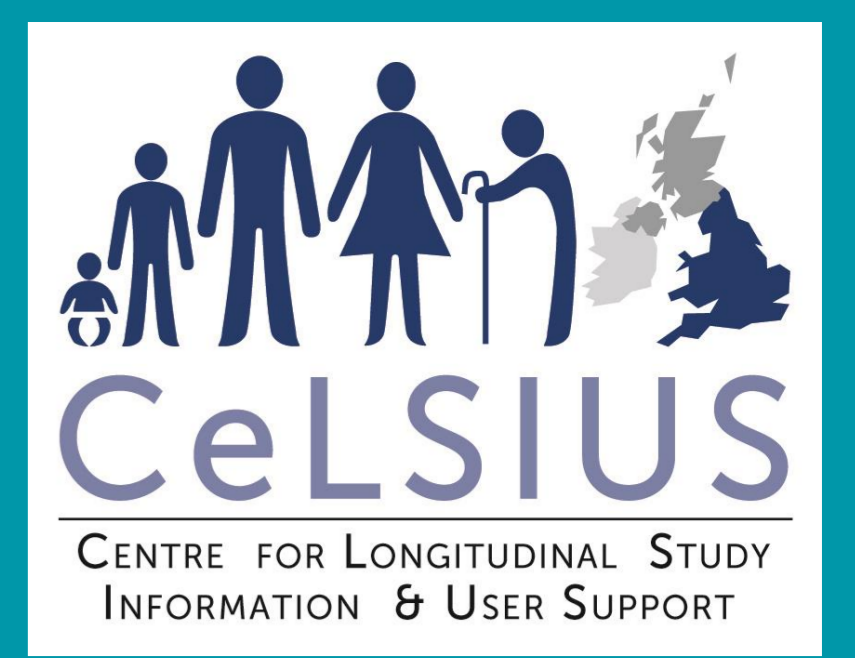
