Institute of Education

NCDS

Data Note: Pre-2000 geographically linked data

David Church

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

March 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

ISBN XXXX

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

This Data Note explains the pre-2000 geographical units which have been linked to sweeps 3, 4, and 5 of the National Child Development Study (NCDS).

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 AFD[1]. Unit postcode at interview is the variable used to link to a range of higher level areal units. The UK has had full coverage of postcodes since 1974 (Raper et al, 1992). This means that accurate geographically linked data is available for NCDS from sweep 3 onwards.

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 website[2]. This dataset has been released quarterly since 2004 (every February, May, August and November) and stretches back on an annual basis to 1980. The Postcode Directory 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 centroid[3]. Eastings and Northings are Postcodes held by CLS have been cleaned (i.e. reformatted to the standard 6, 7, or 8 character 'pcd', 'pcd2', or 'pcds' format as present in the Postcode Directory). This postcode data can then be linked to the data in the Directory, and the eastings and northings are then used to generate point data within a GIS. There are a number of licensed and open source GIS packages available (e.g. ArcGIS[4], MapInfo[5] and QGIS[6]). Areal (polygon) representations of unit postcodes are produced in the OS 'Code-Point with Polygons' product, available from Edina Digimap[7]. The projected coordinate system used to display geo-referenced data across Great Britain (i.e. England, Wales and Scotland) is the British National Grid[8].

The point-based longitudinal cohort study data is associated with pre-2001 geographies by means of a 'spatial join' (i.e. based on location) within a GIS. A

[1] www.afd.co.uk/
[2] http://census.edina.ac.uk//pcluts.html
[3] 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.
[4] Licensed software, available from http://www.esriuk.com/
[5] Licensed software, available from http://www.mapinfo.com/
[6] Open-source software, available from http://www2.qgis.org/en/site/
[7] http://digimap.edina.ac.uk/digimap/home
[8] WKID 27700, Authority EPSG

range of 1971, 1981 and 1991 Census geography boundaries are available to download from the UK Data Service[9]. One drawback of the use of mean unit postcode centroids in geo-referenced data is their 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 based on the unit postcodes of cohort members who live in more urbanised areas.

Availability of datasets

Because of the potentially disclosive nature of these datasets, geographic identifiers below Standard Statistical Region (STREG) are being released under Secure Access.

There are two versions of NCDS3 geographically linked data available, one set based on 1971 Census boundaries and the other set based on 1981 Census boundaries. Due to the theoretical 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 1971 and 1981), users can choose either, but not both, of these datasets.

[9] https://census.ukdataservice.ac.uk/get-data/boundary-data

Dataset	Description	Boundary Data Source (UK Data Service unless stated)
Interview Electoral Wards	1971 Electoral Ward Boundaries NCDS Sweep 3	English Electoral Wards, 1971 Welsh Electoral Wards, 1971 Scottish Electoral Wards, 1971
	1981 Electoral Ward Boundaries NCDS Sweeps 3 and 4	English Electoral Wards, 1981 Welsh Electoral Wards, 1981 Scottish Electoral Wards, 1981
	1991 Electoral Ward Boundaries NCDS Sweep 5	English Electoral Wards, 1991 Welsh Electoral Wards, 1991 Scottish Electoral Wards, 1991
Interview Standard Statistical Region (SSR)	Pre-1998 Standard Statistical Region Boundaries NCDS Sweeps 3, 4, and 5	English Standard Statistical Regions, pre-1998 Boundaries
Interview Enumeration District (ED)	1971 Enumeration District Boundaries NCDS Sweep 3	English Enumeration Districts, 1971 Welsh Enumeration Districts, 1971 Scottish Enumeration Districts, 1971
	1981 Enumeration District Boundaries NCDS Sweeps 3 and 4	English Enumeration Districts, 1981 Welsh Enumeration Districts, 1981 Scottish Enumeration Districts, 1981

