

Introduction to the Longitudinal Study and census microdata

Dr Oliver Duke-Williams

@oliver_dw

Student Research Using the ONS LS and Census Microdata
(aka SARs/Samples of Anonymised Records)

Census data in the UK / Great Britain

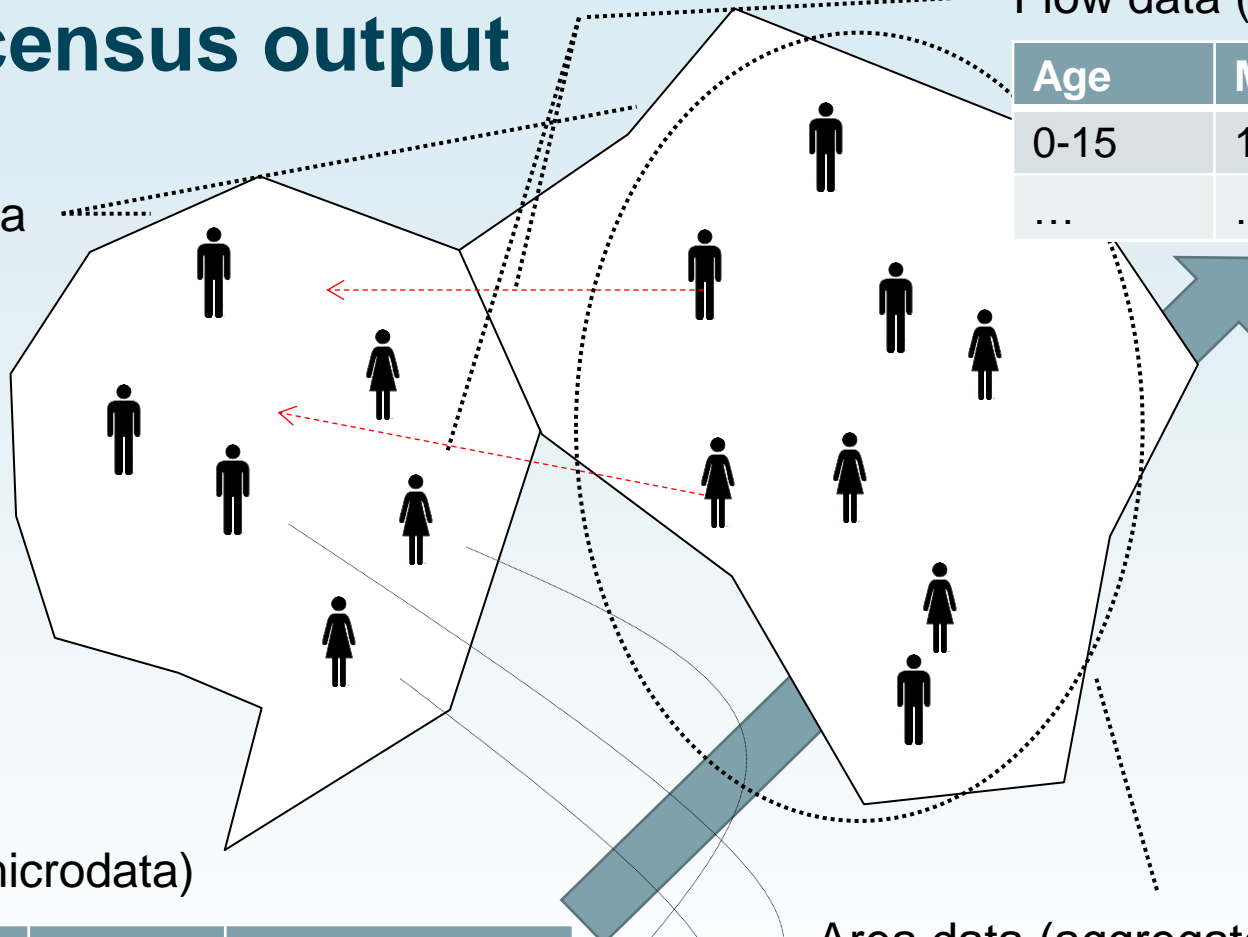
- Censuses held since 1801
- Current arrangement is three separate but coordinated censuses (E&W, S, NI)
- Users are generally most familiar with aggregate data

Types of census output

Boundary data

Flow data (O/D data)

Age	M	F
0-15	1	2
...



