Next Steps

Sweep 9 Geographical Identifiers

User Guide

January 2025

Contact

Data queries: help@ukdataservice.ac.uk

Questions and feedback about this user guide: clsdata@ucl.ac.uk.

Authors

David Church, Liam Curran, Morag Henderson, Adelaida Sanchez-Galvez

How to cite this guide

Church, D., Curran, L., Henderson, M., Sanchez-Galvez, A. (2025) *Next Steps: Sweep 9 Geographical Identifiers*. London: UCL Centre for Longitudinal Studies

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

Data citation and acknowledgement

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

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 UCL Social Research Institute, based at the IOE, UCL's Faculty of Education and Society.

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

Contents

About Next Steps

Next Steps is a longitudinal cohort study, following a nationally representative group of nearly 16,000 people born in England in 1989-90. The study began when cohort members were 14 years old, with the first seven sweeps occurring annually from 2004 to 2010. Sweep 8 documented early adulthood experiences at Age 25 whilst Sweep 9 captures how the cohort members are faring at Age 32. These adulthood sweeps have a wider scope than previously, and include measures of health, wellbeing, family formation and labour market outcomes (among others), providing unparalleled insight into the many different aspects of this millennial generation's lives. Next Steps has captured incredibly rich information about participants' family backgrounds and educational trajectories during adolescence and pathways to adulthood.

A vital source of evidence, Next Steps has had a major influence on national education policy and cast light on a wide range of important social issues, including the effects of zero hours contracts and bullying.

1. Introduction

The Next Steps Age 32 Sweep (Sweep 9) took place between April 2022 and September 2023. It was designed and managed by the Centre for Longitudinal Studies (CLS) at the UCL Faculty of Education and Society (IOE), and fieldwork was carried out by Ipsos. It was funded by the Economic and Social Research Council. The Age 32 Sweep is the ninth sweep of the study. The first seven sweeps of data were collected annually between 2004 and 2010, when the study was run by the Department for Education and known as the Longitudinal Study of Young People in England (LSYPE). The eighth sweep was run when cohort members were aged around 25, between 2015 and 2016, by CLS. In addition, three online surveys were conducted during the COVID-19 pandemic, when data was also collected simultaneously from participants in the MRC National Survey of Health and Development, 1970 British Cohort Study, 1958 National Child Development Study, and the Millennium Cohort Study. For more information see the Next Steps Sweep 9 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, 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 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).

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 principle consideration is protecting the identity of cohort members, particularly in the current era of 'open data' and increasing linkage of previously disparate administrative datasets.

At present, 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

Although the vast majority of Next Steps cohort members live in England, there are a small number who have moved to Scotland, Wales and Northern Ireland over the

years, therefore a short explanation of the extent and nature of the data, and differences across the constituent nations of the UK 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 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 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 Grid. 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 Directory, 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 Grid. Only those cohort members who live in the United Kingdom have been assigned geo-identifiers – there are some participants who live overseas who are not included in the geographical identifiers dataset.

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 AFD[1]. 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. ArcGIS[2],

---

[1] http://www.afd.co.uk/

[2] Licensed software, available from http://www.esriuk.com/

MapInfo[3] and QGIS[4]). 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 website[5]. 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 centroid[6]. The August 2024 ONS Postcode Directory was used to link Next Steps Sweep 9 to the set of geographical indicators described in Section 4.5. The Next Steps Sweep 9 data was linked to the August 2024 ONSPD on postcode (specifically, PCDS) by means of a 'one-to-first' join in ArcPro (i.e. the participant's postcode will be joined to the first matching postcode in the ONSPD data).

4. Linked geographical data

4.1 Licensing and data access

The Next Steps Sweep 9 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 Licence. Details of how to do this are available at ukdataservice.ac.uk/get-data/how-to-access/registration.

The Next Steps Sweep 9 geographical identifiers are potentially disclosive and can be accessed as controlled data from the UK Data Service SecureLab. Applicants

---

[3] Licensed software, available from http://www.mapinfo.com/

[4] Open-source software, available from https://www.qgis.org/

[5] https://www.ons.gov.uk/methodology/geography/geographicalproducts/postcodeproducts

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

wishing to access this data need to abide by the terms and conditions of the UK Data Service Secure Access licence.

In order to gain access to the geographical identifiers, researchers must submit an 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 ukdataservice.ac.uk/find-data/access-conditions/secure-application-requirements/apply-to-access-non-ons-data/

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 Next Steps Sweep 9 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
ns9_2011_geography.sav	Next Steps sweep 9 geographical identifiers: 2011 Census boundaries
ns9_2021_geography.sav	Next Steps sweep 9 geographical identifiers: 2021 Census boundaries

4.3 Data documentation

The following documents contain the following metadata: variable 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
ns9_2011_geography_metadata.xlsx	Metadata spreadsheet containing variable labels, value labels, missing values.
ns9_2021_geography_ metadata.xlsx	Metadata spreadsheet containing variable labels, value labels, missing values.

