

1970 British Cohort Study

Age 51 Geographical Identifiers

User Guide (Version 1)

May 2025





Economic and Social Research Council

Contact

Data queries: <u>help@ukdataservice.ac.uk</u>

Questions and feedback about this user guide: <u>clsdata@ucl.ac.uk</u>.

Authors

David Church, Liam Curran, Adelaida Sanchez-Galvez, George Ploubidis.

How to cite this guide

Church, D., Curran, L., Sanchez-Galvez, A., Ploubidis, G.B. (2025) *1970 British Cohort Study: Age 51 Geographical Identifiers*. London: UCL Centre for Longitudinal Studies

This guide was published in May 2025 by the UCL Centre for Longitudinal Studies.

Data citation and acknowledgement

You should cite the data and acknowledge CLS following the guidance from <u>cls.ucl.ac.uk/data-access-training/citing-our-data/</u>

Centre for Longitudinal Studies

Centre for Longitudinal Studies (CLS) UCL Social Research Institute University College London 20 Bedford Way, London WC1H 0AL cls.ucl.ac.uk

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre. It is home to a unique series of UK national cohort studies. It is part of the <u>UCL Social Research Institute</u>, based at the <u>IOE, UCL's</u> <u>Faculty of Education and Society</u>.

This document is available in alternative formats. Please email the Centre for Longitudinal Studies at clsdata@ucl.ac.uk

Contents

About the 1970 British Cohort Study1		
1. Introduction	2	
2. Geographical indicators	2	
2.1 Background	2	
2.2 Extent and nature of the data	3	
3. Linkage methods	5	
4. Linked geographical data	6	
4.1 Licensing and data access	6	
4.2 List of datasets	6	
4.3 Data documentation	7	
4.4 Identifiers	7	
4.5 Dataset description	7	
4.6 Output disclosure control	8	
5. References	9	
Appendix: Data sources for geographical identifiers	11	

About the 1970 British Cohort Study

The 1970 British Cohort Study (BCS70) is a longitudinal birth cohort study, following a nationally representative sample of over 17,000 people born in Britain in a single week in 1970.

We have surveyed cohort members throughout their childhood and adult lives, mapping their individual trajectories and creating a unique resource for researchers. It is one of very few longitudinal studies following people of this generation anywhere in the world.

Featuring a range of objective measures and rich self-reported data, BCS70 covers an incredible amount of ground and can be used in research on many topics

Evidence from BCS70 has illuminated important issues for our society across five decades. Key findings include how reading for pleasure matters for children's cognitive development, why grammar schools have not reduced social inequalities, and how childhood experiences can impact on mental health in mid-life.

Every day, researchers from across the scientific community are using this important study to make new connections and discoveries.

1. Introduction

The BCS70 Age 51 Survey, (or 'Life in Your Early 50s' Survey as known to study members) was conducted between 2021 and 2024 when participants were aged 51-53 years. It was the 11th sweep of BCS70.

This sweep was designed and managed by the Centre for Longitudinal Studies (CLS) at the UCL Faculty of Education and Society (IOE) and fieldwork was conducted by NatCen and Verian (formerly Kantar).

The main aim of the Age 51 Survey was to gather information to understand midlife outcomes across various life domains and their determinants over the life course, building on the wealth of data collected in previous sweeps. The study was initially planned and designed to be conducted face-to-face with fieldwork commencing in 2020. However, the COVID-19 pandemic meant that face-to-face interviewing was not feasible from May 2020 until early 2022. As such the protocol was changed so that interviews could be conducted by video-call. Fieldwork was launched using a video-only approach in November 2021. Face-to-face interviewing began in April 2022, though participants could still opt to participate via video if they preferred. Once mainstage fieldwork was complete, those who had not participated were invited to complete a short version of the questionnaire via web (known as the 'mop-up' survey). Emigrants were also invited to take part in this short web-survey. For more information see the BCS age 51 user guide.

2. Geographical indicators

2.1 Background

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 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 used 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).

CLS 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).