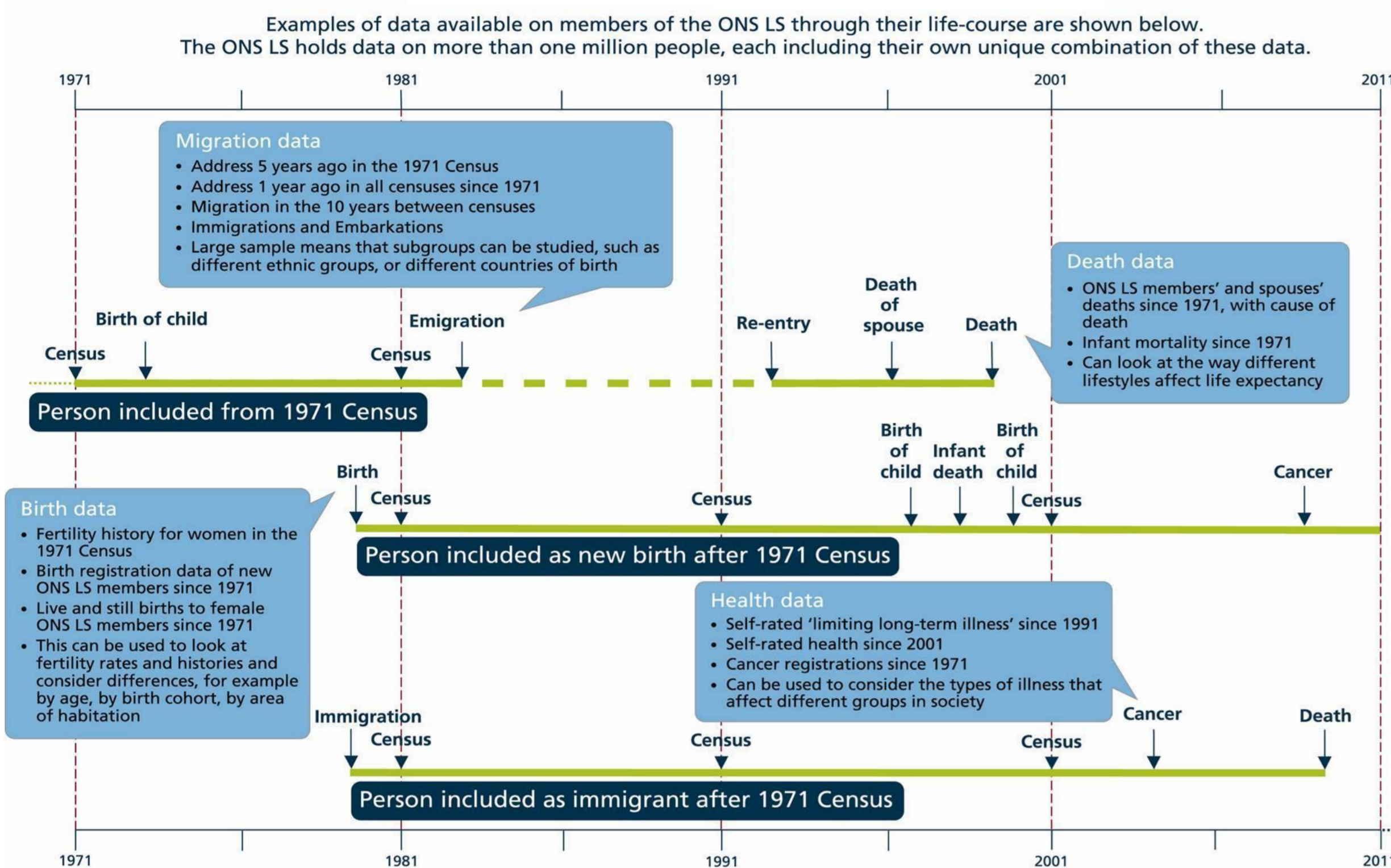


