

EVENTS

Introduction

The LS is a 1% sample of the population of England and Wales. It contains information from the census and vital event registrations. It was started with a sample from the 1971 Census.

This guidance introduces some of the vital events data in the LS; it describes how they are linked to other records in the LS database and how these events can act as entry to and exit from the LS for sample members.

The aims are:

- To inform potential users of the LS of the vital events data available.
- To detail variables that may be used, along with any restrictions.

Having read through the guidance the reader should understand:

- How to use the Data Dictionary when looking for vital event variables.
- The time period in which these variables are valid and how to use each in conjunction with others.

Entry events

The LS is a dynamic sample, meaning new members can be added while existing members 'leave' the sample over time. Detailed composition of the LS can be found in the Study Population guidance.

The original sample members were selected from the 1971 Census on the basis of birthday. New LS sample members can enter the LS in one of three ways:

- At birth (a new birth on one of the four LS birthdays).
- If they register with an NHS doctor (and have one of the LS birthdays).
- If they are enumerated at a census, having one of the LS birthdays and not having entered the LS at a previous census.

The first registration with an NHS doctor after the 1971 Census for those who were not already present in the NHS Central Register is called an 'immigration event', and any subsequent registrations are called 're-entries' (even if there had been no previous record of an 'emigration' - see Exit Events); therefore an LS member can only have a single point of entry, e.g. one record on the Immigrations table (IMMG (registration with NHS)) in the [CeLSIUS Data Dictionary](#), but may have multiple records on the Re-entries table (REEN) in the CeLSIUS Data Dictionary).

Every LS member has a record on the CORE1 table (Core information for each LS member) of the [CeLSIUS Data Dictionary](#), which includes indicator variables such as presence or absence at each census and an overarching unique identifier (CORENO). The CORE1 table acts as an index with which all other information for particular individual LS members can be linked. Values for sex (SEX)

and year of birth (DOBYR) at the entry event are kept on the CORE1 table. If the value for sex or year of birth from the entry event differs with a subsequent event/census, the discrepant value for the subsequent event is stored on the particular event file. A record of the discrepancy is made on the discrepancy indicators (SEXD001 to SEXD006 for sex, and DOBD001 to DOBD006 for year of birth in the CORE1 table) [see the Data Dictionary for further information on these variables]. Sample members can have up to six discrepancies for sex or year of birth.

Exit events

Sample members may have records of 'exit' events in the data, but due to the dynamic nature of the LS, their records are permanently retained in the database and are available for analysis.

- An LS sample member can 'leave' the LS by de-registering from the NHS (by informing their GP or handing in their NHS card) on leaving the country (called embarkations) or on death. These events are recorded in the EMBR and DETH tables respectively (see the [CeLSIUS Data Dictionary](#) for information on what is recorded in these tables). An LS member can 'embark', 're-enter' and 're-embark', so may have more than one embarkation event. If the sample member embarks or dies, their records are retained.
- Enlistment into the armed forces can also be used as an exit point (ENLS table, last updated 2013) since they enter into the Armed forces healthcare system, but should be checked against the re-entry table (REEN).
- Until the end of 1983 entry to a psychiatric hospital was also recorded (in the PSYC table) and could be used as an exit point. Re-entries from long stay psychiatric hospitals are also recorded but only if the subject entered the hospital before 1984.
- LS users should be aware that, apart from the death table (DETH), the data available in these cases is very limited and it is believed that only about 50% of embarkations are reported to NHS Central Register. However, for time dependent analyses these data should be used to exclude subjects at the point of exit as beyond this they are known not to be 'at risk'.

Other events

- The LS also contains information on live births (LBSM table), stillbirths (SBSM table) and deaths of the infant children (IDMI table, last updated in 2013) of female sample members (known as sample mothers). Information for sample fathers is only available for a short period in the 1970s (in tables LBSF, SBSF and IDFI).
- Deaths of husbands and wives of sample members (widowhoods/widowerhoods, in the WIDOW table) are also linked to the surviving spouse using their date of birth given at the registration of the death. The information is not captured if the couple are not legally married.
- Cancer registrations to sample members are also added to the LS database in the CANC table. This information is taken from the eight English cancer registries and the Welsh Cancer Intelligence and Surveillance Unit.

See the [CeLSIUS Data Dictionary](#) for information on the variables in these tables.

How are the events linked to the other records in the LS database?

LS sample members are selected for inclusion in the study if their birthday falls on one of four days within any calendar year. This date of birth along with other personal information (not stored on the LS) is used to link the LS member from one census to another and to the vital events in their lives.

There are two processes by which this happens:

1. The ONS birth and death files and cancer registrations files are searched for individuals with one of the LS birthdates. These records are isolated and an attempt is made to link them to actual LS sample members. Only those event records that match actual LS sample members are linked to their existing records.
2. LS sample members are flagged (i.e. marked as LS members) at the NHS Central Register, and when an event is recorded at the NHS Central Register, the entries are sent to the LS Development Team to be added to the LS database. Information which is recorded at NHS Central Register includes registration with a doctor, deaths to sample members, cancer registrations, and de-registrations from a doctor.

Deaths to sample members and cancer registrations are captured by both these methods and as a result, the linkage rates are very good. An explanation of how sampling fractions and linkage rates are calculated is available from ONS' [Longitudinal Study \(LS\)](#) website.

NHS Central Register

The NHS Central Register (NHSCR) is a database of all those who are or ever have been registered with an NHS doctor in England and Wales. Routine events are added to the database when they are notified to NHS Central Register. Most LS sample members have an entry in the NHS Central Register and have been found and flagged there; this process is called 'tracing'. The Register is used to identify the LS member to whom an event has occurred; this process is called 'matching'. (These processes, with 'tracing', are explained in the Study Population guidance).

At each census, all sample members will be checked to see if they are present at NHSCR. If a sample member is not found on NHSCR, they are referred to as 'not traced'. This is important because events cannot be linked to sample members if they are not traced. Hence a traced sample should be used as the denominator if you are following events in your project. However, the tracing rate - the number of people in the LS who are traced at NHSCR - is very high.

Event data are usually added annually with a two-year delay. For example, information on events in 2015 was added to the LS database in July 2017, and data for 2016 was added in July 2018.

