

Millennium Cohort Study

Guide to geographic identifiers

David Church

CLS Data Note / User guide to the data (Fifth Edition)

June 2017



Centre for Longitudinal Studies

Following lives from birth through the adult years
www.cls.ioe.ac.uk

CLS is an ESRC Resource Centre based at the Institute of Education, London



First published in 2017 by the
Centre for Longitudinal Studies
Institute of Education, University of London
20 Bedford Way
London WC1H 0AL
www.cls.ioe.ac.uk
© Centre for Longitudinal Studies

The Centre for Longitudinal Studies (CLS) is an ESRC Resource Centre based at the Institution of Education. It provides support and facilities for those using the three internationally-renowned birth cohort studies: the National Child Development Study (1958), the 1970 British Cohort Study and the Millennium Cohort Study (2000). CLS conducts research using the birth cohort study data, with a special interest in family life and parenting, family economics, youth life course transitions and basic skills. The views expressed in this work are those of the author(s) (amend as necessary) and do not necessarily reflect the views of the Economic and Social Research Council. All errors and omissions remain those of the author(s).

This document is available in alternative formats.

Please contact the Centre for Longitudinal Studies.

tel: +44 (0)20 7612 6875

email: clsfeedback@ioe.ac.uk

Table of Contents

| | |
|--|-----------|
| Introduction..... | 3 |
| Data and methodology | 3 |
| Extent and nature of the data..... | 4 |
| Availability of Datasets..... | 5 |
| Further Information..... | 5 |
| References | 6 |
| APPENDIX A: Data Sources for Geographical Identifiers available under Secure Access | 7 |
| APPENDIX B: Variables available under Secure Access | 10 |

Introduction

There has been an increasing awareness of the value of geographically linked data in social scientific research, especially since the 'GIS revolution' of the early 1990s (Longley and Batty, 1996). Spatial data can be approached from a number of directions. For example, "longitudinal studies are particularly valuable to geographers because they chart change, collect information across various domains and are spatially referenced" (Ekinsmyth, 1996: 364). On the other hand, economists, particularly those of a more heterodox bent, are beginning to appreciate the value of spatially referenced data, especially in research into the economics of education (e.g. Gibbons et al, 2013 who use the National Pupil Database to estimate the effects of neighbourhood composition on teenagers' behavioural and educational outcomes in England). Epidemiology and its associated disciplines are perhaps most consistently associated with investigating the spatial effects of the type of data collected across the different longitudinal cohort studies. For example, Christakis and Fowler (2007) used data from the Framingham Heart Study in the US to examine the spread of obesity in a large social network over 32 years while Tunstall et al (2010) used data from the Millennium Cohort Study to analyse the health outcomes of pregnant women who moved house. Two particularly fruitful fields are, firstly, the investigation of so-called 'neighbourhood effects' across a number of socio-economic domains (e.g. Lupton and Kneale (2012) used data from the 1970 British Cohort Study to investigate neighbourhood influences on teenage parenthood) and, secondly, network-based analyses of particular issues such as obesogenic environments (e.g Burgoine et al, 2014), accessibility to health-promoting community resources (e.g. Wolch et al, 2011) and the impact of built environment (morphological) characteristics on health and well-being (e.g. Sarkar et al, 2014).

However, balancing the obvious advantages of incorporating the spatial dimension within longitudinal social scientific research, there are a number of important limitations to be borne in mind when dealing with this type of data. The principal consideration is protecting the identity of cohort members, particularly in the current era of 'open data' and increasing linkage of previously disparate administrative datasets. It is recognised that there is a cultural dimension to the issue of data confidentiality, with, for example, Scandinavian countries taking a more laissez-faire approach, adopting the perspective that data gathered using public funds should be available for public consumption. More socially conservative states like the US and UK, on the other hand, have tended to take a much more protectionist approach to personal data (Exeter et al, 2014). At present, the UK Data Archive takes the approach that access to geo-referenced data below Government Office Region (GOR) level should be subject to increasing access restrictions the more likely the data is to reveal the identity of cohort members. Other limitations include non-uniformity of geo-identifiers used across different sweeps of the various cohort studies and varying levels of accuracy in terms of the geo-identifiers collected (a particular problem of early sweeps before the standardisation of unit postcodes).