The ONS Longitudinal Study



What is the Longitudinal Study (LS)?

The ONS Longitudinal Study (LS) is a linked set of individual census responses, for a sample of 1% of the population of England and Wales, together with records for other people in their households. Results from the censuses of 1971, 1981, 1991, 2001 and 2011 are included and life event information has been added to the LS since 1971. This includes birth and immigration (entry events) and death and emigration (exit events) of individuals. At each census, data on just over 500,000 sample members are collected. Similar studies also exist in Scotland and Northern Ireland; the three studies differ in a number of ways including the length of time they have been running, the sample fraction, and the amount of additional linked data from non-census sources.



SYLLS – SYNthetic data estimation for the UK Longitudinal Studies

The ONS LS, Scottish LS and Northern Ireland LS are unparalleled resources for social science research in the UK. To maintain confidentiality of the data, it is anonymised at point of use and access is from within a controlled 'safe environment'. SYLLS is a project to generate synthetic versions of the three Longitudinal Studies which look and behave like the real thing, but are not subject to the same access restrictions. It is released under an Open Government Licence and enables researchers to iteratively refine research ideas and update analysis code as well as use data in teaching and expose students to longitudinal data early in their research careers.

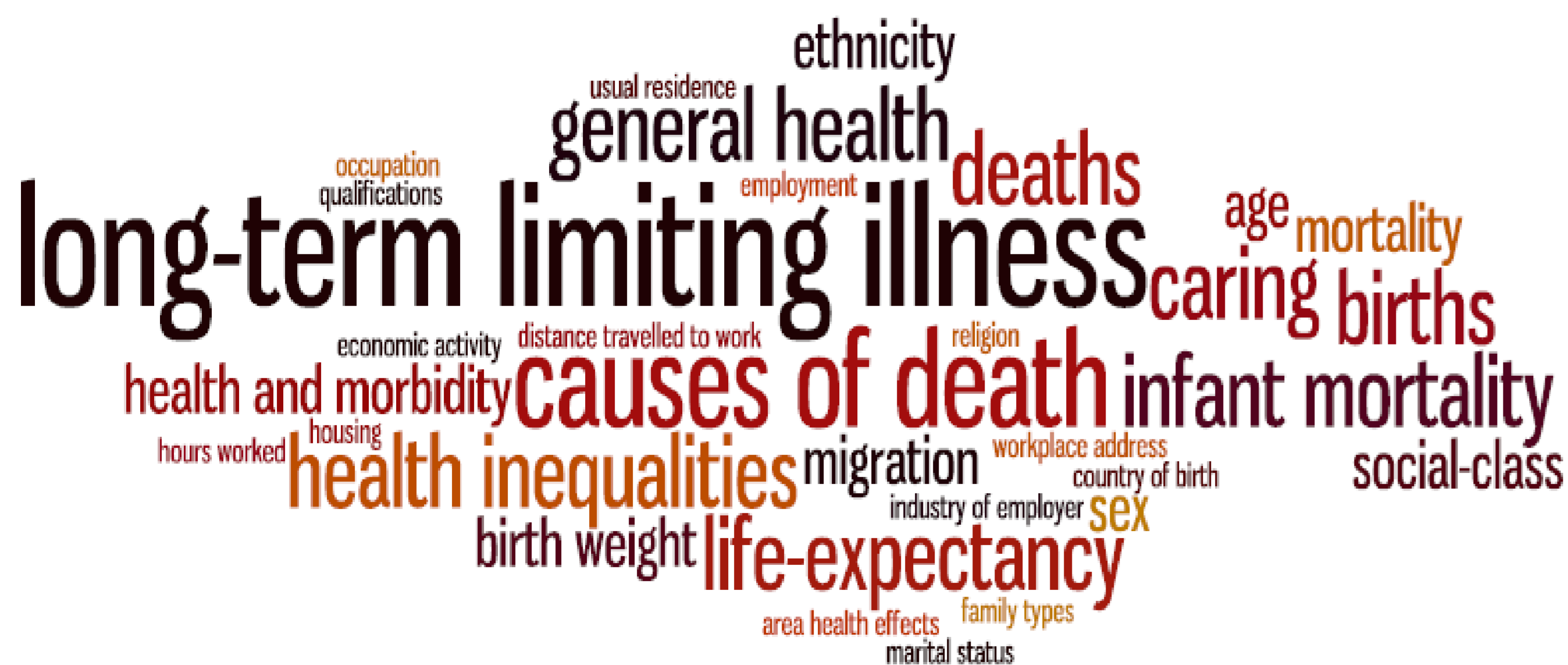
Synthetic Tabulation

| | sex | age | edu | marital | incomenn | ls | wkabin1 | wkabin2ur |
|------------|-------|-----|---------------------------|-----------|----------|---------------------|---------|-----------|
| false data | MAN | 81 | PRIMARY/NO EDUCATION | MARRIED | 1500 | PLEASED | NO | MISS/NA |
| false data | MAN | 54 | VOCATIONAL/GRAMMAR | MARRIED | 1700 | PLEASED | NO | MISS/NA |
| false data | WOMAN | 32 | VOCATIONAL/GRAMMAR | DIVORCED | 870 | MIXED | NO | MISS/NA |
| false data | WOMAN | 61 | PRIMARY/NO EDUCATION | MARRIED | 800 | MOSTLY DISSATISFIED | NO | MISS/NA |
| false data | WOMAN | 50 | PRIMARY/NO EDUCATION | MARRIED | NA | MOSTLY SATISFIED | NO | MISS/NA |
| false data | WOMAN | 37 | VOCATIONAL/GRAMMAR | MARRIED | 158 | PLEASED | NO | MISS/NA |