Individual data (microdata)

ID	Age	Gender	Occupation
0001	23	F	Teacher
0002	17	M	Student
0003	45	F	Doctor
...

Area data (aggregate data)

Age	M	F
0-4	7	6
5-9	8	9
...

Census microdata

- Often referred to as 'individual data'
- Two types
 - Cross-sectional microdata
 - Longitudinal microdata

Census microdata in the UK

- First produced from the 1991 Census; labelled as ‘Sample of Anonymised Records’ or SARs
 - Separate files for Great Britain and for Northern Ireland
 - Separate individual and household samples
- Changes in sample sizes and license arrangements over time

‘Samples of Anonymised Records’

		Open	Safeguarded	Secure
1991	Individuals		2% (GB,NI)	
	Households		1% (GB,NI)	
2001	Individuals		3% (UK)	
	Households			1% (EW)
	‘Small Area Microdata’		5% (UK)	

‘Census Microdata’

		Open	Safeguarded	Secure
1961	Individuals	1% (GB)	5% (GB)	9% (GB)
	Households		0.95% (GB)	
1971	Individuals	1% (GB)	5% (GB)	9% (GB)
	Households		0.95% (GB)	
1981	Individuals	1% (GB)	5% (GB)	9% (GB)
	Households		0.95% (GB)	
2011	Individual		5% (EW,S,NI)	10% (EW,S,NI)
	Households			10% (EW,S,NI)
	Individual (‘Teaching’)	1% (EW,S,NI)		

What is in the microdata files?

- Broadly, all data, except personal identifiers

But!

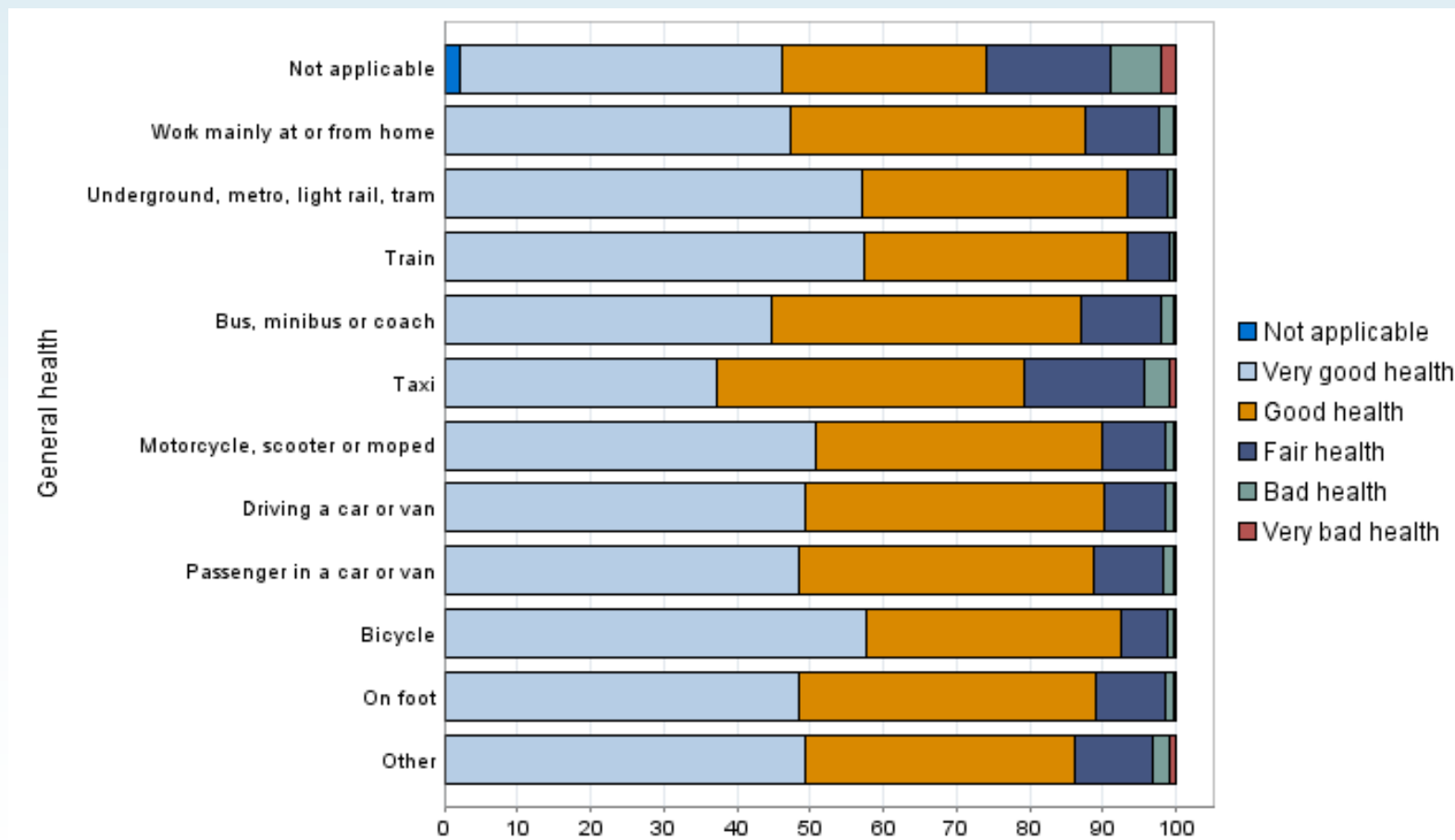
- Trade-offs in level of detail, completeness of data etc with access arrangements and sample size

