

An Introduction to the Office for National Statistics Longitudinal Study (ONS-LS) data 1971-2021 for England and Wales and the Support Offered by CeLSIUS at UCL

The ONS Longitudinal Study (ONS-LS), follows a 1% sample of the England & Wales population from the decennial census data (1971 – 2021), linked to births, deaths and cancer registration data. Sample members are selected on the basis of four confidential birthdays, with new study members entering the study through birth or immigration with those birth dates and leave through death or emigration. The main strength of the ONS-LS is its large sample size (>1 million), making it the largest nationally representative dataset in the UK, and allowing the analysis of small areas and specific population groups.

The ongoing linkage of the 2021 Census data to the Study will extend this follow-up enabling researchers to examine changes that have taken place in the 2011 – 2021 period, which saw Brexit and the Covid-19 pandemic. The paper introduces the data available in the ONS-LS and will summarise some recent examples of research. The support available from CeLSIUS to researchers interested in using the ONS-LS in their research is also highlighted, as are the arrangements for accessing the data. It will also introduce the ONS-LS's sister studies, the Scottish Longitudinal Study (SLS) and the Northern Ireland Longitudinal Study (NILS), and the opportunities that these studies offer for comparative research.

The three UK Longitudinal Studies comprise:

Linked census data for LS members

Unlinked Census data for other household members.

Vital registration data (“events”), including:

Births and deaths of sample members.

Births of children to sample members.

Widow(er)hood of sample members.

Scotland and Northern Ireland have additional components.

Dynamic samples:

New members enter between censuses: by birth or immigration.

Members leave between censuses: by emigration or death.

Census forms are linked for each year members are enumerated.

The ONS Longitudinal Study (ONS LS) covers England and Wales; 1% sample (selected using 4 birthdates); 1971 onwards.

Scottish Longitudinal Study (SLS) covers Scotland; 5.3% sample (selected using 20 birthdates); 1991 onwards.

Northern Ireland Longitudinal Study (NILS): covers Northern Ireland; 28% sample (selected using 104 birthdates); 1981 onwards.

For UK wide analysis there are two options:

Conduct separate analyses and then compare or combine the results on an ad-hoc basis.

eDatashield – developed by Prof Gillian Raab (SLS). Possible to analyse data from 2+ LSs as though they are part of the same dataset. Supports all forms of GLM regression analyses.

The ONS Longitudinal Study

1% sample of population of England & Wales.

Selected if birthday falls on one of four days.

Data on the individual and their household (including other people in the household e.g. parents, siblings, spouse, children).

Uses information from 1971-2011 censuses.

Individual's census forms from 1971 – 2011 are linked, giving up to 40 years of census data on study members.

Includes geographical location of LS members.

Vital events data linked to LS member record comes from:

Routine notifications, e.g. births of new sample members, deaths of sample members, births to sample mothers, cancer registrations. Linkage rates high: >90%.

Date of birth searches on the NHS Central Registry, e.g. immigrations of sample members, emigrations of sample members (but poorly ascertained).

LS members flagged in the NHS Central Register.

When event (birth, death, cancer registration) recorded in NHS Central Register it is linked to the LS member.

CeLSIUS User Support

Advising on design and feasibility of projects.

Guiding researchers through the application process and submitting applications.

Creating data extracts.

Remote analyses and some statistical guidance.

Disclosure control compliance checks of outputs.

Advice and on-site support.

Information on syntax files that are available for LS researchers to use to derive some variables, e.g. causes of death.

Is the LS the best source of data for my research project?

Background to the LS

Data dictionary of variables

Past and present research projects

Thematic guides provide information on using the LS and the variables include:

Defining a study population

Births and fertility

Ethnicity

Events

Socio-economic indicators

Mortality

Households and families

International migration

Discrepancy variables

Imputation flag variables

Cancer

Data Access

The data is available for free to all researchers but research must be in public interest. Protection of confidentiality of data relating to individuals in LS is crucial. The release is subject to disclosure control.

Individual level microdata is never released. You'll need to hold ONS Researcher Accreditation before using the LS. CeLSIUS will work with you to develop your application. This will involve completing a Research Project Accreditation Application form , an ONS Longitudinal Study Supplementary form and an Ethics Self-Assessment sheet.

Once your project has been approved, ONS' Secure Research Service (SRS) will create a project folder in their online Research Accreditation Service (RAS). A CeLSIUS Support Officer will construct your dataset (which they can amend if necessary as your study proceeds) and add it to the project folder. You will then be provided with account details and guidance on accessing your data securely, which can be done in the following ways.

In the UK: Accessing your data in a SafePod, Connecting to the SRS remotely, Accessing your data in an ONS SRS safe room.

Globally: Sending us your code: You may send code in STATA or R to celsius@ucl.ac.uk for us to run for you.

Census 2021/2

The beta test stage of the project to link Census 2021 data to the Office for National Statistics (ONS) Longitudinal Study (LS) is the final test stage prior to release of the new LS Research Database. During this stage, a select group of researchers will run approved projects using the new data to assess whether the new database is fit for purpose. The beta test stage of the project has two primary objectives, which are to:

- test the 2021 data for accuracy, comparability, and consistency through active analytical use
- use the feedback from test participants to inform the decision whether to release the database to the wider research community

Subject to 2021 data release this paper will include some of the results from the 2021 or 2011 analysis.

The NILS database has been updated with 2021 Census data. The NILS 'core' data are drawn from the Health Card Registration data and then linked to data from Census returns (1981, 1991, 2001, 2011 & 2021), Vital Events (births, deaths and marriages), Migration and Property data. In addition to this rich resource there is the potential to link further Health and Social Care data via Distinct Linkage Projects. This includes Primary health care data (supplied by the Business Services Organisation) and secondary health care data available through the HSC Regional Data Warehouse (via the Honest Broker Service).

The Scottish Longitudinal Study (SLS) includes census data from 1991 onwards; vital events data (births, deaths, marriages); NHS Central Register data (gives information on migration into or out of Scotland); and education data (including Schools Census and SQA data). The census was moved to 2022 due to the impact of the COVID-19 pandemic.