The exception is cancer registration, where information may be added with a three-year delay (or more).

Tables in the LS Database

Each event has a table (i.e. a separate dataset), which contains information for that event. Data from different tables can be linked by the CORENO variable, the unique identifier of an LS member. In many tables an LS member may have more than one record - for example, a sample mother may have several births, or a sample member can be widowed more than once.

Below is a list of the principal tables contained in the LS database. The variables contained in each of these tables can be seen in the Data Dictionary.

List of the principal tables contained in the LS database

Table name	Table description
CORE1	Core file (contains all sample members and key indicator variables)
ME71	Sample members at 1971 Census
ME81	Sample members at 1981 Census
ME91	Sample members at 1991 Census
ME01	Sample members at 2001 Census
ME11	Sample members at 2011 Census
NM71	Other members of sample member's household - 1971 Census
NM81	Other members of sample member's household - 1981 Census
NM91	Other members of sample member's household - 1991 Census
NM01	Other members of sample member's household - 2001 Census
NM11	Other members of sample member's household - 2011 Census
LBSM	Live births to sample mothers
LBSF	Live births to sample fathers (1971-78, 1981)
SBSM	Stillbirths to sample mothers
SBSF	Stillbirths to sample fathers (1971-78, 1981)
IDMI	Infant deaths to sample mothers
NBIR	New births of sample members
CANC	Cancer registrations to sample members
DETH	Deaths of sample members
WDOW	Deaths of spouse of sample members
EMBR	Embarkations (deregistration with NHS)
IMMG	Immigrations (registration with NHS)
REEN	Re-entries (re-registration with NHS)

Vital Events Tables

- This guidance introduces seven different events which are part of the LS database.
- Each section includes information on the time period that the data are available for, whether there are multiple records, what the unique identifier is, how the data are linked and what the source of the data is.
- It gives a brief description of the data, including important variables.
- It discusses some analysis issues and the quality of the data, using tables from ONS.
- It shows how many events occur each year.

For each of the events you can see the information that is recorded in the related dataset (table) by searching for it in the [CeLSIUS Data Dictionary](#).

Live Births to Sample Mothers (LBSM)

Years for which data are available: Census Day 1971 to the current year minus two.

Can the LS member have more than one record: Yes.

Key identifier variables: CORENO + EVENTORD.

Number of events a year: 6,000 to 7,000.

Means of linkage: Annual date of birth search of the ONS birth file for female LS sample members who had a live birth.

Source: Birth registration and birth certificate.

Live births to Sample Fathers (LBSF): Information is also available for live births to LS sample fathers, but only for births to fathers in the years 1971-78 and 1981 and where the father's information is recorded on the birth certificate. Variable names are the same as for LBSM but end with the letter F instead of M (e.g. SEXABF and SEXABM). The linkage rate for live births to sample fathers for 1971-78 was 86%.

Basic description:

Every live birth in England or Wales must be registered within six weeks, and a database of these registrations is kept at ONS Titchfield. At regular intervals, details of births to mothers with LS birth dates (LS sample mothers) are extracted from this database and sent to the NHS Central Register. There, the staff attempt to trace each mother in the NHS Central Register. If the birth is 'matched' (i.e. they find that the mother is flagged as an LS member) they will add the LS identification number to the birth details and return them to ONS Titchfield; however, a proportion will be unsuccessful, mainly because the mother's date of birth has been wrongly recorded on the birth registration.

For those mothers successfully matched, a new record is created in LBSM in the LS database. If the mother is not traced in the NHS Central Register, the event will not be entered in the LS database; a fresh attempt will be made to trace the mother in NHS Central Register at a later date. The infant does not become an LS member unless the birth took place on an LS birth date, in which case the procedure for the birth of a new LS Sample Member will also be followed, with a new entry in NBIR (the New births of LS members table).

Live births to sample mothers are not an entry event to the LS for the mother. Therefore, a birth will not be added to the LS database if the mother is not already a sample member, unless the child also has an LS date of birth, when the child only (not the mother) will be added to the LS.

Key documents

- Babb P, Bethune A. Trends in births outside marriage. Population Trends 1995; (81): 17-22
- Grundy E. Women's Migration: Marriage, Fertility and Divorce. LS Series No 4. 1989

- Penhale, B. Associations Between Unemployment and Fertility Among Young Women in the Early 1980s. LS working paper 60. 1987.
- Werner, B. Some examples of fertility analysis from the Office of Population Censuses and Surveys' Longitudinal Study. Population Trends 1984; 35: 5-10.
- Hattersley, L. and Creeser, R. Longitudinal Study 1971-1991: History, organisation and quality of data. LS Series no. 7

[See the Resources section of the CeLSIUS website for links to some of these documents.]

Important variables in the Live Births to Sample Mother's table

[see the LBSM table in the Data Dictionary for more information on these variables and a full list of the variables available]

Variable	Label + comments
CORENO	LS Member unique identifier
EVENTORD	Event number for LS member
BIYRABM	Year of baby's birth (month and day are also available)
SEXABM	Sex of baby
MOAGYRBM	Age of mother at birth of baby. May be imputed (see Data Dictionary) - imputation flagged by MOAGINBM
BWEGRBM	Birth weight in grams (from 1977 onwards; not recommended for use before 1980, when the percentage of birth records with birth weight reached 85.7%. From 1981 to 2006 it was more than 94% each year.)
MLCSABM	Multiple birth indicator. It is possible to link these multiple births, e.g. to study twins.
DOCONBM	Date of conception – this is derived using the actual birth date minus the gestation period. If the gestation period is missing it is calculated by subtracting 231 days from the actual birth date.
In CORE1 not LBSM:	
BRD1-5	Number of live births in first (second, third, fourth, fifth) decade (i.e. 1st decade of the LS = Census Day 1971 to Census Day 1981)

Other variables available in LBSM include geographic classifications for place of birth and parents' usual address. They also include socio-economic indicators for each parent, based on occupation (until 1986 occupation was only collected for the 'working parent', usually the father); however, recording of mother's occupation is not obligatory, and although all cases are coded from 1986, it is thought that many mothers in regular employment are coded as having no occupation. This of course affects the reliability of women's social class coding. There is also a variable in LBSM which gives the number of live and stillborn children previously born to the sample mother (PARABM), from the birth registration, which for migrant LS members may be a good way to check whether a migrant LS member had any children prior to arriving in the UK and being added to the LS. However, prior to 2012, this was only asked of married mothers. In addition, in 2012 the question that was asked changed, specifying that the mother should only report the births prior to the birth that they were

registering¹. This therefore affects the reliability of using the PARABM variable to ascertain whether a LS member had had any births prior to entering the LS sample.