Data and methodology

In order to enable the process of spatial analysis of longitudinal cohort study data, unit postcodes are gathered from the addresses collected during interview, which are then validated by the CLS Cohort Maintenance Team using a range of specialist software products

from AFD¹. This postcode data is then used to generate point data, usually within a GIS. There are a number of licensed and open source GIS packages available (e.g. ArcGIS², MapInfo³ and QGIS⁴). The primary data source for spatialising longitudinal cohort study data within this software is the ONS Postcode Directory, available from the UK Data Service Census Support website⁵. This dataset has been released quarterly since 2004 (every February, May, August and November) and contains Ordnance Survey eastings and northings for each unit postcode centroid. These eastings and northings are spatialised in GIS in the form of 'x', 'y' points, usually to an accuracy of 1 metre of the mean postcode centroid⁶. Areal (polygon) representations of unit postcodes are produced in the OS 'Code-Point with Polygons' product, available from Edina Digimap⁷. The current (November 2014) release of the Postcode Directory contains some 42 different geographies dating back to 1991, encompassing Census, health, administrative, environmental and educational domains. Once again, boundaries for the majority of these geographies are available from Edina Digimap.

The point-based longitudinal cohort study data is associated with these geographies by means of a 'field join' (usually based on unit postcode) or by a 'spatial join' (i.e. based on location) within a GIS. One drawback of the use of mean unit postcode centroids in geo-referenced data is their wide variation in size (typically unit postcodes in rural areas will cover a much larger extent than their urban counterparts). This means that greater accuracy will be achieved in geo-referencing the address at interview of cohort members who live in more urbanised areas. Recent developments in geo-coding with products such as OS 'AddressBase'⁸ mean that it is possible to create point data based on the grid references of individual properties/locations (thus with an accuracy of metres rather than tens/hundreds of metres achieved using unit postcode centroids). This is particularly important for the types of network and accessibility based analyses alluded to above, as well as specific types of environmental analysis, such as the impact on residents of the electromagnetic fields of high voltage overhead power lines (e.g. Swanson et al, 2014).

Extent and nature of the data

England and Wales use the same naming conventions across different geographies. Post-devolution, Scotland has adopted slightly different naming conventions. For example, in both 2001 and 2011 Census geography, what are known as 'Lower Super Output Areas' and 'Middle Super Output Areas' in England and Wales are called 'Data Zones' and 'Intermediate Geographies' respectively in Scotland and the mean populations used to create these areal units also varies between E&W and Scotland⁹. The projected coordinate system used to display geo-referenced data across Great Britain (i.e. England, Wales and Scotland) is the

¹ www.afd.co.uk/

² Licensed software, available from <http://www.esriuk.com/>

³ Licensed software, available from <http://www.mapinfo.com/>

⁴ Open-source software, available from <http://www2.qgis.org/en/site/>

⁵ <http://census.edina.ac.uk//pcluts.html>

⁶ There are, however, a range of 'grid reference positional quality indicators', ranging from 1 ('within the building of the matched address closest to the postcode mean' to 9 ('no grid reference available'). Some 85% of all postcodes have a positional quality indicator of 1.

⁷ <http://digimap.edina.ac.uk/digimap/home>

⁸ <http://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-products.html>

⁹ For example, the mean population of a 2011 LSOA is 1,500 whereas the mean population of a 2011 DZ is 808.

British National Grid¹⁰. The range of Ordnance Survey products (e.g. MasterMap, AddressBase, OpenData) is available for Great Britain (i.e. excluding Northern Ireland).

Northern Ireland uses its own equivalent to the ONS Postcode Directory, called the Central Postcode Directory¹¹, and the process of spatialising geo-referenced data works in exactly the same way as with Great Britain data. Northern Ireland uses a different projected coordinate system from the rest of the UK, the Irish National Grid¹².

Availability of Datasets

Because of the potentially disclosive nature of these datasets, geographic identifiers below Government Office Region (GOR) are being released under Secure Access¹³; these are listed at Appendix A.

Appendix B lists the contents of the Secure Access Datasets.

Appendix C lists the geographies available on the Datasets available under the standard End User Licence.

