

Challenges and solutions – linking additional data to the UK Census Longitudinal Studies

Oliver Duke-Williams

twitter: @oliver_dw email: o.duke-williams@ucl.ac.uk

CLOSER conference:

Preparing for the future II: international approaches to challenges facing the longitudinal population studies

16 January 2020, London

Census data in the UK / Great Britain

- Censuses held since 1801, from 1841 in form that is broadly like current censuses
- Current arrangement is three separate but coordinated censuses (E&W, S, NI)
- Users are generally most familiar with aggregate data

Longitudinal data

- There are three longitudinal studies in the UK
- They have differ sample sizes and cover different time periods
- They also differ in the range and amount of linked data
- All have secure access arrangements

The ONS Longitudinal Study

- Sample selected on the basis of 4 undisclosed birthdays
 - $4/365.25$ sample = 1.1% sample
- Census data from 1971 onwards
- Detailed cohort description in Shelton et al (2018) *International Journal of Epidemiology* 48(2)

<https://doi.org/10.1093/ije/dyy243>

What is in the LSes?

- Similarly to census microdata, all variables apart from individual identifiers
 - More detail than the safeguarded and open microdata files
 - Some (detailed) variables have additionally restricted access
- Imputed fields are included (and identified)
- Imputed records are not included

What is in the LSes?

- As well as LS sample members, the studies also include equivalent records for other persons in the household, referred to as non-members
- Non-members are not linked over time
 - It is sometimes possible to make reasonable assumptions about whether or not a non-member observed at two times is in fact the same person
 - For example, consistent date of birth and relationship to others in the household

Census Data

From each census

- LS sample members
- Other household members
- Age, sex, marital status, country of birth
- Family and household types, communal establishments
- Housing: tenure, rooms and amenities
- Qualifications, economic activity, occupation, industry and social class
- Travel to work and one-year migration
- Geographical information

More recent censuses

- Ethnicity (1991-2011)
- National identity (2011)
- Year of arrival (2011)
- Limiting long-term illness (1991-2011) & self-rated health (2001, 2011)
- Care-giving (2001, 2011)
- Religion (2001, 2011)
- Short-term migration (2011)
- Main language (2011)

Comparison with birth cohorts

- People unfamiliar with the LSes are often more familiar with the idea of birth cohort studies
 - Birth cohorts draw a sample of persons born in a particular year
 - Census longitudinal studies draw a sample across all persons regardless of age
 - LS typically has up to 8k persons per single year of age
 - Total sample size is large; allows analysis of detailed groups

Using the ONS LS

- Two access routes
 - In person at a secure setting
 - Submission of Stata etc scripts to be run remotely
- No data will be transferred out of the secure setting until it has had disclosure clearance

Data dictionaries

CeLSIUS

UCL Home · CeLSIUS · CeLSIUS data dictionary · Details

CeLSIUS

- Home
- About CeLSIUS
- About the ONS Longitudinal Study
- Using the ONS Longitudinal Study
- Projects using the ONS LS
- Resources
- Contact us

Variable details

[Back to Data Dictionary home | previous page]

LS variable metadata

TEND11	
Table	ME11 2011 Census
Short description	Dwelling tenure. 2011.
Source	2011 Census. Questions H12/H13.
Code notes	0-9, -6, -9

Variable coding

Value	Label
0	Owens outright
1	Owens with a mortgage or loan
2	Shared ownership
3	Social rented, Rented from council
4	Social rented, Rented from Registered Social Landlord or housing association
5	Private rented, Private landlord or letting agency
6	Private rented, Employer of a household member
7	Private rented, Relative or friend of household member

Tweets by @celsiusne0s

CeLSIUS @celsiusnews

Looking forward to our conference tomorrow - we have a full programme of student research using #Census data research from the cross sectional and longitudinal anonymised samples eventbrite.co.uk/e/student.