Analysis issues

Each record in the live births table (LBSM) represents one live-born infant; each record in the SBSM table represents one still-born infant. If, for example, a mother gives birth to two live triplets and one dead triplet, there will be two records in LBSM and one in SBSM. The variable MLCSABM (multiple birth indicator) indicates how many births (live births and stillbirths) there were from the pregnancy. This variable is present on both the live birth (LBSM) and stillbirth (SBSM) tables. Live births and stillbirths may not have occurred on the same day; live-born twins may also have been born either side of midnight. Date of birth and the multiple birth indicator (MLCSABM) can be used to link multiple birth records if desired.

The variable EVENTORD in the table only distinguishes births within the table (and therefore cannot be used, for example, to combine data from LBSM and SBSM). A mother who has had six live-born children will have six records in LBSM and the variables CORENO+EVENTORD will uniquely identify each child, though not necessarily in birth order.

Data quality

Virtually all live births in England and Wales are registered. However, births which take place outside England and Wales will not appear in the LS and therefore the number of live births to a sample mother may be underestimated.

Another factor affecting data quality is incorrect recording of mother's date of birth on the birth registration. If the mother is not an LS sample member but has been given an LS date of birth at the birth registration, she will not become an LS member (unless, of course, the same erroneous birthday has been recorded for her on a census form). More seriously, if the mother is an LS sample member but is given a non-LS birth date in the birth register, the birth will not be added to her LS record. However, from time to time a comparison is made between children aged under 10 in the mother's household at census and records of births in the previous decade; by this means, some birth registrations with erroneous maternal dates of birth are identified and belatedly added to the mother's record.

Between 1971 and 1981, births to unmarried mothers had a lower trace rate than did births to married mothers, but this difference has diminished considerably over time. Generally speaking, data quality for births to sample mothers is thought to be better for mothers who were born in the UK than for mothers born elsewhere. Further information is available on the ONS site for the [Longitudinal Study \(LS\)](#). However, data quality (as measured by comparing numbers of births to LS mothers with total

¹ The question specified that only births prior to the birth currently being registered should be included because it was noticed that there was a trend for mothers to include the birth being registered in the number of births prior birth (it appears that at the time of the registration the majority of mothers considered the birth being registered as a prior birth).

numbers of births in England and Wales) has improved substantially over the lifetime of the LS, as have trace rates.

The table produced by ONS below is for live births and stillbirths combined. Linkage rates have increased over time and the rate for 2013 is 101.50. Linkage rates over 100 denote that the actual number of live and still births to sample mothers was higher than expected.

Year of birth	LS births	England and Wales	Sampling Fraction	Births expected in the LS	Linkage Rate
1971-81	65,342	6,454,681	1.01	71,170	91.81
1981-91	68,920	6,696,066	1.03	72,997	94.41
1991-2001	70,707	6,545,635	1.08	71,593	98.76
2001-2011	74,548	6,607,617	1.13	72,894	102.27
2011-2016	45,843	4,090,157	1.12	44,824	102.25

Source: Live and still births to sample mother 1976-2016
Office for National Statistics Longitudinal Study

Number of events

Table 1: Number of Live Births to Sample Mothers (1971-2016) [see Table 1 in Events Data Tables (PDF)].

Stillbirths to Sample Mothers (SBSM)

Years for which data are available: Census Day 1971 to the current year minus two.

Can the LS member have more than one record: Yes.

Key identifier variables: CORENO + EVENTORD.

Number of events a year: 25 to 70.

Means of linkage: Annual date of birth search of ONS birth file for female LS sample members who had a stillbirth.

Source: Stillbirth registration and stillbirth certificate.

Stillbirths to Sample Fathers (SBSF): Information is also available for stillbirths to LS Sample Fathers, but only for stillbirths to fathers in the years 1971-78 and 1981, and (in those years) where the father's information is recorded. Variable names are the same as for Stillbirths to Sample Mothers but end with the letter F instead of M (e.g. MLCSABF compared with MLCSABM).

Basic description

A stillbirth is the birth of a dead baby. Before October 1992, a stillbirth had to follow at least 28 weeks' gestation. This was reduced to 24 weeks or more from October 1992. Stillbirths account for less than 1% of births.

Live births and stillbirths must be registered within 42 days by law. An annual search of the ONS birth file (which includes live births and stillbirths) is carried out by the LS Development Team. The file is checked for LS sample mothers and (up to 1981) fathers, using the date of birth information from the birth registration. Fathers were only linked if the birth was within marriage or jointly registered; if both parents were LS sample members, they were linked separately.

If more than one stillbirth occurs in the pregnancy, each stillbirth will have a separate record. Multiple birth records can be identified using MLCSABM (in the mother's file) and MLCSABF (in the father's file). Note that a stillbirth may take place at some time before a live birth where there is a multiple pregnancy.

Important variables in SBSM

[see the SBSM table in the Data Dictionary for more information on these variables and a full list of the variables available]

Variable name	Short description
CORENO	LS Member unique identifier
EVENTORD	Event number for LS member
BIDYABM	Day of stillbirth
BIMTABM	Month of stillbirth
BIYRABM	Year of stillbirth
EDATEBM	Date of event
DORBM	Date of registration
LEGABM	Whether stillbirth within marriage at time of registration
MLCSABM	Multiple code for birth to female LS member
MOAGMTBM	Age (Months since last birthday) of sample mother at birth of child
MOAGYRBM	Age (years) of sample mother at birth of child
TBHTMBM	Total no. births to mother from this pregnancy (1993 onward)

Key Documents

- Bethune, A. et al. 'Fertility and infant mortality in the OPCS Longitudinal Study' Population Trends (68) Summer 1992 pp24-29
- Sloggett, A. & Joshi, H. 'Deprivation indicators as predictors of life events 1981-92 based on the UK ONS Longitudinal Study' Journal of Epidemiology & Community Health 1998 52(4) pp228-33
- Hattersley, L. and Creeser, R. Longitudinal Study 1971-1991: History, organisation and quality of data. LS Series no. 7

[See the Resources section of the CeLSIUS website for links to some of these documents.]