As you will note in Appendix C, there are two versions of MCS5 and MCS6 geographically linked data available, one set based on 2001 Census boundaries and the other set based on 2011 Census boundaries. Due to the potential for identification of a small number of individual cohort members (i.e. those located in ‘slivers’ between those areal units whose boundaries changed between 2001 and 2011), users can choose either, but not both, of these datasets.

Further Information

If you have any queries, please contact us at clsfeedback@ioe.ac.uk or view our website at <http://www.cls.ioe.ac.uk>.

¹⁰ WKID 27700, Authority EPSG

¹¹ Available, upon application and receipt of password, from <http://www.nisra.gov.uk/geography/postcode.htm>

¹² WKID 29902, Authority EPSG

¹³ <http://ukdataservice.ac.uk/get-data/how-to-access/accesssecurelab>

References

- Burgoine, T., N. G. Forouhi, et al. (2014). "Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study." Bmj **348**.
- Christakis, N. A. and J. H. Fowler (2007). "The spread of obesity in a large social network over 32 years." New England journal of medicine **357**(4): 370-379.
- Ekinsmyth, C. (1996). "Large-scale longitudinal studies: their utility for geographic enquiry." Area: 358-372.
- Exeter, D. J., S. Rodgers, et al. (2014). "'Whose data is it anyway?' The implications of putting small area-level health and social data online." Health Policy **114**(1): 88-96.
- Gibbons, S., O. Silva, et al. (2013). "Everybody needs good neighbours? Evidence from students' outcomes in England." The Economic Journal **123**(571): 831-874.
- Longley, P. A. and M. Batty (1996). Spatial analysis: modelling in a GIS environment, John Wiley & Sons.
- Lupton, R. and D. Kneale (2012). Theorising and measuring place in neighbourhood effects research: The example of teenage parenthood in England. Neighbourhood Effects Research: New Perspectives. M. van Ham, D. Manley, N. Bailey, L. Simpson and D. Maclennan. Berlin, Springer.
- Sarkar, C., C. Webster, et al. (2014). Healthy Cities: Public Health Through Urban Planning, Edward Elgar Publishing.
- Swanson, J., T. Vincent, et al. (2014). "Relative accuracy of grid references derived from postcode and address in UK epidemiological studies of overhead power lines." Journal of Radiological Protection **34**(4): N81.
- Tunstall, H., K. Pickett, et al. (2010). "Residential mobility in the UK during pregnancy and infancy: Are pregnant women, new mothers and infants 'unhealthy migrants'?" Social Science & Medicine **71**(4): 786-798.
- Wolch, J., M. Jerrett, et al. (2011). "Childhood obesity and proximity to urban parks and recreational resources: A longitudinal cohort study." Health & place **17**(1): 207-214.

APPENDIX A: Data Sources for Geographical Identifiers available under Secure Access

| Dataset | Description | Boundary Data Source (UK Data Service unless stated) |
|-------------------------------|---|--|
| Interview Wards | 1998 Ward Boundaries (Sweeps 1 – 5) 2001 Census Area Statistic Ward ¹⁴ (Sweeps 1 – 5) | English Electoral Wards, 1998 Welsh Electoral Wards, 1998 Scottish Electoral Wards, 1998 Northern Ireland Electoral Wards, 2001 ¹⁵ English Census Area Statistic Wards, 2001 Welsh Census Area Statistic Wards, 2001 Scottish Census Area Statistic Wards, 2001 Northern Ireland Electoral Wards, 2001 ¹⁶ |
| Interview Output Area (OA) | 2001 Output Area (Sweeps 1 – 5) 2011 Output Area (Sweep 5 only) | English Output Areas, 2001 Welsh Output Areas, 2001 Scottish Output Areas, 2001 Northern Ireland Output Areas, 2001 English Output Areas, 2011 Welsh Output Areas, 2011 Scottish Output Areas, 2011 |

¹⁴ see <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/electoral-wards-divisions/statistical-wards--cas-wards-and-st-wards/index.html> for definition and http://census.edina.ac.uk/pcluts_download.html?data=pcluts_2014nov for latest ONS Postcode Directory code/name lookups