What can we do with microdata

- Regression modelling etc on individual characteristics
- Multi-level modelling on individual / household / local area characteristics
- Cross tabulation of fields in combinations not found in the standard outputs, e.g....

2011, England and Wales

Travel to work by general health



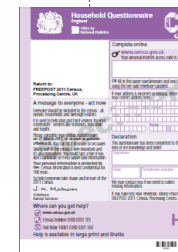
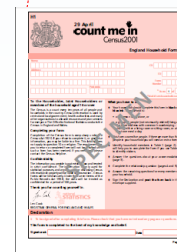
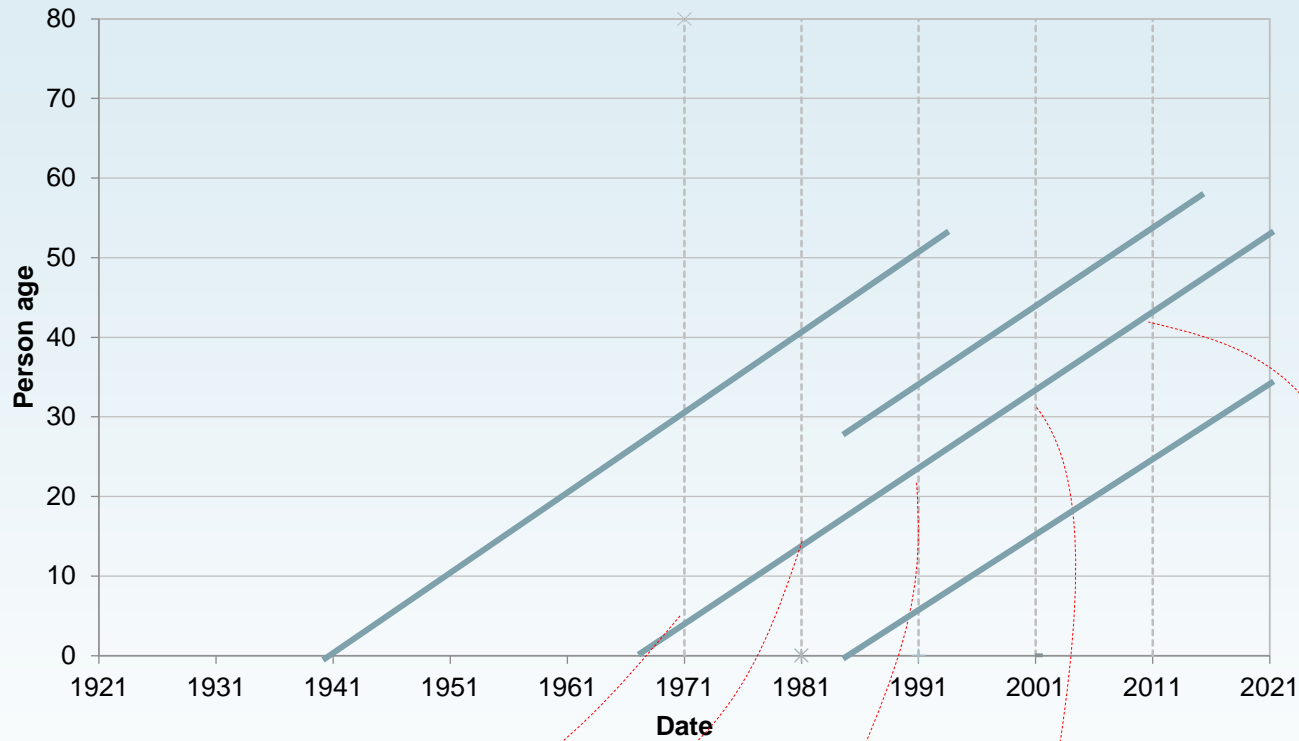
Source: Office for National Statistics, 2011 Census Microdata Individual Safeguarded Sample (Regional) [computer file]. Colchester, Essex: UK Data Archive [distributor], December 2014. SN: 7605: [10.5255/UKDA-SN-7605-1](https://beta.ukdataservice.ac.uk/datacatalog/7605/10.5255/UKDA-SN-7605-1)