Data Quality

Linkage rates for stillbirths have not been produced by ONS. However, ONS has produced sampling fractions and linkage rates for all births (live births and stillbirths) (see table above).

Number of events

Table 2: Number of Stillbirths to Sample Mothers (1971-2016) [see Table 2 in Events Data Tables (PDF)].

Infant Deaths to Sample Mothers (IDMI)

Years for which data are available: Census Day 1971 to end of 2013².

Can the LS member have more than one record: Yes.

Key identifier variables: CORENO + EVENTORD.

Number of events a year: 25 to 95.

Means of linkage: Annual date of birth search of the Infant Death file for LS sample mothers.

Source: birth registration and certificate, and death registration and certificate.

Change in the scope of the data:

Until January 1993, only deaths of infants up to one-year old were included in the ONS database, and therefore recorded in the LS. From January 1993, the death of any child born in 1993 or later was included, so that by 2006 the maximum age of a child death that was included was 13. The LS data follows this change in the national database. As most deaths to children occur before the age of one, there are small numbers of child deaths over that age - only 110 for the period 1994 to 2006. However, numbers will increase as more children are added to the ONS database. In the LS, deaths to children over the age of one can be identified in the IDMI dataset using the variable AGCUTIM.

Basic description

A national database was set up for infant deaths in England and Wales in 1976. Information on infant deaths to LS sample mothers is available from 1971 but until 1976 the linkage was carried out manually and the quality of the data suffered as a result. Caution should be used if using the data for this period.

Information on deaths of infants of LS sample mothers is created by linking birth and death registration data for children who died under the age of 1 year (or, from January 1993, who were born in 1993 or later). The LS sample mother is identified by the mother's date of birth from the birth registration.

Information for sample fathers is only available for a limited period (1976-78).

Important variables in IDMI

[see the IDMI table in the Data Dictionary for more information on these variables and a full list of the variables available]

Variable name	Short description
CORENO	LS Member unique identifier (i.e. of the mother)
EVENTORD	Event number for LS member
AGCUTIM	Calculated age unit (1993 onwards); missing (blank) if child born before 1 January 1993.

² IDMI records are no longer processed and therefore data is only available up to 2013.

Variable name	Short description
AGDC3DIM	Age of child at death (1993 onwards) in units defined by AGCUTIM
AGDC5DIM	Age of infant at death (to 1992)
BIWGRIM	Birthweight (g) of child
BIDYAIM	Day of birth of child (1976 onwards)
BIMTAIM	Month of birth of child from birth registration (or BIYRCDIM from death registration, but this has missing values whereas BIMTAIM does not from 1976 onwards)
BIYRCDIM	Year of birth of child (1976 onwards)
DEDYBIM	Day of child death
DEMTBIM	Month of child death
DEYRBIM	Year of child death
EDATEIM	Combined date of death (1993 onwards)
DORBIM	Date of registration of birth (1993 onwards)
DORDIM	Date of registration of death (1993 onwards)
REYRABIM	Year of registration of birth (1971-1992)
REYRADIM	Year of registration of infant death (1971-1992)
MMAGMTIM	Age (months since last birthday) of sample mother at birth of child
MMAGYRIM	Age of sample mother at birth of child (years)
PARAIM	Number of live/stillborn children previously born to sample mother.
SEXABIM	Sex of child from birth registration (or SEXADIM from death registration, but this has some missing cases whereas SEXABIM does not)
MLCSAIM	Multiple birth code from birth registration
TBTHTMIM	Number of children born at time of birth of child who subsequently dies (1993 onwards)

Analysis issues:

Birthweight in grams: available from 1977 onwards but not recommended for use before 1980, when the percentage of birth records with birth weight reached 85.7%. From 1981 to 2006 it was more than 94% each year.

The occupation of parents of the children is coded to the most recent occupational classification at the time of the infant/child death. It is therefore coded to C070, CO80, SOC90 or SOC2000.

Cause of death information is coded to the most recent ICD classification at the time of the infant/child death. It is therefore coded to ICD8, ICD9 or ICD10.

For children born in 1976 or later it is possible to link cases in IDMI with cases in LBSM, i.e. to study the birth record for children that subsequently die. There is no unique identifier for the children in the dataset as they are not LS members, but an effective match can be made using the mother's identifier (CORENO) in conjunction with the child's sex (SEXABIM), birthweight (BIWGRIM), complete date of birth (BIDYAIM, BIMTCDIM, BIYRCDIM) and whether singleton, twin etc. at birth (MLCSAIM). Of

course, some children will have been born outside England and Wales and therefore will not have a record in LBSM.

Key Documents

- Bethune, A. et al. (1992) 'Fertility and infant mortality in the OPCS Longitudinal Study' Population Trends (68) Summer 1992 pp24-29
- Botting, Rosato & Wood 'Teenage mothers and the health of their children' Population Trends (93) Autumn 1998 pp19-28
- Corbin, T. (2004) 'Mortality in children aged under 8' Health Statistics Quarterly (24) Winter 2004 pp30-37
- Leon, D.A & Macran, S. (1995) 'Infant mortality and maternal circumstances in childhood: the OPCS Longitudinal Study of England and Wales, 1971-1981' LS Working Paper 75 (1995)
- Hattersley, L. and Creeser, R. Longitudinal Study 1971-1991: History, organisation and quality of data. LS Series no. 7

[See the Resources section of the CeLSIUS website for links to some of these documents.]

