

Getting started with the data

CENTRE FOR LONGITUDINAL STUDIES



Economic and Social Research Council

CENTRE FOR LONGITUDINAL STUDIES

This section

- Available resources
- Accessing the data
- Key IDs and other data protocols
- Merging data within and across sweeps
- Study design and sample weights
- Non-response and attrition
- Where to go for more information

Available resources https://cls.ucl.ac.uk/

- User guides
 - Overview of measures
 - Response and weights
- Questionnaires
 - Exact question wording
 - Questionnaire routing
 - Variable names
- Data documentation
 - Data notes
 - Coding frames
 - Variables lists, including derived variables
- Technical reports
 - Sample and questionnaire design, development
 - Fieldwork, response, ethics
 - Coding, editing
- Data dictionaries
- Cohort profiles e.g.
 - Power, C and Elliott, J (2006) Cohort Profile: 1958 British birth cohort (National Child Development Study) International Journal of Epidemiology, 35(1), 34-41
 - Joshi, H and Fitzsimons, E (2016) The Millennium Cohort Study: the making of a multi-purpose resource for social science and policy. Longitudinal and Life Course Studies, 7(4), 409-430.
- Previous journal publications

	≜UC	L
co	MD-19 Our studies Our research O Publications and resources O Data access and training	
• Ho	ne	
Ρ	ublications and	
re	sources	
Sear stud pape	ch the complete documentation from each of our four es. You can also explore our extensive library of working est, binefings, impact documents and videos.	
Ρι	ublications and resources	
	Bibliography Ov bibliography is an excellent resource for finding publications based on data from our four studies. It includes over 4.000 publications and is warchable by study year, author, journel, title and abstract.	⊳
	Data documentation Here you will find all the documentation for our four studies. This includes survey questionnaires, interviewer instructions, showcards, technical reports, survey reports, and derived variables.	⊳
	Guide to the documentation Here you can find out more about the different types of documents we publish on our website for each of our studies.	⊳
	Briefings and impact Ox briefings and impact library includes summaries of our research findings as well as reports highlighting the impact of our cohort studies.	⊳
	CLS working papers Over working papers series features research based on our four cohort studies and dates back to 1983. Topics range from social inequatities and mobility. Io physical health and cognitive development. Other papers in the series seek to improve the practice of	⊳

CENTRE FOR

STUDIES

LONGITUDINAL

https://cls.ucl.ac.uk/publications-and-resources/

Resources available: UK Data Service

Details Documentation Resou	rces	Access data
ocumentation		
Title	File name	Size (MB)
UCL Cohorts: COVID-19 Data Variable Catalogue	covid-19_w1-3_variable_lookup.xlsx	0.05
UCL Cohorts: COVID-19 Survey Technical Report	ucl_cohorts_covid-19_survey_technical _report.pdf	0.86
UCL Cohorts: COVID-19 Survey User Guide	ucl_cohorts_covid-19_survey_user_guid e.pdf	1.26
UCL Cohorts: COVID-19 Wave 1 Survey Questionnaire	covid-19_online_survey_questionnaire_ wave_1_april_2020.pdf	0.49
UCL Cohorts: COVID-19 Wave 2 Survey Questionnaire	covid-19_online_survey_questionnaire_ wave_2_september_2020.pdf	0.62
UCL Cohorts: COVID-19 Wave 3 Survey Questionnaire	covid-19_online_survey_questionnaire_ wave_3_february_2021.pdf	0.95
UK Data Archive Citation File for Study 8658	UKDA_Study_8658_Information.htm	0
UK Data Archive Data Dictionaries	ukda_data_dictionaries.zip	0.23
UK Data Archive ReadMe File for Study 8658	read8658.htm	0