Longitudinal data

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

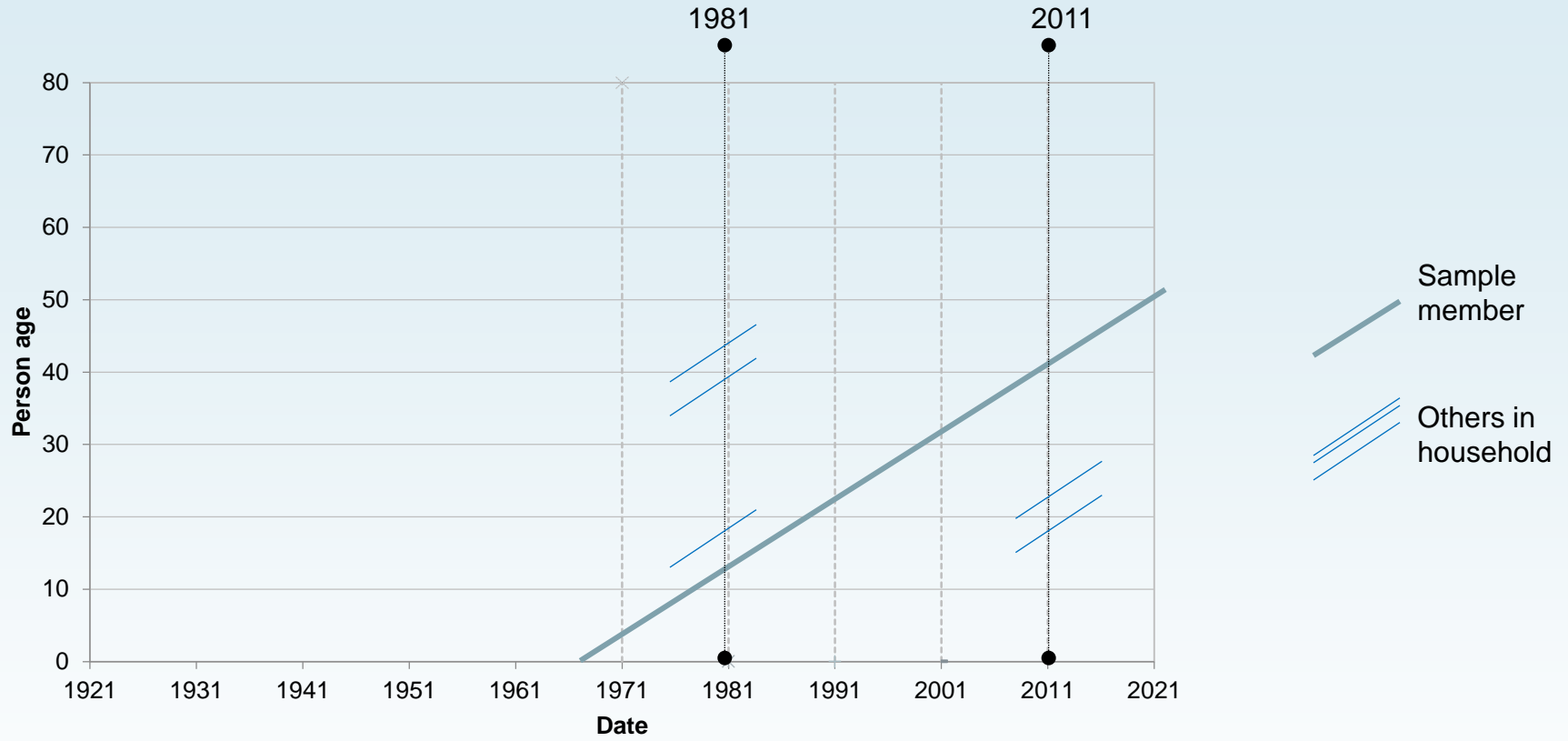
UK Census Longitudinal Studies

	Sample size	1971	1981	1991	2001	2011
ONS Longitudinal Study	4/365.25	✓	✓	✓	✓	✓
Scottish Longitudinal Study	20/365.25		✓	✓	✓	✓
Northern Ireland Longitudinal Study	104/365.25			✓	✓	✓



What is in the LS?

- 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
- Sample members and others in the household



