual imputed. This compares with 8.1 per cent of females. For males, 10.6 per cent were in a family in which a relationship had been imputed (between family members, not necessarily involving themselves) and 11.6 per cent lived in households with imputed relationships. Turning to age, of all the different ages of LS members, those who were 12 years old were most likely to have a relationship with another individual imputed (14.4 per cent). This rises to 17.4 per cent for those living in a household where a relationship is imputed. By contrast, only 3.1 per cent of 93 year olds have relationships imputed, and this falls to 0.3 with relationships in their family imputed. The most likely explanation for the latter is that 93 year olds have a larger chance than the average person of most other ages of living in a small and family-centred household where the relationships are clear cut.

Among ethnic groups, Bangladeshis were most likely to have a relationship imputed (30.1 per cent). The percentage of Bangladeshis living in households where one or more relationship(s) were imputed is particularly high, at 42.4 per cent. This contrasts with White people, 9.5 per cent of whom lived in households where relationships had been imputed. Further analysis could disentangle the extent to which high imputation rates for Bangladeshi households are explained by their larger size. Table 2 also shows wide regional variations, for example between the City of London, where only 3.9 per cent of households had a relationship imputed, compared to the London Borough of Tower Hamlets, at 28.3 per cent. Again, these differences may be explained by household sizes in those areas.

Platt, Simpson and Akinwale in their comparison of ethnic identity in 1991 and 2001 conclude that imputed records appearing to be for minority ethnic groups should be excluded from analysis.

One Number Census

The ‘One Number Census’ (ONC) used donor imputation to identify and adjust for the number of people and households not counted in the 2001 Census. The adjustment factors were based on a Census Coverage Survey (CCS) conducted three and a half weeks after Census Day. Individuals and households revealed to have been missed by census were imputed, their records ‘donated’ by households with matching characteristics. The
LS members with more than one 2001 enumeration, by type of multiple enumeration

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of LS members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple enumerations:</td>
<td></td>
</tr>
<tr>
<td>Students enumerated at home and at a term-time address (and not a student)</td>
<td>3,691</td>
</tr>
<tr>
<td>Enumerated at home and in a communal establishment</td>
<td>451</td>
</tr>
<tr>
<td>Enumerated at two addresses</td>
<td>412</td>
</tr>
<tr>
<td>Children enumerated at two parental homes</td>
<td>262</td>
</tr>
<tr>
<td>LS member entered twice on one form</td>
<td>201</td>
</tr>
<tr>
<td>Same address enumerated twice</td>
<td>198</td>
</tr>
<tr>
<td>Others</td>
<td>540</td>
</tr>
<tr>
<td>Total</td>
<td>5,755</td>
</tr>
</tbody>
</table>

Source: ONS Longitudinal Study

LS sample was deliberately drawn from census before the ONC process, so no LS members were created in this way. However, the absence of ONC imputation in the LS has implications for its consistency with published 2001 Census tables. This is described in more detail in the discussion of LS sampling fractions below.

De jure enumeration

A further change in census methodology in 2001 was the requirement that people only filled in Census forms at their usual address (de jure enumeration), rather than at their whereabouts on Census night (de facto enumeration). The single exception to this rule was students with a different term-time address, who completed a full census return at the term-time address and for whom a rather shorter return was completed at their vacation address. Thus in 2001, students living away from home were the only legitimate type of multiple enumeration. However, the LS tracing process identified other types of multiple enumeration, listed in Table 3.

Multiple enumerations other than those for students reflect ambiguity in what might be considered some people’s ‘usual residence’ and errors resulting from census form processing. For some individuals their ‘usual residence’ is not easy to define, for example some elderly people were enumerated in a nursing home as well as by their spouse or child in the family home. Some children were included on the census forms of both separated parents. Some individuals used the multiple person pages in the census form to make new, slightly changed versions of their entry and the minor changes between entries prevented these duplicates from being identified in the automatic validity checks carried out by census.

The post-back method used for the 2001 Census allowed enumerator effort to be concentrated in those areas known to be difficult to enumerate. However, it also meant that some people moving house around census time completed and submitted forms for both old and new addresses. Duplicate forms for the same address were missed in validity checks where, for example, one form was incomplete. Other types of multiple enumeration included children or young adults enumerated in a parental home and with other adults, sometimes grandparents, and people enumerated in primary and second homes, for example in town and country. For each multiple enumeration, a combination of automatic and clerical processes identified the primary record, using different criteria for each type. Primary records are stored with single enumerations in the LS database, but secondary records are also held in a separate file.