CENTRE FOR	
LONGITUDINA	L
STUDIES	

Millennium Cohort Study: Seventh Survey, 2018						
Details Documentation Resource	5	Access data				
Documentation						
Title	File name 🌲	Size (MB)				
Information about the MCS Longitudinal Data Dictionary	mcs longitudinal data dictionary 2021-02- 03.xlsx	1.35				
MCS Data Handling Guide with syntax in R, STATA and SPSS, August 2020	<u>mcs data handling guide ed1 2020-08-1</u> <u>0.pdf</u>	1.2				
MCS7 Derived Variables User Guide, 2nd Edition, December 2020	mcs7 dv user guide age17 ed2 2020 12 08.pdf	0.4				
MCS7 Parent Online (CAWI) Questionnaire	<u>mcs7-parent-online-cawi-questionnaire.pd</u> <u>f</u>	1.11				
MCS7 Technical Report	mcs7 technical report.pdf	1.65				
MCS7 User Guide, 2nd Edition, December 2020	mcs7 user guide age17 ed2 2020 12 08. pdf	1.03				
MCS7 Young Person Interview	mcs7-young-person-interview.pdf	1.18				
MCS7 Young Person Online (CAWI) Questionnaire	<u>mcs7-young-person-online-cawi-questionn</u> <u>aire.pdf</u>	0.93				
MCS7 Young Person Self-completion Questionnaire	mcs7-young-person-self-completion-questi onnaire.pdf	0.95				
UK Data Archive Citation File for Study 8682	UKDA Study 8682 Information.htm	0				
UK Data Archive Data Dictionaries	ukda data dictionaries.zip	0.1				
UK Data Archive ReadMe File for Study 8682	read8682.htm	0				

Data freely available to researchers, government analysts and third sector



CENTRE FOR

STUDIES

UK Data Service



ukdataservice.ac.uk

Access to different types of data



Access to data held by the UK Data Service varies depending on how the data is classified:

Tier 1: <u>End User Licence</u> (EUL) for access to data with a low level of sensitivity and disclosivity.

- Most of our data are available under this licence.
- Your application is authorised directly by the UK Data Service, and you can download the data directly from there.

•Tier 2a: <u>Special Licence</u> (SL) for access to moderately sensitive or disclosive data. Access through the UK Data Service and application approved by CLS before you can download the data.

•Tier 2b: <u>Secure Access Licence</u> (SA) for access to the most sensitive and/or potentially disclosive data. Access through the UK Data Service and attend a specialised training course.

CENTRE FOR Registering with the UK Data Service LONGITUDINAL **STUDIES** 🕅 UK Data Service Q Search Login | Registe Find data Deposit data Training and events About Help Login If you are in the UK and your organisation is part Select your organisation to continue of the UK Access Management Federation (UKAMF) and on this list of federation members. You previously selected: UCL (University College London) you can use the username and password issued to you by your organisation to register with the UK Data Service. Start typing the name of you If you are not in the UK, or your organisation is not on the list of federation members, request a username Continue If you have a username beginning 'ukd', use t If you are a Secure Lab user, login to Secure Lab 📈 UK Data Service Q Search the site. 🗊 Login | Register Training and events Deposit data Learning hub Help Application for a username If you are from an organisation not using Your contact details federated access management complete this form to request a username. Title: Please select First name: * her education or your organisation uses federated access management and is on this list Last name: * of federation members, do not complete this form log in to the UK Data Service using your loc Address line 1: * ame we issue username and pe will only give you access to resources managed by Address line 2: the UK Data Archive and UK Data Service. Address line 3: Town / City: *

1. Click 'Login/Register'

2. On the Login page begin typing your organisation name. Select the organisation required

3. Click 'Continue'

4. Your own organisation login page will then be displayed. Login with your usual username and password

5. Complete the registration form with your details, selecting other options as required

6. Agree to the End User Licence (EUL), which outlines the terms and conditions of use of the Service

7. Click 'Register'.

Once registration is complete you will be able to download/order or request access to data

Accessing CLS cohort data I

Cohort	Link
NCDS	http://discover.ukdataservice.ac.uk/series/?sn=2000032
BCS70	http://discover.ukdataservice.ac.uk/series/?sn=200001
Next Steps	http://discover.ukdataservice.ac.uk/series/?sn=2000030
MCS	http://discover.ukdataservice.ac.uk/series/?sn=2000031



The Millennium Cohort Study (MCS), which began in 2000, is conducted by the Centre for Longitudinal Studies (CLS). It alms to chart the conditions of social, economic and health advantages and disadvantages facing children born at the start of the 21st century The study has been tracking the 'Millennium children' through their early childhood years and plans to follow them into adulthood. It also provides a basis for comparing patterns of development with the preceding cohort studies (the National Child Development Study (NCDS) and the 1970 Birth Cohort Study (BCS70).