Census Data

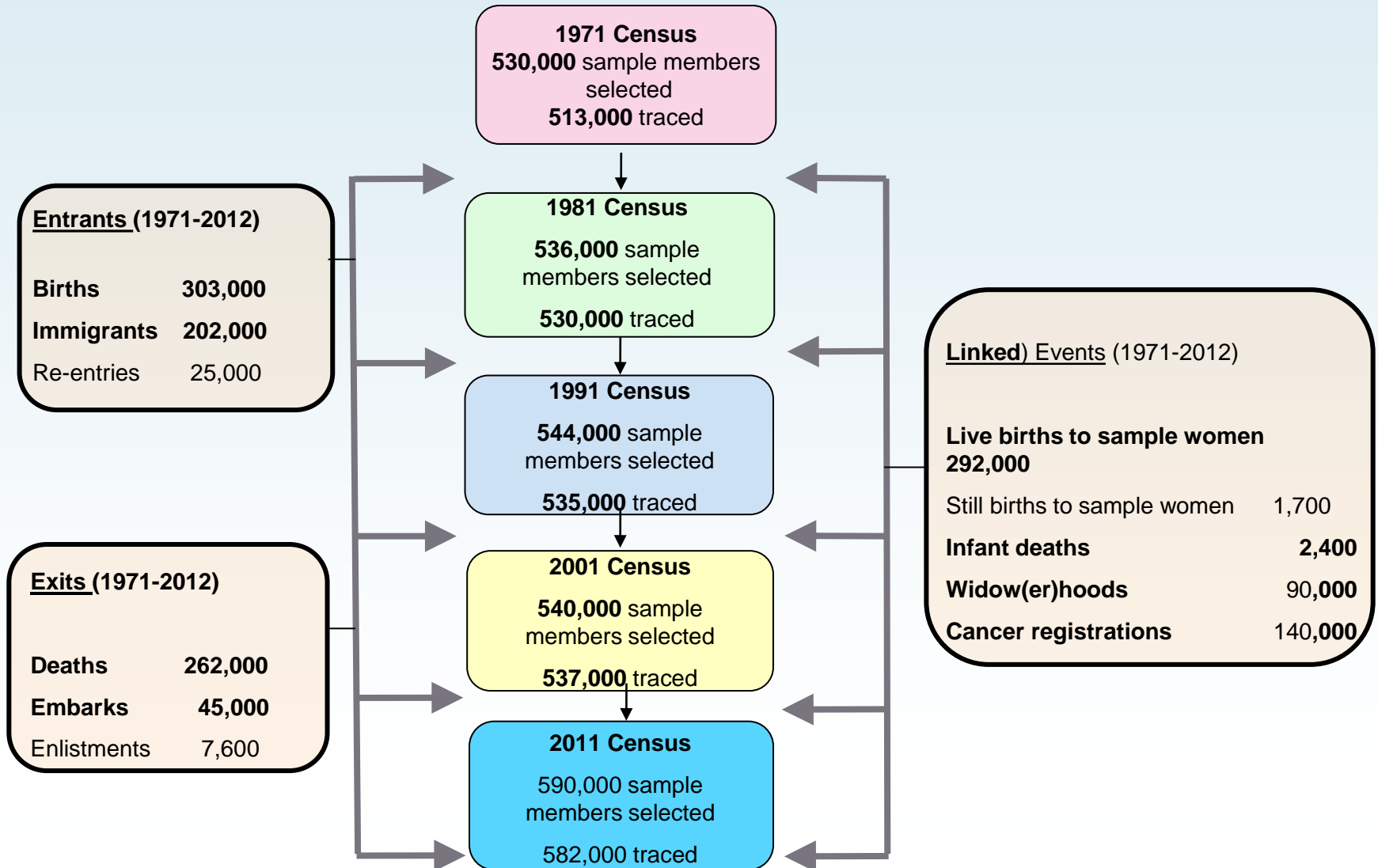
From each census

- 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