| false data | MAN | 28 | VOCATIONAL/GRAMMAR | MISS/NA | 1500 | MOSTLY SATISFIED | NO | MISS/NA |
| false data | WOMAN | 62 | PRIMARY/NO EDUCATION | MARRIED | 830 | MOSTLY SATISFIED | NO | MISS/NA |
| false data | MAN | 78 | PRIMARY/NO EDUCATION | MARRIED | 1200 | PLEASED | NO | MISS/NA |
| false data | WOMAN | 29 | SECONDARY | MARRIED | 580 | MOSTLY SATISFIED | NO | MISS/NA |
| false data | MAN | 59 | PRIMARY/NO EDUCATION | MARRIED | 1300 | MOSTLY SATISFIED | NO | MISS/NA |
| false data | MAN | 41 | SECONDARY | UNMARRIED | 1500 | MIXED | NO | MISS/NA |
| false data | MAN | 58 | SECONDARY | MARRIED | -8 | PLEASED | NO | MISS/NA |
| false data | WOMAN | 73 | PRIMARY/NO EDUCATION | WIDOW(ER) | 1350 | MOSTLY SATISFIED | NO | MISS/NA |
| false data | WOMAN | 70 | SECONDARY | WIDOW(ER) | 1313 | MOSTLY SATISFIED | NO | MISS/NA |
| false data | MAN | 54 | VOCATIONAL/GRAMMAR | UNMARRIED | 800 | MOSTLY SATISFIED | NO | MISS/NA |
| false data | WOMAN | 74 | PRIMARY/NO EDUCATION | MARRIED | 1300 | MOSTLY SATISFIED | NO | MISS/NA |
| false data | WOMAN | 48 | POST-SECONDARY AND HIGHER | MARRIED | 1500 | MOSTLY SATISFIED | NO | MISS/NA |

To find out more about SYLLS and download the data: calls.ac.uk/guides-resources/synthetic-ls-data/

Research potential of the LS

With the large sample sizes, range of census variables and annual data on life events (births, deaths, widowhoods and cancer registrations), the list of potential research topics using the LS is large.



Accessing the LS

The CeLSIUS user support team helps researchers to apply for access to the LS by providing advice on whether the LS is appropriate for their study, helping with sample and variable selection and preparing data extracts ready for researchers to analyse.

website: ucl.ac.uk/celsius/
 email: celsius@ucl.ac.uk
 phone: 020 7679 1995

For information about all UK Census Longitudinal Studies: calls.ac.uk

Jo Tomlinson, CeLSIUS, UCL
 With thanks to Nicky Rogers, James Warren, Kevin Lynch, Longitudinal Study Development Team, Office for National Statistics

The permission of the Office for National Statistics to use the Longitudinal Study is gratefully acknowledged. CeLSIUS is supported by the ESRC Census of Population Programme (Award Ref: ES/K000365/1). The authors alone are responsible for the interpretation of the data. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