Quality tables

Sampling fractions and linkage rates for infant deaths 1976-2016

Year of death	LS infant deaths	England and Wales infant deaths	Sampling Fraction	Infant deaths expected in the LS	Linkage Rate
1976-81	369	41,267	0.89	451	81.82
1981-91	608	61,234	0.99	668	90.95
1991-2001	445	39,188	1.14	429	103.62
2001-2011	332	31,328	1.06	346	95.89
2011-2016	164	15,384	1.07	169	97.28

Source: Infant Mortality to sample mother 1976-2016
Office for National Statistics Longitudinal Study

Number of Events

Table 3: Number of infant/child deaths to sample mothers (1971-2016) [see Table 3 in Events Data Tables (PDF)]

Births of New LS Sample Members (NBIR)

Years for which data are available: Census Day 1971 to the current year minus two.

Can the LS member have more than one record: No.

Number of events a year: 5,900 to 8,200.

Means of linkage: Annual date of birth search of the ONS birth file.

Source: Birth registration and birth certificate.

Basic description

New births are babies born on one of the four LS birthdays, who therefore become LS sample members.

Every live birth in England or Wales must be registered within six weeks and a database of these registrations is kept at ONS Titchfield. At regular intervals, details of births on LS birth dates are extracted from this database, allocated LS identification numbers and sent to the NHS Central Register. There, each ONS record is matched with the infant's NHS Central Register record and the latter is flagged as an LS member record. (Infants whose birth took place in England or Wales but where the mother's usual address is outside these countries are not eligible for the LS.)

At the same time in ONS Titchfield, a temporary record for each infant is created in the LS database; if they are successfully traced in NHS Central Register, the infant is added permanently to the LS sample, with a record of the birth in the NBIR table and a central record in the CORE1 table.

Important variables in NBIR

[see the NBIR table in the Data Dictionary for more information on these variables and a full list of the variables available]

Variable name	Label and comments
CORENO	LS Member unique identifier (i.e. of the mother)
BIYRANB	Year of new birth (on an LS birth date). This is invariably the same as DOBYR in the CORE1 table, since the latter represents the year of birth given at entry to the LS.
MLCSANB	Indicator of multiple birth (e.g. a twin). NB: it is not possible to link records of multiple births, e.g. to study both twins as a pair
BIWGRBNB	Birth weight in grams (from 1977 onwards; not recommended for use before 1980, when the percentage of birth records with birth weight reached 86.8%. From 1981 to 2006 it was more than 94% each year.)
MOAGYRNB	Age in years of mother
INSTBNB	Where birth took place (1971-1992; from 1993 on, replaced by ESTYPBNB)
LEGANB	Whether infant's parents were married at time of registration. From 1986, the variable includes additional codes which indicate whether the registration was by one or both parents, and if both, whether they lived at the same address.

Variable name	Label and comments
FAAGYRNB	Age in years of father; can be used to indicate whether registration of birth includes a father, as this variable will be coded -9 if it does not (from 1986, LEGANB is a more convenient way).

Other variables available include geographical classifications for place of birth and parents' usual address. They also include socio-economic indicators for each parent, based on occupation (until 1986 occupation was only collected for the 'working parent', usually the father); however, recording mother's occupation is not obligatory, and although all cases are coded from 1986, it is thought that many mothers in regular employment are coded as having no occupation. This of course affects the reliability of women's social class coding.

Key Documents

The most important documentation relating to the births of new LS members is to be found in:

- Hattersley, L. and Creeser, R. Longitudinal Study 1971-1991: History, organisation and quality of data. LS Series no. 7, pp. 25, 94-95, 109-114

[See the Resources section of the CeLSIUS website for a link to this document.]

Analysis issues:

After the 1981 Census, a number of new births were added to the LS sample; these children were shown at the census as present in the mother's household and having an LS birth date, but had not been identified from the birth registration as the date of birth stated there was not an LS birth date. It is doubtful whether these sample members should be included in analysis; code 8 in TRACE81 (in the CORE1 table) will identify them.

New births should be used when following cohorts over a period of time, when it is necessary to add sample members born between censuses. It is probably unwise to try and use them for fertility studies as there will only be limited information about the mother; the preferred source would be to use the Live and Still Births to Sample Mothers tables (LBSM and SBSM respectively).

Data Quality

Virtually all live births in England and Wales are registered, and all registered births on LS birth dates (where the usual address of the mother is in England or Wales) enter the LS. It is unlikely that a new birth will not be successfully traced in the NHS Central Register as the NHS Central Register records are created from the same sources (electronic returns from Register Offices) as the ONS records. The coverage quality is therefore excellent.

The usual measures of quality, e.g. sampling fraction, will suggest a slightly less satisfactory picture (although still very good) but this is due to measurement methods; for example, births are unequally distributed between seasons and between days of the week, and this can lead to a difference between the 'actual' and 'expected' numbers of LS births.

Researchers who need to study the child's environment in more depth will wish to look at the child's first census record, and here the data become less complete. In the 1981 Census, 7.2% of LS members born in the previous decade were not found (and were not known to have died or emigrated). This percentage rose to 8.7% for children born in the decade before the 1991 Census and rose again to 14.3% for children between the 1991 and 2001 Censuses. Children not found in the 2001 Census were disproportionately more likely to have been born in London, to have very young parents, to have no father mentioned on the birth certificate, to have a parent in the armed forces or to have a parent born overseas. It is likely that young children not found in any census have been taken abroad by their parent(s), who may be in the Armed Forces or not long-term residents of England or Wales; such migrations are only recorded in the LS if the migrant's GP is notified, which commonly does not occur.

Sampling fractions and linkage rates for new births 1971-2016

Year of birth	LS births	England and Wales births	Sampling Fraction	Births expected in the LS	Linkage Rate
1971-81	71,887	6,390,481	1.12	70,461	102.02
1981-91	73,005	6,660,260	1.1	72,606	100.55
1991-2001	70,548	6,511,226	1.08	71,217	99.06
2001-2011	72,845	6,572,218	1.11	72,593	100.35
2011-2016	44,607	4,070,889	1.10	44,612	99.99

Source: Live births in England and Wales 1971-2016
Office for National Statistics Longitudinal Study

Number of events

Table 4: Number of New Births of LS Members in each year (1971 to 2016) [see Table 4 in Events Data Tables (PDF)]

Deaths of LS Sample Members (DETH)

Years for which data are available: Census Day 1971 to the current year minus two.

Can the LS member have more than one record: No.

Number of events a year: 5,700 to 6,700.

Key identifier variable: CORENO

Means of linkage: Annual date of birth search of the ONS death file and flags at NHS Central Register.