CeLSIUS data dictionary

ucl.ac.uk/celsius

Variable	LS	Variable Description	Table	Table Description	Similar	More	Save
NS_DEP_TENH0	NILS	Census 2001 - National Statistics Deprivation Indicator - Tenure A household is defined as deprived in the housing dimension if it is social housing	C012	2001 Census - Household Information	Yes	↔	+
TENUREH0	NILS	Census 2001 - Tenure of Household	C012	2001 Census - Household Information	Yes	↓	+

Variable Name TENUREH0

Table name C012

Full Description Census 2001 - Tenure of Household

Similar variables in other LSs (Guidance)

- 6 **ONS LS - TENH0** (table: ME01) - Household tenure. 2001.
Comments: ONS category 9 - 'Lives rent free' not present in NILS. ONS 'Other' category appears more general than NILS' 'Private Rented: Other'
- 6 **SLS - TENH0** (table: C22) - Household tenure. 2001.
Comments: SLS category 9 - 'Lives rent free' not present in NILS.

Similar variables across time within NILS

- TENUREH8** (table: C812) - Census 1981 - Tenure
- TENUREH9** (table: C912) - Census 1991 - Tenure
- TENH1** (table: C112) - Census 2011 - Tenure of Household (NI)
- TENUREH8** (table: C812) - Census 1981 - Tenure
- TENUREH9** (table: C912) - Census 1991 - Tenure

CALLS-Hub data dictionary

calls.ac.uk

Linking data to the LSes

- Existing linkages
- Individual linkages
- Contextual linkages



Comparison of non-census data available



<u>England/Wales (LS)</u>	<u>Scotland (SLS)</u>	<u>Northern Ireland (NILS)</u>
Civil registration system	Civil registration system	Civil registration system
Births of sample members	Births of sample members	Births of sample members
Births to sample mothers	Births to sample mothers	Births to sample mothers
	Births to sample fathers	Births to sample fathers
Stillbirths/infant deaths	Stillbirths/infant deaths	Infant deaths
Deaths of sample members	Deaths of sample members	Deaths of sample members
Widow(er)hoods	Widow(er)hoods	
	Marriages	
NHS Patient Register	NHS Patient Register	Health card registration system
Immigration into Eng/Wales	Immigration into Scotland	Immigration into N. Ireland
Emigration from Engl/Wales	Emigration from Scotland	Emigration from N. Ireland
	NHS postcode histories	Internal migration
Cancer registrations	Education data from ScotXed	Land and Property Services
Cancer data	Individual-level data from Schools Census, attendance, absences/exclusions, SQA attainment, qualifications	Type of accommodation, value in 2005, urban/rural etc.
Other data		
Other data can be linked at low level geographies e.g. climate, deprivation quintiles	Special linkage, subject to approval Hospital attendances Maternity data Cancer data Prescribing data Also: Weather and pollution data	Special linkage, subject to approval Health data, inc breast screening, dental treatments, prescription of antibiotics

Individual linkages

- Have to have agreement from data owners
- Have to be done by trusted third party
- Cannot reveal the sample selection birth dates
- Problems
 - Institutional intransigence
 - SLS example: some linkage is done without names: relies on fairly unique key variables, but this assumes linked data are valid for census date

Contextual linkages

- Can link to spatial identifiers (or to any other categorised variable)
 - Users can import data fairly easily
- Problems
 - Geography changes over time
 - Cannot risk disclosure of sub-local authority locations
 - This is problematic for very small areas

An example of a difficult contextual linkage

- User wanted to attach low level data to grid-reference locations in 1971 data
- Giving a list of all locations would reveal very small area locations of 1971 sample members
 - User could not be given a list of all locations
 - We could not be given a list of all locations

An example of a difficult contextual linkage

- Solution
 - Outside the secure environment
 - User gave us target low-level geography
 - We created all possible grid references (some are at 100m level, some are at 1km level)
 - We overlaid these on the user data, generating a point data set with attached properties
 - Transferred the new point data set back into the secure environment
 - Inside the secure environment
 - ONS matched actual records to our superset at grid-ref level
 - Analysis performed on the matching records, attaching user geography
 - Results could be given to user