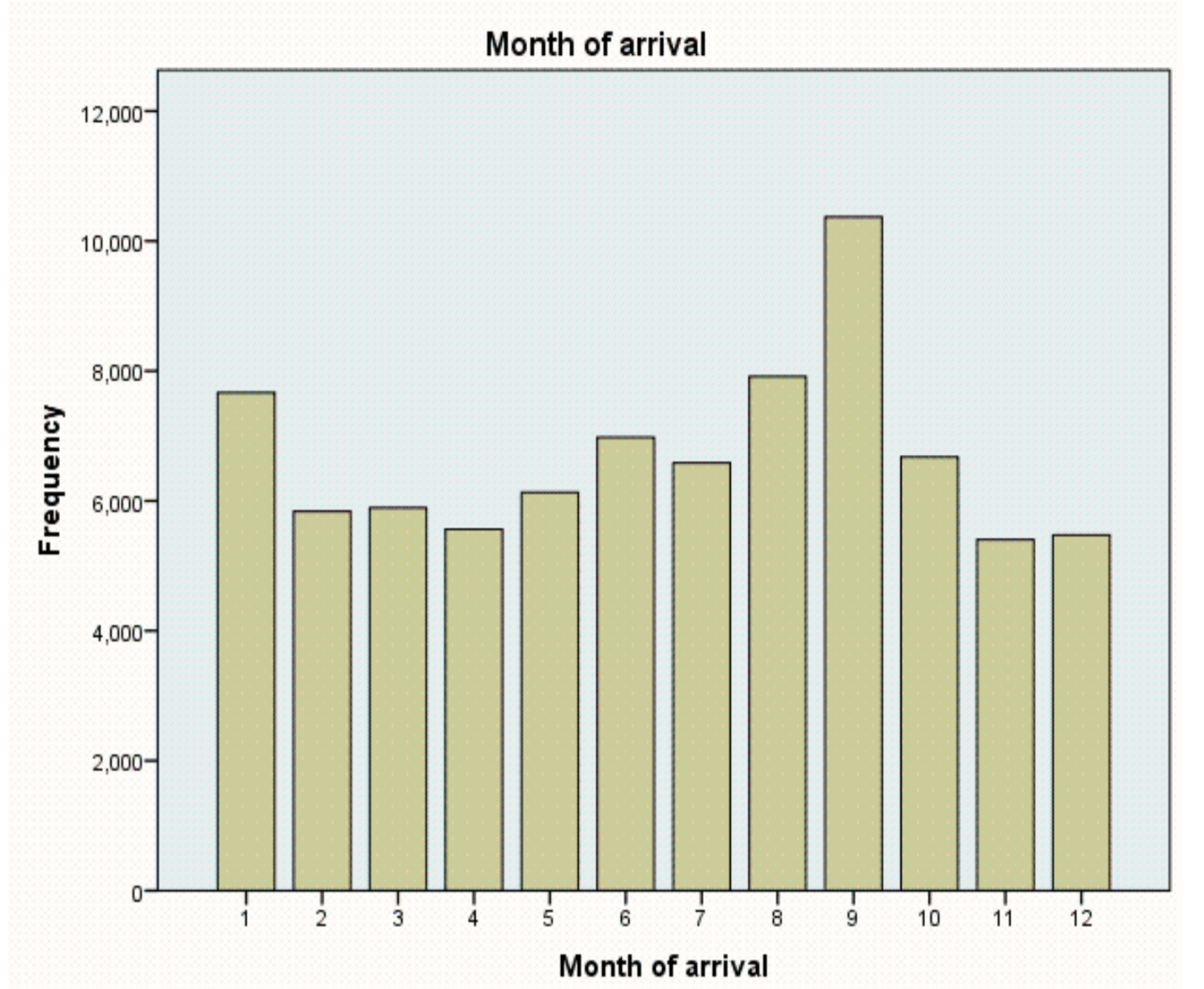


Research example: relationships between migration / time since entering the UK and main language used - Oliver Duke-Williams, Dept of Information Studies and CeLSIUS, UCL

This research formed part of a *beta test* project examining the addition of data from the 2011 Census to the ONS Longitudinal Study. It looked at patterns revealed by results of some new census questions, and considered what they might tell us about migration and language.

Month of arrival in the UK

A question about the **year and month of** (most recent) **entry to the UK** (for persons born outside the UK) is of interest in looking at when new migrants arrive.