Source: Death registration and death certificate.

For more information: See the Mortality guidance.

Basic description

A death must be registered within 42 days. Information from the death certificate and the death registration form is included in the ONS Deaths database. Additionally, a shorter version is sent to the NHS Central Register. The ONS Deaths database is scanned for deaths of people with LS birthdays; notification is also sent to ONS of deaths to people who are flagged in the NHS Central Register as LS members. Thus, the LS has two sources of deaths notification and these combine to give very high data quality.

Neo-natal deaths have a separate range of variables to deaths at other ages.

A doctor can record more than one cause of death on the death certificate. The causes of death are coded using the World Health Organization [ICD](#) (International Classification of Diseases) by the Mortality branch at ONS. Three versions are used in the LS (ICD8, ICD9 and ICD10). There is a procedure in the ICD to identify one underlying cause of death. The remaining causes are called contributory causes. Up to eight contributory causes of death may also be recorded for each death. However, there is a gap in these entries from 2nd December 1986 to 31st December 1992 when these were not coded and these values are not recoverable.

The appropriate Cause of Death (COD) LS variables for each ICD Classification are shown in the table below.

Cause of Death LS variables by ICD version (versions 8-10) [See the DETH table in the Data Dictionary for further information on these variables.]

	ICD8	ICD9		ICD10
	1971-4/4/1981	1/1/1979-31/12/1992		1/1/2001 onwards
Underlying COD - 1st 3 digits	UCDE3CDE	UCDE3DDE	ICD10 code - Underlying COD	IC10UDE
Underlying COD - 4th digit	UCDEXCDE	UCDEEXDDE	ICD10 code - Final Underlying COD	IC10UFDE
		1/1/1993-2001		
Underlying COD - 4th digit		ICD9UDE		
Final underlying COD - 4th digit		ICD9UFDE		

A new year of death is added to the LS database each year. This is done by year of registration with a two-year delay (so deaths registered in 2016 were added in 2018). It is important to note that some deaths that occurred in, for example, 2016, but were registered in 2017, may not be added to the LS database until 2019.

Important variables in DETH

[see the DETH table in the Data Dictionary for more information on these variables and a full list of the variables available]

Variable name	Label and comments
CORENO	Unique identifier
DEYRBDE	Year of death
DEMTBDE	Month of Death (Note: special permission to access needed)
DEDYBDE	Day of Death (Note: special permission to access needed)
INSTADE	Where birth took place (1971-1992; from 1993 on, replaced by ESTTYPDE)
UCDE3CDE	Underlying cause of death code - ICD8 (1971 - 1981)
UCDE3DDE	Underlying cause of death code - ICD9 (1981 - 1992)
ICD9UDE	Underlying cause of death code - ICD9 (1993 - 2000)
ICD9UFDE	Final underlying cause of death code - ICD9 (1993 - 2000)
ICD10UDE	Underlying cause of death code - ICD10 (2000 onwards)
IC10UFDE	Final underlying cause of death code - ICD10 (2000 onwards)

Other variables available include geographical classifications for place of birth and parents' usual address. They also include socio-economic indicators for each parent (see the Socio-economic indicators guidance)³.

³ Parent's usual address and socio-economic indicators are only coded for children <15 years old.

Key Documents

- World Health Organisation International Statistical Classification of Diseases and Related Health Problems 10th Revision, version for 2016
- Guide to Presenting and Using ICD-10 Mortality data
- Fox, A. J. 1971-77 Longitudinal Study: Social Class and Occupational Mortality. LS Series No. 2 (1982)
- Fox, A. J. and Goldblatt, P. 1971-75 Longitudinal Study: Socio-demographic Mortality Differentials. LS Series No.1 (1982)
- Goldblatt, P. Mortality and Social Organisation. LS Series No. 6 (1990)
- Britton, M. Mortality and Geography: Decennial Supplement HMSO (1990)
- Trends in life expectancy by social class 1972-2005. Health Statistics Quarterly 2007; 36(4):3
- Scott, A. and Johnson, B. Social Mortality rates by social class by age and sex 2007
- Trends in social class differences in mortality by cause 1986 to 2000. Health Statistics Quarterly 2003; 20(4): 25-37
- Longitudinal Study 1971-1991: History, organisation and quality of data Hattersley L, Creeser R (1995). LS Series no 7

[See the Resources section of the CeLSIUS website for links to some of these documents.]

Analysis issues

There are differences between the three ICD classifications used in the LS; in particular, there were changes in coding and in rule interpretation between ICD9 and ICD10. This means that trend analysis may not be continuous for some conditions (e.g. pneumonia) and a break in the line where the change from ICD9 to ICD10 occurs should be made unless correction factors have been applied.

As already mentioned, there is a gap in the recording of contributory causes in the LS from 1987 to 1992. This is a particular problem when carrying out cause-specific mortality analysis and the cause of death is often recorded as a contributory cause. For some causes, up to half the cases may be lost during this period.

Where there was an inquest or post mortem, the cause of death may be altered and a new death certificate issued. These Final Underlying Cause of Death codes are only present if there was a change in code from that originally assigned. They need to be added into the cause of death variables for analysis. There are separate variables for the Final Underlying Cause of Death.

The DETH file contains date of death; it is therefore possible to carry out survival analysis using person years at risk. One of the earliest uses of the LS was occupational mortality. By employing the occupational information recorded at the census (from the members' files), you have occupation for the numerator (deaths) and denominator (people at risk) in order to create a rate. This also overcomes the known unreliability of occupational information on death certificates. An example of the use of occupational information in mortality analysis is life expectancy by social class. Life expectancy

by social class for different periods has been produced over a number of years by ONS using the LS (see the Resources section of the CeLSIUS website for links to some of these documents).

Deaths of LS members that were registered in Scotland or the Isle of Man are also included in the DETH file, but those registered in Northern Ireland are not and therefore some LS member deaths may have been missed. A further issue is that deaths overseas will only be included on the LS when and if the NHS Central Register is notified of the death. This may not occur for a considerable time (or never) if the subject is not returned to the UK.