¹⁵ Northern Ireland Electoral Wards remained unchanged between 1998 and 2001

¹⁶ Northern Ireland does not have CAS wards but the electoral wards have alternative codes in Census outputs.

| Dataset | Description | Boundary Data Source (UK Data Service unless stated) |
|---|---|---|
| | | Northern Ireland Small Areas, 2011 |
| Interview Lower Super Output Area (LSOA) | 2001 Lower Super Output Area (Sweeps 1 – 5) 2011 Lower Super Output Area (Sweep 5 only) | English Lower Super Output Areas, 2001 Welsh Lower Super Output Areas, 2001 Scottish Datazones, 2001 Northern Ireland Lower Super Output Areas, 2001 English Lower Super Output Areas, 2011 Welsh Lower Super Output Areas, 2011 Scottish Lower Super Output Areas, 2011 Northern Ireland Super Output Areas, 2011 ¹⁷ |
| Interview Middle Super Output Area (MSOA) | 2001 Middle Super Output Area (Sweeps 1 – 5 only) 2011 Middle Super Output Area (Sweep 5 only) | English Lower Super Output Areas, 2001 Welsh Lower Super Output Areas, 2001 Scottish Intermediate Geographies, 2001 English Lower Super Output Areas, 2011 Welsh Lower Super Output Areas, 2011 Scottish Intermediate Geographies, 2011 |
| Interview Local Authority District | 2001 Local Authority District/Unitary Authority (Sweeps 1 – 5 only) | English Administrative Districts, 2001 English Unitary Authorities, 2001 Welsh Unitary Authorities, 2001 Scottish Council Areas, 2001 |

¹⁷ In both 2001 and 2011 Census Geography, Northern Ireland does not have Middle Super Output Areas

| Dataset | Description | Boundary Data Source (UK Data Service unless stated) |
|--|--|---|
| | <p>2011 Local Authority District/Unitary Authority (Sweep 5 only)</p> <p>2015 Local Authority District/Unitary Authority (Sweep 6 only)</p> | <p>Northern Ireland District Councils, 2001</p> <p>Ordnance Survey Boundary-Line, November 2013¹⁸</p> <p>Ordnance Survey Boundary-Line, November 2015¹⁹</p> |
| Interview Westminster Parliamentary Constituency | <p>2001 Westminster Parliamentary Constituency (Sweep 1)</p> <p>2005 Westminster Parliamentary Constituency (Sweeps 2 – 4)</p> <p>2011 Westminster Parliamentary Constituency (Sweep 5 only)</p> <p>2015 Westminster Parliamentary Constituency (Sweep 6 only)</p> | <p>English Westminster Parliamentary Constituencies, 2001</p> <p>Welsh Westminster Parliamentary Constituencies, 2001</p> <p>Scottish Westminster Parliamentary Constituencies, 2001</p> <p>Northern Ireland Westminster Parliamentary Constituencies, 2001</p> <p>Scottish Westminster Parliamentary Constituencies, 2005²⁰</p> <p>Ordnance Survey Boundary-Line, November 2011²¹</p> <p>Ordnance Survey Boundary-Line, November 2015²²</p> |

¹⁸ Available from <http://digimap.edina.ac.uk>

¹⁹ Available from <http://digimap.edina.ac.uk>

²⁰ 2005 Westminster Parliamentary Constituency boundaries remained the same as in 2001 for England, Wales and Northern Ireland

²¹ Available from <http://digimap.edina.ac.uk>

²² Available from <http://digimap.edina.ac.uk>

APPENDIX B: Variables available under Secure Access

Geographical Identifiers (based on address at interview), MCS Sweep 1

| Variable name | Description |
|---------------|--|
| award98 | S1 Ward Code (1998 Boundaries) |
| acasward | S1 Census Statistic Ward Code 2001 |
| alsoa | S1 Lower Super Output Area Code 2001 |
| amsoa | S1 Middle Super Output Area Code 2001 |
| aoa | S1 Output Area 2001 |
| aladist | S1 LA District/Unitary Authority 2001 |
| awestcon | S1 Westminster Parliamentary Constituencies 2001 |

