Quality of ONS Longitudinal Study (LS) Event Linkage 1971-2013

Three types of events are collected for the LS:

- Entry events causing the entry of new members to the study (birth on LS dates or immigration)
- Exit events causing exits from the study (death of LS members or emigration)
- Events occurring to LS members while alive and active in the study (birth of children, cancer registration, widow(er)hood)

The quality of events data within the LS depends on two components: firstly, on the quality of the original event data collected for the England and Wales population; secondly, on the quality of the sampling of these events and of their linkage into the LS.

Methods

Events sampling and linkage quality are measured by sampling fractions (Figure 1) and linkage rates (Figure 2). By definition, entry events (Figure 3) cannot be linked to existing LS members, so an entry rate is used for them rather than a linkage rate.

Figure 1: Calculating sampling fractions

$$\text{Sampling fraction} = \frac{\text{Events occurring to LS members in the calendar year}}{\text{Events occurring to the E\&W population in the calendar year}} \times 100$$

Figure 2: Calculating linkage rates

$$\text{Linkage rate} = \frac{\text{Events occurring to LS members in the calendar year}}{\text{Events expected to occur to the LS population in the calendar year}} \times 100$$

Figure 3: Calculating entry rates into the LS

$$\text{Entry rate} = \frac{\text{New members entering the LS in the calendar year}}{\text{Expected number of new members entering the LS in the calendar year}} \times 100$$

Because the LS is a sample of around 1.1 % of the population of England and Wales, sampling fractions for each event should be close to this figure. If they are larger than 1.1 per cent, then the LS is "oversampling" that event; if they are smaller, there is "undersampling". Ideally, linkage and entry rates should be as close as possible to 100.

Sampling fractions and linkage/entry rates are normally calculated for each calendar year. The exception to this is during census years, when estimates are derived separately for two periods: from the start of the year up to the day before census night (pre-census) and from the day after census night until the end of the year (post-census). The formulae (Figures 1, 2 & 3) are then applied using the number of events occurring or expected to occur in these two separate parts of the census year (see Figure 6).

While the number of actual events occurring to the LS members in the 'pre' and 'post' census periods of census years can be accurately calculated, national figures of number of events occurring to the England and Wales population are normally published only on a calendar year basis. National figures for the pre and post-census night periods are estimated by apportioning the England and Wales calendar year figures. The apportionment factors used are the number of days preceding and following the census (see Table A). Figure 4 gives the number of events for the part of the year before and after census day.

Figure 4: Calculating the number of events in the first and second part of the year

$$\text{Events occurring to the E\&W population in first part-year} = \frac{\text{Number of days in year before census (including census day)}}{365} \times \text{Events occuring to the E\&W population in the calendar year}$$

$$\text{Events occurring to the E\&W population in second part-year} = \frac{\text{Number of days in year after census (including census day)}}{365} \times \text{Events occuring to the E\&W population in the calendar year}$$

Table A shows the Census dates and the number of days pre and post census for the five censuses included in the LS.

Table A. Census dates and number of days in each part of the year

Year	Census night between	Days in first part-year*	Days in second part-year**
1971	25th and 26th April	115	250
1981	5th and 6th April	95	270
1991	21st and 22nd April	111	254
2001	29th and 30th April	119	246
2011	27th and 28th March	86	279

Notes:

* Number of days from 1st January up to and including census day.
** Number of days from the day after census day up to and including 31st December.

The LS includes linked census and events data for people born on one of four selected dates in a calendar year. Assuming an even distribution of events (births, deaths, births to sample mothers (BSM), widow(er)hoods, infant deaths) across the days of the year, the LS basically samples 4/365 events in non-leap years and 4/366 events in leap years. The expected numbers of events into the LS are given by Figure 5; the example shown is for new births into the LS but can be substituted for deaths, births to sample mothers (BSM), widow(er)hoods, infant deaths. Figure 6, shows the expected number of LS events during a census year (pre-census and post-census period).

Figure 5: Calculating the number of expected LS events (births, deaths, BSM, widow(er)hoods, infant deaths)

$$\text{Expected number of new births into the LS (non-leap years)} = \frac{\text{Total births in E\&W in the calendar year}}{365} \times 4$$

$$\text{Expected number of new births into the LS (leap years)} = \frac{\text{Total births in E\&W in the calendar year}}{366} \times 4$$

Figure 6: Calculating the number of expected LS events (births, deaths, BSM, widow(er)hoods, infant deaths) in census year

$$\text{Expected number of new births into the LS (first part of year)} = \frac{\text{Total births in E\&W in the part-year before census night}}{\text{days in part year}} \times 1$$

$$\text{Expected number of new births into the LS (second part of year)} = \frac{\text{Total births in E\&W in the part-year after census night}}{\text{days in part year}} \times 3$$

Users should note that the methods to calculate the expected number of events and the census year apportionment factors have been revised from previous publications. All linkage and entry rates published here have been derived using the new method.

New births into the Longitudinal Study (LS)

In England and Wales all live births must be registered within six weeks of the birth occurrence by law. New births are entered into the study directly from birth registration data and entry rates are expected to be very close to 100%, the quality of births data for the whole of England and Wales is known to be extremely high.

All individuals born in England and Wales on one of four dates of birth enter as "new births" into the ONS Longitudinal Study (LS). The LS includes linked census and events data for people born on one of four dates in a calendar year. The LS basically samples 4/365 births in non-leap years and 4/366 births in leap years, using the formula in Figure 5. In census years, Figure 6 is used which allows us to derive one pre and one post-census estimate of expected LS births.