Data quality

Sampling fractions and linkage rates have increased over time and the linkage rate for 2005 was 104.24. Deaths are captured in the LS by two methods: an annual date of birth search of the ONS deaths file and by flags at NHS Central Register. The linkage rates are therefore very high. If a person gives their date of birth incorrectly at the entry event to the LS, the record will be retained in the LS and their death will be attached when they die.

Sampling fractions and linkage rates for Deaths 1971-2016

Year of birth	LS deaths	England and Wales deaths	Sampling Fraction	Deaths expected in the LS	Linkage Rate
1971-1981	60,990	5,821,191	1.05	64,080	95.18
1981-1991	64,734	5,781,244	1.12	63,049	102.67
1991-2001	64,782	5,590,707	1.16	61,113	106.00
2001-2011	61,316	5,075,095	1.21	56,107	109.28
2011-2016	35,313	2,932,490	1.20	32,137	109.88

Source: Deaths of LS members in England and Wales 1971-2016
Office for National Statistics Longitudinal Study

Number of events

Table 5: Number of deaths to LS Members each year 1971 to 2016 [see Table 5 in Events Data Tables (PDF)]

Widow(er)hoods to LS Sample Members (WDOW)

Years for which data are available: Census Day 1971 to the current year minus two.

Can the LS member have more than one record: Yes.

Number of events a year: around 2000.

Key identifier variable: CORENO+EVENTORD.

Means of linkage: Annual search of the ONS death file to identify LS birthdays in the date of birth of surviving spouse - thus showing that an LS sample member has been widowed.

Source: Death registration and death certificate.

Basic description

It is possible to identify the death of a spouse of a sample member by the date of birth of the surviving spouse in the death registration. This date of birth will only be recorded at death registration if the couple is legally married.

The date of birth of the husband is always recorded. However, up to the mid-1990s the name and date of birth of a surviving wife was not recorded at death registration unless she was the informant, which meant that some widowhoods were not recorded in the LS. From the mid-1990s data capture has steadily improved.

Cause of death information is included in the widow(er)hood record in the LS, but as with death to sample members, there is a gap in contributory causes of death for the period 1987 to 1992. Date of death of the spouse is also contained in the dataset.

Important variables in WDOW

[see the WDOW table in the Data Dictionary for more information on these variables and a full list of the variables available]

Variable name	Label and comments
CORENO	Unique identifier
DEYRBWI	Year of death of LS member's spouse
DEMTBWI	Month of death of LS member's spouse
DEDYBWI	Day of death of LS member's spouse
AGDC3DWI	Age at death of LS member's spouse
AGS03FWI	Age of LS member at death of spouse (years)
<i>Cause of death: Underlying cause of death variables</i>	
IC10UWI	ICD10 Underlying COD of spouse of LS member (not before 2001)
ICD9UWI	ICD9 Underlying COD of spouse of LS member (1993 - 2001)
UCDE3DWI	ICD9 Underlying COD of spouse of LS member (1979 - 1992) 3 digit
UCDE3CWI	ICD8 Underlying COD of spouse of LS member (1971 - 1981) 3 digit
<i>Cause of death: Other Cause of Death Variables:</i>	

Variable name	Label and comments
IC101WI-IC108WI	ICD10 codes for multiple cause of death (not before 2001)
ICD91WI-ICD98WI	ICD9 codes for multiple cause of death (1993-2001)
MLC941WI-MLC948WI	ICD9 codes for multiple cause of death (1981-1986)
MLC841WI-MLC848WI	ICD8 codes for multiple cause of death (4/1971-4/1981)

Key Documents

- Jones, D.R. and Goldblatt, P. Cancer and mortality following widow(er)hood. Some further results from the Office of Population Censuses and Surveys Longitudinal Study. *Stress Medicine* 1986; 2: 129-140.
- Jones, D. R. et al. Heart disease mortality following widowhood: some results from the OPCS Longitudinal Study. *J. Psychosomatic Research* 31:3 325-33, 1987.
- Grundy E. Divorce, widowhood, remarriage and geographic mobility among women. *Journal of Biosocial Science* 1985; 17 (4): 415-435.
- Hattersley, L. and Creeser, R. Longitudinal Study 1971-1991: History, organisation and quality of data 1995. LS Series no. 7.

[See the Resources section of the CeLSIUS website for links to some of these documents.]

Analysis issues

Some LS members have been widowed more than once and therefore will have more than one record in the file. The variable EVENTORD can be used to establish date order and to generate a dataset with one record per LS member.

Data relating to occupation are known to be unreliable on death registers. For elderly people, the deceased may have been retired for a number of years, and the informant may not know the exact occupation of the deceased.

The LS widowhood records only cover people who were legally married. It is not possible to analyse the deaths of partners or cohabitants.

Quality tables

The linkage rates for widowhoods have increased each decade and the rate for 2013 is 95.87. Sampling fractions are below 1.1. The overall number of widow(er)hoods is falling in England and Wales (see table below). The rates vary by sex; in the 1970s and 1980s, the linkage rates were lower for widowers.

Sampling fractions and linkage rates for Widow(er)hoods

Year of widow(er)hood	LS widow(er)hoods	England and Wales widow(er)hoods	Sampling Fraction	Widow(er)hoods expected in the LS	Linkage Rate
1971-81	22,297	2,638,998	0.84	29,069	76.70
1981-91	22,105	2,428	0.91	26,472	83.50
1991-2001	15,579	1,553,095	1.00	16,978	91.76
2001-2011	19,812	1,919,291	1.03	21,220	93.36
2011-2016	11,470	1,106,870	1.04	12,130	94.56

Source: Widow(er)hoods to LS members in England and Wales 1971-2016
Office for National Statistics Longitudinal Study

Number of events

Table 6: Number of Widow(er)hoods to LS Members each year 1971 to 2016 [see Table 6 in Events Data Tables (PDF)]

Cancer Registrations to LS Sample Members (CANC)

Years for which data are available: Census Day 1971 to the current year minus two⁴.

Can the LS member have more than one record: Yes.

Number of events a year: 1200 to 4500

Key identifier variable: CORENO + EVENTORD.

Means of linkage: Annual date of birth search of the ONS cancer file and flags at the NHS Central Register

Source: Eight English Cancer registries and Welsh Cancer Intelligence and Surveillance Unit.