2.2 Extent and nature of the data

BCS70 cohort members live throughout England, Scotland and Wales, therefore a short explanation of the extent and nature of the data, and differences across the constituent nations of Great Britain may be helpful. England and Wales use the same naming conventions across different geographies. Post-devolution, Scotland has adopted slightly different naming conventions. For example, in both 2011 and 2021 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 England & Wales and Scotland. The projected coordinate system used to display geo-referenced data across Great Britain (i.e. England, Wales and Scotland) is the British National Grid¹. Ordnance Survey products (e.g. MasterMap, AddressBase, OS OpenData) are available for Great Britain (i.e. excluding Northern Ireland). Only those cohort members who live in Great Britain have been assigned geo-identifiers – there are some participants who live overseas who are not included in the geographical identifiers dataset.

¹ https://epsg.io/27700

3. Linkage methods

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 CLS using a range of specialist software products from AFD². This postcode data is then used to generate point data, usually within a GIS. There are several licensed and open source GIS packages available (e.g. ArcGIS³, MapInfo⁴ and QGIS⁵). The primary data source for spatialising longitudinal cohort study data within this software is the ONS Postcode Directory, available from the Office for National Statistics website⁶. This dataset has been released guarterly 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 centroid⁷. The August 2024 ONS Postcode Directory was used to link BCS70 sweep 11 to the set of geographical indicators described in Section 4.5. The BCS70 sweep 11 data was linked to the August 2024 ONSPD on postcode (specifically, PCDS) by means of a 'one-to-first' join in ArcGIS Pro (i.e. the participant's postcode will be joined to the first matching postcode in the ONSPD data).

² <u>http://www.afd.co.uk/</u>

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

⁴ Licensed software, available from <u>http://www.mapinfo.com/</u>

⁵ Open-source software, available from <u>https://www.qgis.org/</u>

⁶ <u>https://www.ons.gov.uk/methodology/geography/geographicalproducts/postcodeproducts</u>

⁷ 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'). Only 0.68% of current postcodes in the August 2024 ONSPD have a positional quality indicator greater than 1.

4. Linked geographical data

4.1 Licensing and data access

The BCS70 age 51 geographical data have been processed by CLS and supplied to the UK Data Service. All data users need to be registered with the UK Data Service and sign the UK Data Service End User License. Details of how to do this are available at <u>ukdataservice.ac.uk/get-data/how-to-access/registration</u>.

The BCS70 age 51 geographical identifiers are potentially disclosive and can be accessed as controlled data from the UK Data Service SecureLab. Applicants wishing to access this data need to abide by the terms and conditions of the UK Data Service Secure Access licence.

To gain access to the geographical identifiers, researchers must submit a Secure Access application to the UK Data Service detailing the intended analysis and provide a justification as to why this data is requested. Application guidance can be found at <u>ukdataservice.ac.uk/find-data/access-conditions/secure-application-requirements/apply-to-access-non-ons-data/</u>

Subject to approval, researchers can upload their own datasets to their UK Data Service SecureLab account in order to link to these geo-identifiers (e.g. linking Census, meteorological, crime, housing or other socio-economic data to the cohort studies).

4.2 List of datasets

Datasets are flat in structure, with one row per cohort member.

There are two versions of BCS70 age 51 geographically linked data available, one set based on 2011 Census boundaries and the other set based on 2021 Census boundaries.

Table 1: List of available datasets

Name of the dataset	Content summary
bcs11_age51_geographical_identifiers_restricted_2011.sav	BCS70 age 51 geographical identifiers: 2011 Census boundaries
bcs11_age51_geographical_identifiers_restricted_2021.sav	BCS70 age 51 geographical identifiers: 2021 Census boundaries

4.3 Data documentation

The following data dictionaries contain the following metadata: variable names and labels, value labels, missing values. The order of the variables in the metadata documents is consistent with the variable ordering in the datasets.

Table 2: Data documents

Name of the document	Content summary
bcs11_age51_geography_2011_data_dictionary.xlsx	Variable names and labels, value labels, missing values.
bcs11_age51_geography_2021_ data_dictionary.xlsx	Variable names and labels, value labels, missing values.

4.4 Identifiers

Individual identifiers

The data are identified with the variable BCSID. This is the same research ID used for the rest of the cohort data available at the UK Data Service. This enables the data to be easily merged with one another.