Abstra	ct FAQ's Resources		Access data
ccess o	data		
N 33359			
illenniu	m Cohort Study – Survey and Biomeasures Data		
5N	Study description	Explore online	Select
3756	Millennium Cohort Study, Sweeps 1-7, 2001-2019: Exact Participation Dates: Secure Access		
3755	Millennium Cohort Study, Sweeps 1-7, 2001-2019: Demographics, Language and Religion: Secure Access		
3754	Millennium Cohort Study, Sweeps 1-7, 2001-2019; Self-Reported Health, Behaviour and Fertility; Secure Access		
3753	Millennium Cohort Study, Sweeps 1-7, 2001-2019: Socio- Economic, Accommodation and Occupational Data: Secure Access		
3682	Millennium Cohort Study: Seventh Survey, 2018		
3172	Millennium Cohort Study: Longitudinal Family File, 2001-2018		
3156	Millennium Cohort Study: Sixth Survey, 2015		
7464	Millennium Cohort Study: Fifth Survey, 2012		
7261	Millennium Cohort Study: First Survey, Health Visitor Survey, 2002-2003		
7238	Millennium Cohort Study: Fourth Survey, Physical Activity Data, 2008		
5411	Millennium Cohort Study: Fourth Survey, 2008		
5795	Millennium Cohort Study: Third Survey, 2006		
5559	Millennium Cohort Study: Survey of Mothers who Received Assisted Fertility Treatment, 2003		
5350	Millennium Cohort Study: Second Survey, 2003-2005		
4683	Millennium Cohort Study: First Survey, 2001-2003		

CENTRE FOR LONGITUDINAL STUDIES

Accessing CLS cohort data II

Assign dataset to a project						
Before you can download a dataset or request access, you must assign it to one of your projects or create a new project for it. Once assigned, you can access datasets via the Projects section.						
Awaiting assignment to projects Si			Select all datasets			
SN	¢	Dataset	·			
8172		Millennium Cohort Study: Longitudinal Family File, 2001-2018	3			
		Millennium Cohort Study: Seventh Suprey, 2018				

Title: *	UK Data Service
	235 characters remainin
Project type: *	Non-commercial -
Abstract: *	Project created for demonstration purposes to show the process of creating new projects and downloading datasets.
	Please include a short description of the project and its benefits (100

Before downloading the data:

- Click on Request Access
- Click on Complete actions
- Agree to standard 'End User Licence'

CENTRE FOR

STUDIES

LONGITUDINAL

Read and agree extra conditions

Choose data format and download zip file

- SPSS
- STATA
- TAB
- Data sent to your email address

Name	Contents	Structure	Identifier
NS8_2015_Main_Interview	Modules 1 to 7	Flat	NSID
NS8_2015_Self_Completion	Module 8	Flat	NSID
NS8_2015_Partnerships	Relationship histories	Hierarchical	NSID, W8RELID
NS8_2015_Children	Details of children of CM	Hierarchical	NSID, W8CHID
NS8_2015_Household_Members	Details of members living in same household as CM	Hierarchical	NSID, W8HHMID
NS8_2015_Activity_History	Activities and Employment histories	Hierarchical	NSID, W8HISTID
NS8_2015_Benefits	Details of individual benefits received	Hierarchical	NSID, W8BENID
NS8_2015_Income_Unfolding_brackets	Unfolding brackets questions for payments and income	Flat	NSID
NS8_2015_Benefits_Unfolding_brackets	Unfolding brackets questions for benefits	Hierarchical	NSID, W8BENID
NS8_2015_Derived_variables	Derived variables	Flat	NSID

Key identifiers (IDs)