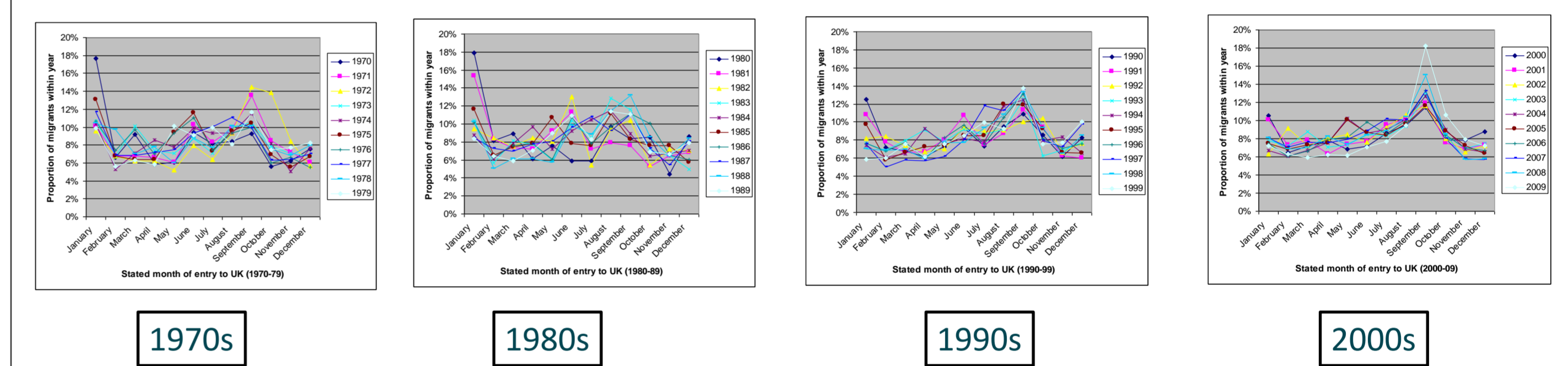
When looking at this for all persons who answered the question (that is, persons born outside the UK) we can see a limited amount of variation by month, with two apparent focuses: arrival in September, and arrival in January.

A September peak may perhaps be indicative of arrivals of students. A January peak is less obvious, but might speculatively be related to recall error: that someone cannot remember the exact month of entry so simply writes a year and month '01'.

This can be looked at more closely by considering both month and year of entry. The four graphs show month of entry by year, for the entry to the UK in the 1970s, 1980s, 1990s and 2000s.

The September entry peak is most apparent in the 1990s and 2000s, and may be associated with an increase in non-UK domiciled students. It is unsurprising that the highest levels are associated with the most recent years: these people are more likely to be current students; after completion of their studies some will leave the UK (and not be recorded in subsequent censuses) whereas others may remain in the UK. The January peak is more pronounced for older entry years, and this might support the partial recall hypothesis.