4.5 Dataset description

Table 3: BCS70 Age 51 Geographical Identifiers (2011)

Variable Name	Description
BCSID	BCS70 research ID
B11CTRY	Country of interview
B11RGN	December 2020 Region of interview

Variable Name	Description	
B110SWARD	May 2023 Ward	
B11CASWARD	January 2003 Census Area Statistics Ward	
B11OSLAUA	April 2023 Local Authority	
B11PCON	July 2024 Westminster Parliamentary Constituency	
B11IMD	IMD Overall Rank - Eng 2019, Sco 2020, Wal 2019, NI 2017	
B11IMDD	IMD Overall Rank Decile	
B11OA11	2011 Output Area	
B11LSOA11	2011 Lower Super Output Area	
B11MSOA11	2011 Middle Super Output Area	
B11WZ11	2011 Workplace Zone	
B110AC11	2011 Output Area Classification	
B11RUC11	2011 Rural-Urban Classification	

Table 4: BCS70 Age 51 Geographical Identifiers (2021)

Variable Name	Description
BCSID	BCS70 research ID
B11CTRY	Country of interview
B11RGN	December 2020 Region of interview
B11OSWARD	May 2023 Ward
B11CASWARD	January 2003 Census Area Statistics Ward
B11OSLAUA	April 2023 Local Authority
B11PCON	July 2024 Westminster Parliamentary Constituency
B11IMD	IMD Overall Rank - Eng 2019, Sco 2020, Wal 2019, NI 2017
B11IMDD	IMD Overall Rank Decile
B11OA21	2021 Output Area
B11LSOA21	2021 Lower Super Output Area
B11MSOA21	2021 Middle Super Output Area

See the Appendix for the data sources from which the BCS70 Age 51 Geographical Identifiers are derived.

4.6 Output disclosure control

Access to this controlled data is only available via the UK Data Service SecureLab.

The UK Data Service will always perform a certain level of disclosure control on the outputs generated by researchers, as outlined in their SDC Handbook which can be downloaded from: securedatagroup.org/sdc-handbook/

The two UK Data Service Secure Lab rules of thumb that will be applied to all outputs are:

- Threshold rule: No cells should contain less than 10 observations
- Dominance rule: No observation should dominate the data to a huge extent

5. 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.

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: Data sources for geographical identifiers

(Please see overleaf)

Dataset	Description	Boundary Data Source (ONS Postcode Directory February 2024)
		English Electoral Wards, May 2023
		Welsh Electoral Wards, May 2023
	2023 Ward Boundaries	Scottish Electoral Wards, May 2023
Interview Wards		
	2003 Census Area Statistic Ward ⁸	English Census Area Statistic Wards, 2003
		Welsh Census Area Statistic Wards, 2003
		Scottish Census Area Statistic Wards, 2003
		English Output Areas, 2011
		Welsh Output Areas, 2011
	2011 Output Area	Scottish Output Areas, 2011
Interview Output Area (OA)		
	2021 Output Area	English Output Areas, 2021
		Welsh Output Areas, 2021
		Scottish Output Areas, 2021
Interview Lower Super	2011 Lower Super Output Area	English Lower Super Output Areas, 2011
		Welsh Lower Super Output Areas, 2011
		Scottish Datazones, 2011

⁸ Please see the ONS publication '<u>A Beginner's Guide to UK Geography 2023'</u> for definitions and an overview of the various geographical units.

Dataset	Description	Boundary Data Source (ONS Postcode Directory February 2024)
	2021 Lower Super Output Area	
		English Lower Super Output Areas, 2021
		Welsh Lower Super Output Areas, 2021
		Scottish Data Zones , 2022
Interview Middle	2011 Middle Super Output Area	English Lower Super Output Areas, 2011
		Welsh Lower Super Output Areas, 2011
		Scottish Intermediate Geographies, 2011
Super Output Area		
(MSOA)	2021 Middle Super Output Area	English Lower Super Output Areas, 2021
		Welsh Lower Super Output Areas, 2021
		Scottish Intermediate Geographies, 2022
	2022 Local Authority District/Unitory	English Administrative Districts April 2022
Interview Local Authority District		
	Authority	English Unitary Authorities, April 2023
		Welsh Unitary Authorities, April 2023
		Scottish Council Areas, April 2023