Dual occupation coding

Because longitudinal analysis often requires continuity over time in classificatory systems, the LS has dual coded occupational data collected at each census, to both the contemporary scheme and to that used in the previous census. Thus there was a requirement to code 2001 Census data to both the 2000 Standard Occupational Classification (SOC2000) and to the 1990 Standard Occupational Classification (SOC90). The continuation of SOC90 coding in the 2001 data means that, for example, a proxy for Registrar General’s social class variable has been derived for LS members, which provides some comparability over time in longitudinal and time series research. The SOC90 coding was achieved through a combination of automatic and clerical coding. A customised version of coding software developed at Warwick University (CASOC2K) was used to automatically code 70 per cent of records. ONS clerical coders coded the remaining 30 per cent of records.22

Quality of the 2001 Census-LS link

Tracing rates

‘Tracing’ involves finding LS members’ records on the NHSCR. The latter enables records for LS sample members to be linked to various life events for those individuals and facilitates the matching of records collected at different times. Tracing rates are therefore an important measure of LS quality.

The tracing rate is defined as:

\[
\text{Tracing rate} = \frac{\text{number of LS members traced in the NHSCR in a subgroup } i}{\text{total number of LS members in subgroup } i} \times 100
\]

Table 1 shows how tracing rates varied at each census, rising from 96.8 per cent in 1971 to 99.3 per cent in 2001. Since 1981, tracing rates for females have been higher than for males. The reverse, found in 1971, was largely attributed to tracing failure as women changed their name upon marriage. Changing marriage patterns and computerisation may help to explain increased tracing of women in subsequent censuses.

Tracing rates for different subgroups within the population in 2001 and 1971, 1981 and 1991 Census-LS samples are available on the National Statistics website.22 Tracing failure in 2001, as at previous censuses, was particularly associated with:

- being a young adult (2.4 per cent of 20–24 year-olds were not traced)
- being divorced (0.6 per cent of divorces were not traced)
- having been born outside the UK (4.7 per cent of those born outside of the UK were not traced)
- being a full-time student (2.4 per cent not traced)
- being in a minority ethnic group (for example, 5.3 per cent of Black Africans and 5.4 per cent of Chinese people were not traced)
- being long-term unemployed or never having worked (0.8 per cent of the long-term unemployed were not traced)
- living in London (2.4 per cent not traced)
- living in certain types of communal establishment (for example, 13.3 per cent of people living in defence establishments and 9.3 per cent of prisoners were not traced)
- serving in the armed forces (for example, 12.8 per cent of non-commissioned officers were not traced).

Sampling fractions

The LS was designed to include four birth dates in each year. Including leap years in the calculation, this should lead to a theoretical model
Sampling fraction of 1.10 per cent (4/365.25). Sampling fractions based on traced LS members should approximate to this, that is:

\[
\text{Sampling fraction} = \frac{\text{Traced LS members in subgroup } i}{\text{Enumerated Census sample in subgroup } i} \times 100.
\]

Table 4 shows that the sampling fractions achieved at each census approximate to the level expected from the study design. Sampling fractions approach 1.10 in 2001. However, with denominators that include the additional people added to census counts through One Number Census imputation (totalling 3.1 million, 2.5 million in imputed households and 0.6 million individuals added to enumerated households), sampling fractions fall to 1.03 in 2001. This difference varies for sub-groups in the population, reflecting the differing degrees of underenumeration that the ONS process accounted for. Sampling fractions with both adjusted and unadjusted denominators and for different groups are available on the National Statistics website. As an example, among males aged 35–39, the sampling fraction is 1.10. However, with a denominator that includes ONS imputation, this falls to 1.02.

Differences in sampling fractions by sex reflect differences in tracing rates by sex. Groups with lowest observed sampling fractions included:
- young adults
- divorced or widowed people
- people born in the USA or serving in the armed forces
- those living in communal establishments.

These patterns largely reflect the patterns of tracing failure described above.

A different pattern, however, is found among LS members in minority ethnic groups, who have tracing rates that are below those for white people, falling to 98.5 per cent for Asian or Asian British and 93.1 per cent for the Chinese and other ethnic group. By contrast, these groups have sampling fractions that are higher than for white people, at 1.39 and 1.23 respectively (Table 5). This table shows sampling fractions based on all LS members and on those who were traced. The level of oversampling among all LS members, traced and untraced, is unaffected by the effect of tracing failure for minority groups. This over-sampling has been attributed to date of birth preferences, which are largely observed among older people born overseas. Sampling fractions for members of ethnic minority groups who were born in England and Wales approach those for White people.

### Linkage rates

Census-to-census linkage rates take into account deaths and emigrations known to have taken place in the inter-censal decade. LS members whose records cannot be linked between censuses are said to be ‘lost to follow-up’. Loss to follow-up in 1981, 1991 and 2001 is summarised in Table 6.