Sampling fractions and entry rates by sex for new births occurring since the 1971 Census are shown in the download table. Part years are shown to reflect the occurrence of a census. Between 2001 Census day and the end of 2013, the overall achieved sampling fraction for the LS was very close to 1.1. Annual sampling fractions are more variable. This happens because births are actually not equally distributed throughout the year or by day of the week. The number of new births of LS members and sampling fractions may vary from year to year. This impacts on the observed variability of the annual entry rates around the 100 target, given that the expected number of LS events used as denominator of the rate assumes an even distribution of births. Over a period of years, differences tend to even out, resulting in a total entry rate very close to 100 for 2001-2013.

While all new birth entrants to the LS must be resident in England and Wales, the estimates for England and Wales include a small number of births to mothers who are not usually resident in England or Wales. This would appear as an undersampling in the LS.

Deaths of Longitudinal Study (LS) Members

Death certificates are required by law before the burial or cremation of a body, and as a result, virtually all deaths occurring in England and Wales are registered and the quality of death data for England and Wales is known to be extremely high. However, delays in certification can occur if an inquest is required or a person died while abroad. Deaths may also be missed if the death is not registered in England and Wales.

Deaths are linked into the ONS Longitudinal Study (LS) both through routine notification of deaths of traced LS members from the National Health Service (NHS) registration systems and through a supplementary search for deaths of people born on an LS date. The use of these two different methods to identify deaths of LS members greatly improves the linkage rates. The accuracy of the linkage is dependent on the quality of data supplied by informants and discrepancies in the dates of birth affect around 8% of all LS members.

The LS includes linked census and events data for people born on one of four selected dates in a calendar year. The number of expected deaths occurring to LS members in any calendar year will be given by using the formula in Figure 5. In census years, the formula in Figure 6 is used which allows us to derive one pre and one post-census estimate of expected LS deaths.

The total numbers of deaths of LS members by age and sex since 1971 Census are shown in the download table together with the relevant sampling fractions and linkage rates.

Births to LS sample mothers (live and still births)

Births to ONS Longitudinal Study (LS) women are generated from England and Wales birth registration data by using confidential particulars collected on the birth registration draft, which is completed by the informant (normally one of the parents) when registering either a live or a stillbirth. The informant is asked to state the mother's date of birth; if this is one of the LS dates, the birth registration is then matched to the LS mother.

If there are differences between the mother's birth date given at the census or at immigration and that given when registering the birth of a child, then further research will be undertaken before linkage can be made between the LS member and the birth.

The LS includes linked census and events data for people born on one of four selected dates in a calendar year. The number of expected births to LS mothers in any year will be given by Figure 5. In census years, the formula used in Figure 6 is applied, this allows us to derive one pre and one post-census estimate of expected births to LS mothers.

The total numbers of live and still births to LS mothers since 1971 Census are shown in the download table together with the relevant sampling fractions and linkage rates.

Widow(er)hoods

Widow(er)hoods or surviving civil partner of ONS Longitudinal Study (LS) members are generated from England and Wales death registration data by using confidential particulars collected on the death registration draft, which is completed by the informant registering a death. The informant is asked whether the deceased was married or in a civil partnership (the Civil Partnership Act 2004 came into force in December 2005); if so, the date of birth of the surviving spouse/civil partner is requested. If the date given for the spouse or civil partner is an LS date, the death registration is then matched to the LS member.

This method of linkage is known to be incomplete. For example, if the informant did not register the deceased as married or did not give a correct LS date of birth for the surviving spouse, no linking to the LS member can be attempted.

The LS includes linked census and events data for people born on one of four selected dates in a calendar year. The number of expected widow(er)hoods occurring to LS members in any calendar year will be given by the formula in Figures 5, depending on whether the death occurred in a leap year or non-leap year. In census years, the formula in Figure 6 is used, this allows us to derive one pre and one post-census estimate of expected LS widow(er)hoods.

Sampling fractions and linkage rates by sex of widow(er)hoods of LS members occurring since 1971 Census are shown in the download table. The sampling fractions and linkage rates suggest that widow(er)hoods are under-represented in the LS. This result is expected to some extent, given the known limitations of the widow(er)hood linking method. As shown linkage rates for widow(er)hoods have consistently improved since the inception of the LS. The linkage rate has increased from just over 77% in the first decade of the LS to just over 94% between 2001 and 2013

Infant Mortality

Infant Mortality data are generated from the ONS Statistical Infant Mortality System. Since 1975, ONS has been producing an infant mortality file which links information from the death records of children who were under the age of one at time of death to the corresponding birth record. From 1993, the age limit has increased gradually to include children under 16 years, but published results here include only infant deaths (i.e. under one year old).

The infant mortality linked file, compiled centrally for all infant deaths in England & Wales, is searched annually for children of LS sample mothers. The extract is then matched to the annual LS sample file of Live Births to LS Sample Mothers. If a corresponding record is found, the infant death is then linked into the study.

The LS includes linked census and events data for people born on one of four selected dates in a calendar year. The number of expected infant deaths for LS mothers in any calendar year (leap year and non leap year) will be given by using the formula in Figure 5. In census years, the formula in Figure 6 is used, this allows us to derive one pre and one post-census estimate of expected LS infant deaths.

Sampling fractions and linkage rates of infant deaths to LS sample mothers occurring since 1971 Census are shown in the download table.

Sampling fractions and linkage rates would suggest a tendency to undersample infant deaths in the LS. However, actual and expected numbers of infant deaths are very small and are therefore more sensitive to sampling variability.