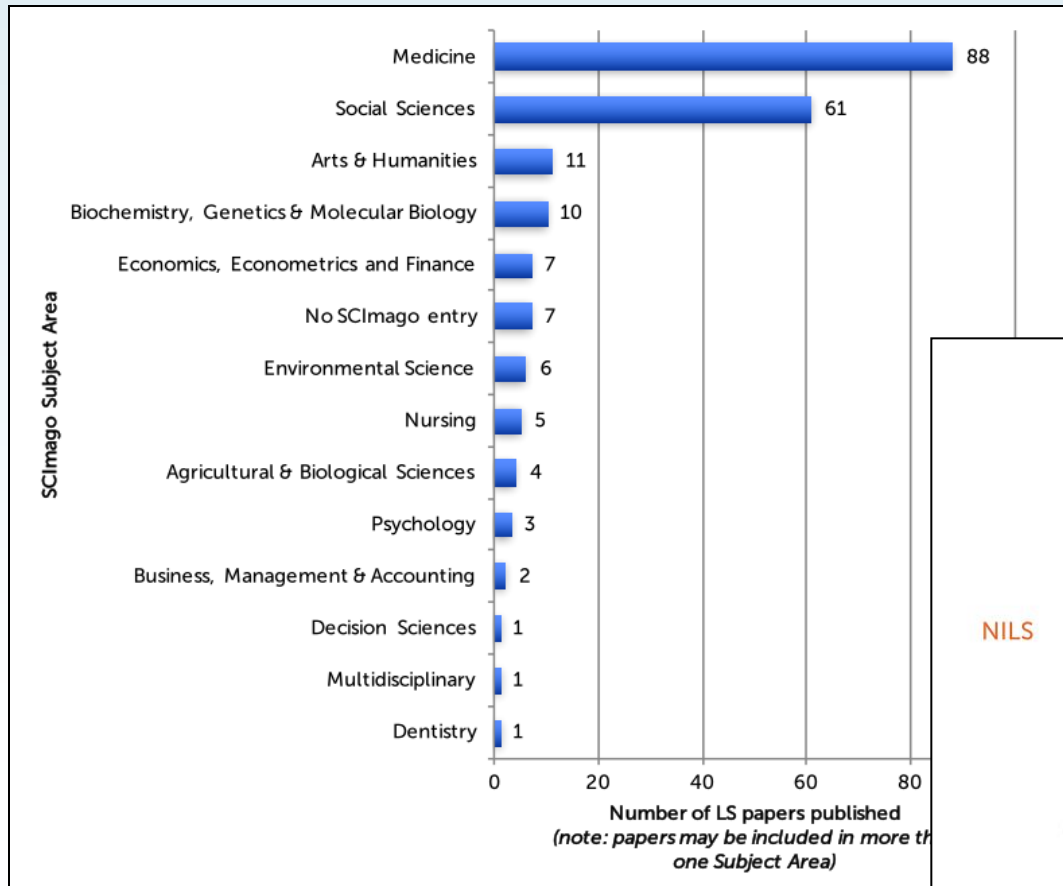
More recent censuses

- 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)