Dataset	Description	Boundary Data Source (UK Data Service unless stated)
	1991 Enumeration District Boundaries NCDS Sweep 5	English Enumeration Districts, 1991 Welsh Enumeration Districts, 1991 Scottish Enumeration Districts, 1991
Interview District	1971 District Boundaries NCDS Sweep 3	English Districts, 1971 Welsh Districts, 1971 Scottish Districts, 1971
	1981 District Boundaries NCDS Sweeps 3 and 4	English Districts, 1981 Welsh Districts, 1981 Scottish Districts, 1981
	1991 District Boundaries NCDS Sweep 5	English Districts, 1991 Welsh Districts, 1991 Scottish Districts, 1991
Interview Westminster Parliamentary Constituency	1973 Westminster Parliamentary Constituency Boundaries NCDS Sweep 3	English Westminster Parliamentary Constituencies, 1973 Welsh Westminster Parliamentary Constituencies, 1973 Scottish Westminster Parliamentary Constituencies, 1973
	1983 Westminster Parliamentary Constituency Boundaries NCDS Sweeps 3 and 4	English Westminster Parliamentary Constituencies, 1983 Welsh Westminster Parliamentary Constituencies, 1983 Scottish Westminster Parliamentary Constituencies, 1983

Dataset	Description	Boundary Data Source (UK Data Service unless stated)
	2011 Westminster Parliamentary Constituency Boundaries NCDS Sweep 5	English Westminster Parliamentary Constituencies, 1991 Welsh Westminster Parliamentary Constituencies, 1991 Scottish Westminster Parliamentary Constituencies, 1991
Interview County	1971 County Boundaries NCDS Sweep 3	English Counties, 1971 Welsh Counties, 1971 Scottish Counties, 1971
	1981 County Boundaries NCDS Sweep 3 and 4	English Counties, 1981 Welsh Counties, 1981 Scottish Counties, 1981
	1991 County Boundaries NCDS Sweep 5	English Counties, 1991 Welsh Counties, 1991 Scottish Districts, 1991

Variables

NCDS Sweep 3 linked to geographical variables (based on 1971 Census boundaries):

SERIAL	NCDS research id
N3COUNTRY	Sweep 3 country of residence
N3STREG	Standard Statistical Region (pre-1998 boundaries)
N3CTY71	Counties (1971 boundaries)
N3DT71	Districts (1971 boundaries)
N3PCON73	Parliamentary constituencies (1973 boundaries)
N3WARD71	Wards (1971 boundaries)
N3ED71	Census Enumeration Districts (1971 boundaries)

NCDS Sweep 3 linked to geographical variables (based on 1981 Census boundaries):

SERIAL	NCDS research id
N3COUNTRY	Sweep 3 country of residence
N3STREG	Standard Statistical Region (pre-1998 boundaries)
N3CTY81	Counties (1981 boundaries)
N3DT81	Districts (1981 boundaries)
N3PCON83	Parliamentary constituencies (1983 boundaries)
N3WARD81	Wards (1981 boundaries)
N3ED81	Census Enumeration Districts (1981 boundaries)

NCDS Sweep 4 linked to geographical variables (based on 1981 Census boundaries):

SERIAL	NCDS research id
N4COUNTRY	Sweep 4 country of residence
N4STREG	Standard Statistical Region (pre-1998 boundaries)
N4CTY81	Counties (1981 boundaries)
N4DT81	Districts (1981 boundaries)
N4PCON83	Parliamentary constituencies (1983 boundaries)
N4WARD81	Wards (1981 boundaries)
N4ED81	Census Enumeration Districts (1981 boundaries)

NCDS Sweep 5 linked to geographical variables (based on 1991 Census boundaries):

SERIAL	NCDS research id
N5COUNTRY	Sweep 5 country of residence
N5STREG	Standard Statistical Region (pre-1998 boundaries)
N5CTY81	Counties (1991 boundaries)
N5DT81	Districts (1991 boundaries)
N5PCON83	Parliamentary constituencies (1991 boundaries)
N5WARD81	Wards (1991 boundaries)
N5ED81	Census Enumeration Districts (1991 boundaries)

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.

Raper, J., et al. (1992). Postcodes: the new geography, Longman Scientific, Harlow, Essex.

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.

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