Geographical Identifiers (based on address at interview), MCS Sweep 2

| Variable name | Description |
|---------------|--|
| bward98 | S2 Ward Code (1998 Boundaries) |
| bcasward | S2 Census Statistic Ward Code 2001 |
| blsoa | S2 Lower Super Output Area Code 2001 |
| bmsoa | S2 Middle Super Output Area Code 2001 |
| boa | S2 Output Area 2001 |
| bladist | S2 LA District/Unitary Authority 2001 |
| bwestcon | S2 Westminster Parliamentary Constituencies 2005 |

Geographical Identifiers (based on address at interview), MCS Sweep 3

| Variable name | Description |
|---------------|--|
| cward98 | S3 Ward Code (1998 Boundaries) |
| ccasward | S3 Census Statistic Ward Code 2001 |
| clsoa | S3 Lower Super Output Area Code 2001 |
| cmsoa | S3 Middle Super Output Area Code 2001 |
| coa | S3 Output Area 2001 |
| cladist | S3 LA District/Unitary Authority 2001 |
| cwestcon | S3 Westminster Parliamentary Constituencies 2005 |

Geographical Identifiers (based on address at interview), MCS Sweep 4

| Variable name | Description |
|----------------------|--|
| dward98 | S4 Ward Code (1998 Boundaries) |
| dcasward | S4 Census Statistic Ward Code 2001 |
| dlsoa | S4 Lower Super Output Area Code 2001 |
| dmsoa | S4 Middle Super Output Area Code 2001 |
| doa | S4 Output Area 2001 |
| dladist | S4 LA District/Unitary Authority 2001 |
| dwestcon | S4 Westminster Parliamentary Constituencies 2005 |

Geographical Identifiers (based on address at interview), MCS Sweep 5 (2001 Census)

| Variable name | Description |
|----------------------|--|
| eward98 | S5 Ward Code (1998 Boundaries) |
| ecasward | S5 Census Statistic Ward Code 2001 |
| elsoa | S5 Lower Super Output Area Code 2001 |
| emsoa | S5 Middle Super Output Area Code 2001 |
| eoas | S5 Output Area 2001 |
| eladist | S5 LA District/Unitary Authority 2001 |
| ewestcon | S5 Westminster Parliamentary Constituencies 2005 |

Geographical Identifiers, (based on address at interview), MCS Sweep 5 (2011 Census)

| Variable name | Description |
|----------------------|--|
| eward98 | S5 Ward Code (1998 Boundaries) |
| ecasward | S5 Census Statistic Ward Code 2001 |
| elsoa11 | S5 Lower Super Output Area Code 2011 |
| emsoa11 | S5 Middle Super Output Area Code 2011 |
| eoas11 | S5 Output Area 2011 |
| eladist13 | S5 LA District/Unitary Authority 2013 |
| ewestcon11 | S5 Westminster Parliamentary Constituencies 2011 |

Geographical Identifiers (based on address at interview), MCS Sweep 6 (2001 Census)

| Variable name | Description |
|----------------------|--|
| fward98 | S5 Ward Code (1998 Boundaries) |
| fcasward | S5 Census Statistic Ward Code 2001 |
| flsoa | S5 Lower Super Output Area Code 2001 |
| fmsoa | S5 Middle Super Output Area Code 2001 |
| foa | S5 Output Area 2001 |
| fladist | S5 LA District/Unitary Authority 2015 |
| fwestcon | S5 Westminster Parliamentary Constituencies 2015 |

Geographical Identifiers, (based on address at interview), MCS Sweep 6 (2011 Census)

| Variable name | Description |
|----------------------|--|
| fward98 | S5 Ward Code (1998 Boundaries) |
| fcasward | S5 Census Statistic Ward Code 2001 |
| flsoa11 | S5 Lower Super Output Area Code 2011 |
| fmsoa11 | S5 Middle Super Output Area Code 2011 |
| foa11 | S5 Output Area 2011 |
| fladist13 | S5 LA District/Unitary Authority 2015 |
| fwestcon | S5 Westminster Parliamentary Constituencies 2015 |

UCL Centre for Longitudinal Studies
Institute of Education
20 Bedford Way
London WC1H 0AL
Tel: 020 7612 6860
Fax: 020 7612 6880
Email: clsfeedback@ioe.ac.uk
Web: www.cls.ioe.ac.uk