<table>
<thead>
<tr>
<th>Census</th>
<th>Males (per cent)</th>
<th>Females (per cent)</th>
<th>All LS members (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>1.06</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>1981</td>
<td>1.07</td>
<td>1.09</td>
<td>1.09</td>
</tr>
<tr>
<td>1991</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
</tr>
<tr>
<td>2001*</td>
<td>1.09</td>
<td>1.10</td>
<td>1.09</td>
</tr>
</tbody>
</table>

* 2001 Sampling fractions calculated using traced LS members and census populations without adjustment for underenumeration

Source: ONS Longitudinal Study

<table>
<thead>
<tr>
<th>Ethnic group</th>
<th>Traced (per cent)</th>
<th>Sampling fraction* based on:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All LS members</td>
<td>Traced LS members</td>
</tr>
<tr>
<td>White</td>
<td>99.5</td>
<td>1.09</td>
</tr>
<tr>
<td>Mixed</td>
<td>98.2</td>
<td>1.10</td>
</tr>
<tr>
<td>Asian or Asian British</td>
<td>98.5</td>
<td>1.41</td>
</tr>
<tr>
<td>Black or Black British</td>
<td>96.9</td>
<td>1.15</td>
</tr>
<tr>
<td>Chinese or other</td>
<td>93.1</td>
<td>1.23</td>
</tr>
<tr>
<td>Total</td>
<td>99.3</td>
<td>1.10</td>
</tr>
</tbody>
</table>

* 2001 Sampling fractions calculated using traced LS members and census populations without adjustment for underenumeration

Source: ONS Longitudinal Study

It is difficult to disentangle these various reasons. However, completeness of death registration and LS linkage of death data imply that increased loss to follow-up over time is attributable to unobserved embarkation and underenumeration at census. A comprehensive account of loss to follow-up in 2001 and previous censuses found that the 1991 characteristics associated with linkage failure in 2001 included:
- being young and male
- being an immigrant
- living in London
- being single, divorced, cohabiting or in a lone parent household
- being born outside England and Wales
- being in a minority ethnic group
- serving in the armed forces.

### Conclusion

Use of the LS for research into a wide range of social policy issues is supported by two teams of user support officers, one for academics and another for other researchers (including from government departments, local authorities and overseas). In early 2006, similar record linkage from 1991 onwards will be available for Scotland. This will extend the scope for comparative analysis.
Box one

**Accessing LS Data**

Support in the extraction and analysis of LS data is currently provided by two user support teams. Academic users working in the UK Higher or Further Education sectors are assisted by the Centre for LS Information and User Support (CeLSIUS) who are based at the London School of Hygiene and Tropical Medicine. In addition to a broad range of descriptive information on the LS, CeLSIUS provides a range of online training modules to assist new and existing users. CeLSIUS also maintains a list of publications based on LS research, which is a useful source for potential analysts as it illustrates the broad range of research topics that the LS supports.

**Academic users:**
E-mail: celsius@lshtm.ac.uk  
Website: [http://www.celsius.lshtm.ac.uk](http://www.celsius.lshtm.ac.uk)

Support for non-academic users is provided by staff within ONS. The ONS LS webpages provide background information on LS processes and describe the quality of the data contained in the study. Both the National Statistics and CeLSIUS websites provide access to the LS data dictionary, which describes the 4,500 variables currently held in the LS. The dictionary contains, for each variable, a short description, list of codes and their labels, which groups of LS records are coded, notes about usage, start and end dates for event variables, derivations and whether access is restricted.

**Government and other non-academic users:**
E-mail: ls@ons.gsi.gov.uk  
Website: [http://www.statistics.gov.uk/ls](http://www.statistics.gov.uk/ls)

**Notes and References**


**Key findings**

- Linkage of 2001 Census data to the ONS Longitudinal Study (LS) allows analysis of change in individuals’ circumstances through three decades, as well as some analysis of generational effects.
- The LS includes around half a million LS members at each census since 1971. Over 400,000 were present at two consecutive censuses and around 330,000 at three consecutive censuses. Over a quarter of a million were present at all four.
- LS data have been used to support research in a broad range of social policy areas, including social inequalities, equal opportunities, social exclusion, family policy and the economic integration of migrants. New census questions in 2001 on religion, unpaid care and self-reported health offer new avenues for investigation.
- Census to census and events data linkage in the LS is achieved through tracing LS members on the NHS Central Register. Tracking rates in 2001, at 99.3 per cent, were the highest achieved so far.
- As well as providing an estimate of males missing from the 2001 Census, the tracing process also identified some overenumeration. The LS shows that multiple enumeration arose because of ambiguity in the concept of ‘usual residence’, in the way people live and, to a lesser extent, processing error.
- Dual coding of 2001 occupations to both the SOC90 and SOC2000 classifications means that an approximation to Registrar General’s social class can be derived for 2001. A similar proxy for NS-SEC in 1991 allows researchers to establish continuity in time series and longitudinal analysis.


25. See, for example, the sampling fractions by age, sex and ethnic group: http://www.statistics.gov.uk/about/data/methodology especific/population/LS/downloads/sf01.pdf


27. The Longitudinal Studies Centre Scotland website contains details of the forthcoming Scottish LS: http://www.lscs.ac.uk/