4.4 Identifiers

Individual identifiers

The data are identified with the variable NSID. 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: Next Steps Sweep 9 Geographical Identifiers (2011)

Variable Name	Description
NSID	Next Steps cohort member identifier
W9CTRY	Wave 9 Country
W9RGN	Wave 9 Region
W9DIST23	Wave 9 District (Local Authority) as at 04 2023
W9WARD23	Wave 9 Electoral Ward as at 05 2023
W9PCON14	Wave 9 Westminster Parliamentary Constituency
W9CASWARD	Wave 9 Census Area Statistics Ward
W9OA11	Wave 9 2011 Output Area
W9LSOA11	Wave 9 2011 Lower Super Output Area
W9MSOA11	Wave 9 2011 Middle Super Output Area
W9RU11IND	Wave 9 2011 Rural-Urban Indicator
W9OAC11	Wave 9 2011 Output Area Classification Code
W9IMD	Wave 9 Index of Multiple Deprivation 2019 (Overall Rank)
W9WZ11	Wave 9 2011 Workplace Zones

Table 4: Next Steps Sweep 9 Geographical Identifiers (2021)

Variable Name	Description
NSID	Next Steps cohort member identifier
W9CTRY	Wave 9 Country
W9RGN	Wave 9 Region
W9DIST23	Wave 9 District (Local Authority) as at 04 2023
W9WARD23	Wave 9 Electoral Ward as at 05 2023
W9PCON14	Wave 9 Westminster Parliamentary Constituency
W9CASWARD	Wave 9 Census Area Statistics Ward
W9IMD	Wave 9 Index of Multiple Deprivation 2019 (Overall Rank)
W9OA21	Wave 9 2021 Output Area
W9LSOA21	Wave 9 2021 Lower Super Output Area
W9MSOA21	Wave 9 2021 Middle Super Output Area

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](#))
Interview Wards	2023 Ward Boundaries 2003 Census Area Statistic Ward[7]	English Electoral Wards, May 2023 Welsh Electoral Wards, May 2023 Scottish Electoral Wards, May 2023 Northern Ireland Electoral Wards, May 2023 English Census Area Statistic Wards, 2003 Welsh Census Area Statistic Wards, 2003 Scottish Census Area Statistic Wards, 2003 Northern Ireland Electoral Wards, 2011[8]
Interview Output Area (OA)	2011 Output Area 2021 Output Area	English Output Areas, 2011 Welsh Output Areas, 2011 Scottish Output Areas, 2011 Northern Ireland Output Areas, 2011 English Output Areas, 2021 Welsh Output Areas, 2021 Scottish Output Areas, 2021

---

[7] Please see the ONS publication 'A Beginner's Guide to UK Geography 2023' for definitions and an overview of the various geographical units.

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

Dataset	Description	Boundary Data Source ([ONS Postcode Directory February 2024](#))
		Northern Ireland Small Areas, 2021
Interview Lower Super Output Area (LSOA)	2011 Lower Super Output Area 2021 Lower Super Output Area	English Lower Super Output Areas, 2011 Welsh Lower Super Output Areas, 2011 Scottish Datazones, 2011 Northern Ireland Lower Super Output Areas, 2011 English Lower Super Output Areas, 2021 Welsh Lower Super Output Areas, 2021 Scottish Lower Super Output Areas, 2021 Northern Ireland Super Output Areas, 2021[9]
Interview Middle Super Output Area (MSOA)	2011 Middle Super Output Area 2021 Middle Super Output Area	English Lower Super Output Areas, 2011 Welsh Lower Super Output Areas, 2011 Scottish Intermediate Geographies, 2011 English Lower Super Output Areas, 2021 Welsh Lower Super Output Areas, 2021 Scottish Intermediate Geographies, 2021

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

Dataset	Description	Boundary Data Source ([ONS Postcode Directory February 2024](#))
Interview Local Authority District	2023 Local Authority District/Unitary Authority	English Administrative Districts, April 2023 English Unitary Authorities, April 2023 Welsh Unitary Authorities, April 2023 Scottish Council Areas, April 2023 Northern Ireland District Councils, 2001
Interview Westminster Parliamentary Constituency	2014 Westminster Parliamentary	English Westminster Parliamentary Constituencies, December 2014 Welsh Westminster Parliamentary Constituencies, December 2014 Scottish Westminster Parliamentary Constituencies, December 2014 Northern Ireland Westminster Parliamentary Constituencies, December 2014