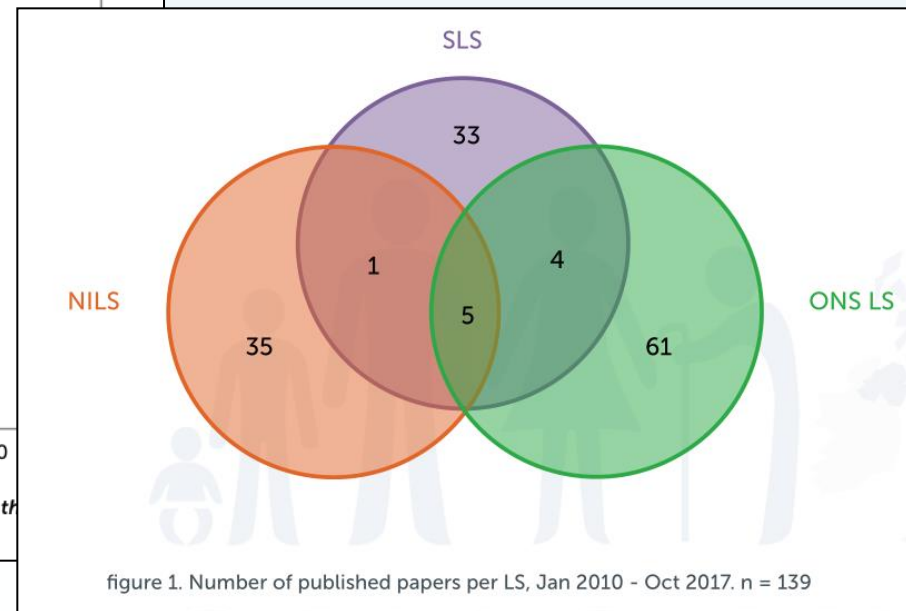
LS structure: England & Wales



Who uses the LS?



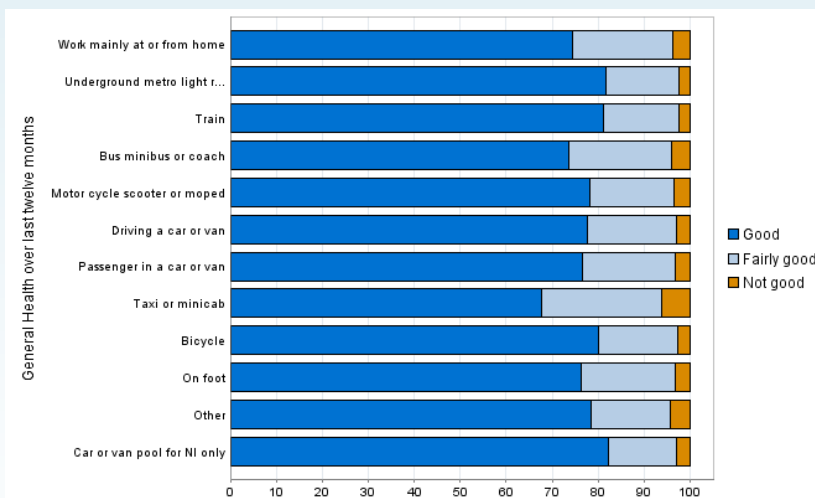
Source: Cox, F (2017) CALLS Hub Citation Analysis
<https://calls.ac.uk/research-blog/>