Year of arrival in the UK – comparing different sources



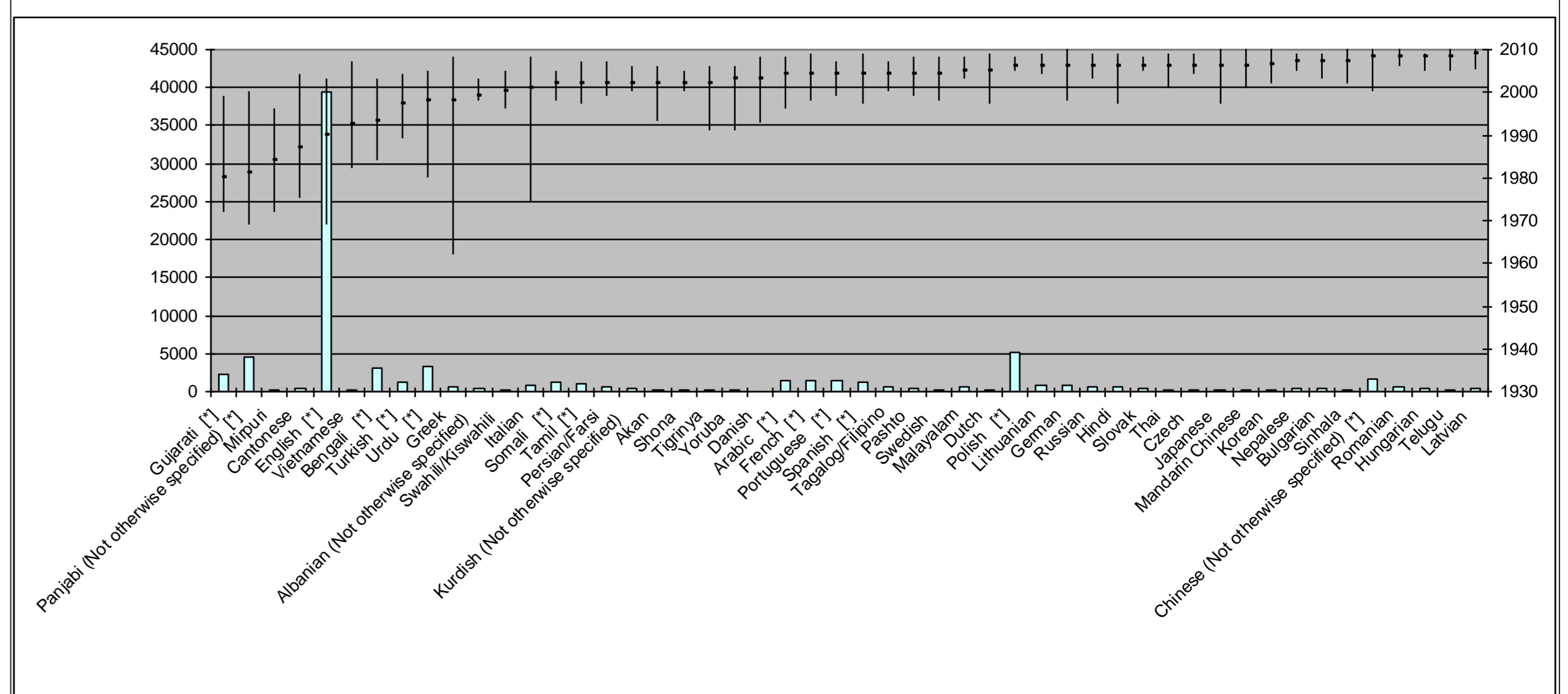
A strength of the LS is that it can be compared for individuals with other sources of data about entry to the UK, including a similar question about *first entry to the UK*, asked in the 1971 Census, and *entry into the NHS administrative records*. It might be assumed that recall error occurs for entry longer ago, but research with the LS suggests that this is not so.

NHS entry data are available from 1971 onwards; we can therefore compare the year of entry (*i.e.* ignoring month) for persons who indicated in the 2011 Census question that they had entered the UK in 1971 or later with the year they first appear in NHS records. There is a correlation of 0.930 (significant at 0.01 level) between these two observations.

What about earlier entry to the UK? Is the same still true? For those people who indicated in the 2011 Census that they entered the UK prior to 1971, and for whole there is also a 1971 Census record in the LS, we can look at their response to the 1971 question about first entry to the UK. This also shows a highly significant strong correlation (0.924) between year of entry, even though the question wording means that the two questions could legitimately be given different responses.

Main language used

Questions about **main language** used and level of **proficiency in English** allow us to look at the rich diversity of languages in the UK, and also to explore possible problems of linguistic isolation. Limited ability in English does not necessarily mean limited proficiency in other languages, nor does it necessarily mean that someone is limited in their day to day activities. Further, the fact that a person has indicated that their main language is not English does not necessarily mean that their English proficiency is poor. By using the data in the LS on length of time in the UK, we can look at languages spoken and time since immigration. The graph below shows the mean and inter-quartile range of year of entry for **persons born outside the UK**, by main language used, and numbers of persons in each group. Data shown for 50 languages with largest number of users; those languages with >1000 users marked [*].



Jo Tomlinson, CeLSIUS, UCL

The permission of the Office for National Statistics to use the Longitudinal Study is gratefully acknowledged. CeLSIUS is supported by the ESRC Census of Population Programme (Award Ref: ES/K000365/1). The authors alone are responsible for the interpretation of the data. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