Basic description

Cancer registrations are collected by the eight English Cancer Registries and the Welsh Cancer Intelligence and Surveillance Unit. Since 1995, ONS has only published incidence rates for England. Only information on primary cancers is kept in the LS, although some secondaries and metastases are known to be recorded. Users of the LS must remember that an individual may have more than one cancer registration and must take this into account during the analysis. It will be essential to get the registrations in the correct order.

The cancer file is a dynamic file, with late registrations being added, dates of diagnosis changing and some information being deleted. Cancer registration only became mandatory in 1993. Data in the LS may differ from published sources, because of the changes to the data over time. Cancer registrations are identified by an annual search of the cancer file and by flags at NHS Central Register.

Information is collected for all malignant neoplasms, reticuloses, carcinoma in situ, neoplasms of uncertain behaviour, benign neoplasms and hydatidiform moles. Information available includes date of diagnosis, type of cancer, whether a multiple cancer, treatment type and grade/stage. The date of registration can be created from the variables (AVDDYACC, AVDMTACC and AVDYRACC) [see the CANC table (Cancer registrations) in the [CeLSIUS Data Dictionary](#)].

Three classifications of the International Classification of Disease (ICD) are used in the LS. The appropriate LS variables for each classification are listed in the table below.

⁴ August 2019: Since 2016 the Cancer Registry has been managed by the National Cancer Registration and Analysis Service (NCRAS) within Public Health England, A data sharing agreement with PHE for the linkage of cancer registrations has not yet been put in place and therefore cancer registrations after 2015 have not been received by the ONS LS.

ICD variables for cancer registrations by ICD version (versions 8-10)

	ICD8	ICD9	ICD10
Years coded to ICDx	1971-1978	1979-1994	1995 onwards
Chapter headings	140-208	140-208	C00-C97
Site of growth variable	ICCS3BCC	SIDC3CC	SICD30CC
Sub-site of growth variable	ICCSXBCC	SICD1CC	SICD10CC

Important variables in CANC

[see the CANC table in the Data Dictionary for more information on these variables and a full list of the variables available]

Variable name	Label and comments
CORENO	Unique identifier
EVENTORD	Event order for this LS member
AVDDYACC	Day of 1st cancer (anniversary day)
AVDMTACC	Month of 1st cancer (anniversary day)
AVDYRACC	Year of 1st cancer (anniversary day)
EDATECC	Event date
AGACC	Age at registration of cancer - 1st cancer (1971-1989)
AGDIAGCC	Age at diagnosis (1993 onwards)
BEHA10CC	Behaviour of growth - ICD-02
BEHAVCC	Behaviour of growth - ICD-0
GRADECC	Grade of differentiation of tumour cells (1993 onwards)
ICCS3BCC	Site of growth (ICD8)
ICCSXBCC	Subsite of growth (ICD8)
NOREGCC	Number of registrations (1993 onwards)
PLBICC	Expanded place of birth code
REYRBCC	Registration details - registration year
SICD10CC	Subsite of growth (ICD10)
SICD1CC	Subsite of growth (ICD9)
SICD30CC	Site of growth (ICD10)
SIDC3CC	Site of growth (ICD9)
SIDECC	Side of growth (1993 onwards)
TICD10CC	Histology of growth (ICD-02)
TICDCC	Histology of growth (ICD-0)
TRTCHECC	Treatment type - Chemotherapy (1993 onwards)
TRTHORCC	Treatment type - Hormonal (1993 onwards)
TRTOTHCC	treatment type - Other (1993 onwards)
TRTRADCC	Treatment type - Radiotherapy (1993 onwards)
TRTSURCC	Treatment type - Surgery (1993 onwards)
TY1ACC	Behaviour of growth (to 1/1/1979 only)
TY3ACC	Histology of growth (to 1/1/1979 only)

Key Documents

- Donkin, A. & Hattersley, L. Using the Longitudinal Study cancer data for research LS User Guide 19
- Quinn, M. et al. Cancer trends in England and Wales 1950-99 Studies on Medical and Population Subjects No.66, The Stationery Office (2001)
- Cancer Registrations Statistics, England 2016
- Cancer Incidence in Wales 2003-07 (Welsh Cancer Intelligence and Surveillance Unit)
- Sloggett A, Young H, Grundy E. The association of cancer survival with four socioeconomic indicators: a longitudinal study of the older population of England and Wales 1981-2000 BioMed Central Cancer 2007; 7 (20)
- Hattersley, L. & Creaser, R. Longitudinal Study 1971-1991: History, organisation and quality of data LS Series no 7, 1995

[See the Resources section of the CeLSIUS website for links to some of these documents.]

Analysis issues

It is possible to perform analysis on cancer incidence (the rate of new cases), cancer mortality (mortality rates for different cancers/all cancers), and cancer survival (length of time someone survives after a diagnosis). However, it may only be possible to look at the more common cancers separately for men and women, because of small numbers.

There are differences in the codes for types of cancer between ICD9 and ICD10. The coding in the LS changed from ICD9 to ICD10 in 1995. Care should be taken when using time periods that cover the two classifications. Users should also note that deaths are coded using ICD10 from 2001, which will affect the coding of cancer deaths.

Data quality

Readers are referred to the [Cancer registration statistics, England](#) pages on the ONS website for the latest available national information. Separate information for the LS is not available.

Number of events

Table 7: Number of Cancer registrations to LS Members each year 1971 to 2016 [see Table 7 in Events Data Tables (PDF)]

Summary

This guidance has introduced researchers to the vital events data available in the LS. The most commonly used events included:

- Live or still births to sample mothers
- Infant deaths
- Births, deaths and widow(er)hoods of sample members
- Cancer registrations

The guidance has also described the key variables used in the analysis of events data and identified a few of the issues associated with their use.

As well as the information contained in this guidance, you could also look at the following sources on the [CeLSIUS](#) website:

- The Research section of the CeLSIUS website will show you what has been done before.
- The Mortality guidance gives more information on deaths and widowhoods (to be available in the future).
- There may be existing projects using these events data, which you can see in the Current LS projects section of the CeLSIUS website.

The [CeLSIUS Data Dictionary](#) will give you more information about the variables for a particular event.