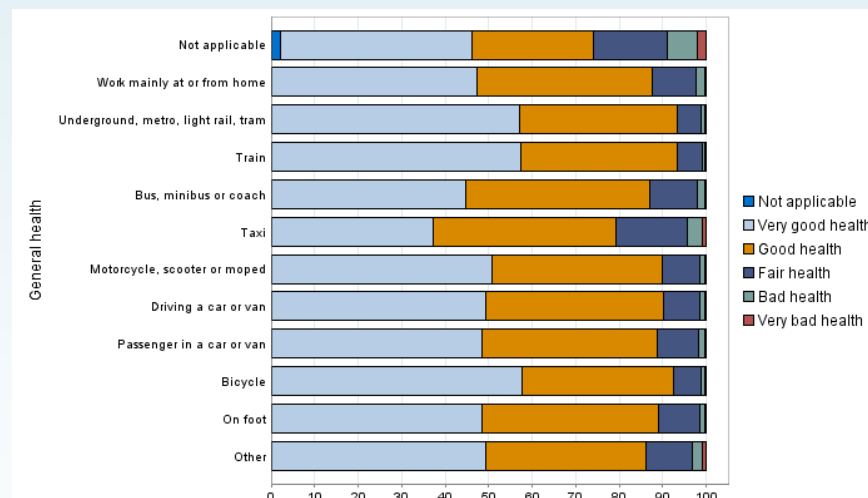
Travel to work by general health 2001 v 2011

Comparing cross-sectional microdata

2001



2011



Sources:

Office for National Statistics. Census Division and University of Manchester. Cathie Marsh Centre for Census and Survey Research, Census 2001: Individual Licenced Sample of Anonymised Records (I-SAR) [computer file]. Colchester, Essex: UK Data Archive [distributor], February 2013. SN: 7205: [10.5255/UKDA-SN-7205-1](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-7205-1)

Office for National Statistics, 2011 Census Microdata Individual Safeguarded Sample (Regional) [computer file]. Colchester, Essex: UK Data Archive [distributor], December 2014. SN: 7605: [10.5255/UKDA-SN-7605-1](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-7605-1)

- Previous slide shows travel to work in 2001 and in 2011
- We can compare cross-sections easily enough
- But: we don't know whether those that used (mode x) in 2001 were the same people that used (mode x) in 2011, unless we use longitudinal data...

Travel to work 2001 v 2011: longitudinal transitions

For each 2001 mode

- most common 2011 outcome
- second most common 2011 outcome

		Mode in 2011								Total	
		Train	Tube	Bus	Driving a car or van	Passenger in a car or van	Motorcycle	Bicycle	On foot		Work mainly at or from home
Mode in 2001	Train	41%	6%	4%	32%	2%	1%	2%	5%	7%	6324
	Tube/Metro/Light rail/tram (E&W)	17%	33%	8%	24%	2%	1%	3%	5%	7%	3849
	Bus/Minibus/Coach	5%	3%	27%	39%	8%	0%	2%	12%	3%	10638
	Drive car/van	2%	1%	2%	82%	2%	1%	1%	4%	5%	93087
	Passenger car/van	3%	1%	8%	53%	20%	1%	2%	10%	3%	10522
	Motor cycle/Scooter/Moped	4%	2%	3%	56%	4%	19%	6%	4%	3%	1840
	Bicycle	3%	1%	4%	43%	4%	2%	30%	10%	3%	4415
	On foot	3%	2%	7%	40%	6%	1%	3%	34%	4%	14621
	Works at/from home	3%	1%	2%	54%	3%	0%	1%	7%	28%	14035
	Total	7381	3428	7323	105507	6879	1248	4053	13075	10437	159331

- 30% retention rate for cyclists
- 34% retention rate for walkers

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

Comparison with birth cohorts

- Sample sizes
 - LS total samples are much bigger
 - LS individual year-of-age samples are smaller in England and Wales, but less subject to attrition
 - Starting cohort sizes 17-19K
 - ONS LS, 2011 per single year of age < 65 c. 5-8K
- Content
 - Cohort studies have much broader content
 - Sample size allows LS to have more detailed geography etc

Questions?

Acknowledgements

•The permission of the Office for National Statistics (ONS) to use the Longitudinal Study is gratefully acknowledged, as is the help provided by staff of the Centre for Longitudinal Study Information & User Support (CeLSIUS). 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.