Cohort	Key cohort identifier	Key cohort member/family identifier format	
NCDS 1958	NCDSID	7 characters: N followed by 5 digits, and a single character e.g. N10016V	Cohort member/family identifier: Every cohort member (or for the MCS family) has the same ID across sweeps
BCS70 1970	BCSID	7 characters: B followed by 5 digits, and a single character e.g. B25819Z	Use these IDs to link datasets
Next Steps 1989/90	NSID	8 characters: NS followed by 5 digits and a single character e.g. NS21140C	W8xxID used in particular files to denote relationship, child, HH member etc
MCS 2000/02	MCSID	7 characters: M followed by 5 digits, and a single character e.g. M10029A	 CNUM Cohort members, 1, 2 (twins) or 3 (triplets) PNUM Person number, for everyone else in the family apart from cohort members: parents, siblings, grandparents, etc

CENTRE FOR

STUDIES

LONGITUDINAL

File structures: Flat v hierarchical

Type of file	Format	Examples in the cohorts	BCSID	Sex	Country	Emp
structure			B567689A	1	1	1
Flat	1 record per case	NCDS, BCS70, Next Steps – main data files	B567689A	1	1	3
		MCS – family files	B567689A	1	1	3
	1 or more records per case	Household files	23070037	-	-	3
		Activities of a compleximent histories	B467921B	2	1	1
		Activities e.g. employment histories	B879255C	2	2	5
Hierarchical		Relationship histories	B879255C	2	2	2
		Time use diaries	B297614D	1	1	6
		MCS - Person within family	B297614D	1	1	1
			B349725E	1	3	3

	BCSID	Sex	Country	Emp1	Emp2	Emp3
	B567689A	1	1	1	3	3
	B467921B	2	1	1		
CENTRE FOR	B879255C	2	2	5	2	
LONGITUDINAL	B297614D	1	1	6	1	
	B349725E	1	3	3		

Merging data within and across sweeps

- 1. Identify appropriate files
 - Establish number of cases in target population
- 2. Check file structure: flat v hierarchical
 - Transform if necessary
- 3. Identify merging variables:
 - 1. unique 'key' cohort ID (member or family)
 - 2. other IDs depending on merge (appear in both files)
 - Check the same variable name (case sensitive, changed across sweep etc)
 - Create identical variable name if necessary
- 4. Check merged correctly

CENTRE FOR LONGITUDINAL

STUDIES

	Institute of Education				
	Millennium Cohort Study				
sweep etc)	Data Handling Guide with syntax in R, STATA and SPSS				
	August 2020				
	CENTRE FOR LONGITUDINAL STUDIES				

Study design (sampling and sample weights) Studies are representative of...

CENTRE FOR LONGITUDINAL STUDIES

Cohort	Population	Sample and study design	Design weight(s)		
NCDS 1958	All born in GB in one week	Total sample: Captured 98% of the total births in	None		
BCS70 1970	Il born in GB in one week Total sample: Captured 95-98% of the total births in GB in the target week			None	
Next Steps 1989/90	Young people in England in 2004 born between 01/09/89 and 31/8/90	Complex sample design: probability Maintained schools* 2-stage sample Stage 1: Schools Stage 2: Pupils within schools.	ty sampling ling procedure lext_Steps_Longitudinal_	psu (SampPSU) strata (SampStratum) _File	
MCS 2000/02	Children born in 2000 /2002 and living in the UK at age nine months,	Complex sample design: probat Stage 1: A selection of areas of resi Stage 2: Disproportionately stratifie disadvantage, and UK country	Longitudinal Family File		

Non-response and attrition

CENTRE FOR

NGITUDINAL

- Distinction between unit (respondents') non-response and item nonresponse
 - Unit non-response (not responding to a particular sweep)
 - Non-response is common in longitudinal surveys
 - Item non-response i.e. not answering some questions
 - · tends to be less of an issue in the cohorts
- Missing data threat to representativeness
 - Potential for bias since respondents are often systematically different from nonrespondents



Dealing with unit non-respons

- Case-wise deletion i.e. ignoring non-response (unle
 - Any individual in a data set is deleted from an ana variable in the analysis
 - Straightforward, but doesn't deal with any non-res
- Non-response weights

CENTRE FOR

STUDIES

LONGITUDINAL

- Adjust the sample composition to take account or respondents.
- Provided in MCS (govwt2 = overall in MCS7 for v Steps (W8FINWT = final weight for age 25 surve https://cls.ucl.ac.uk/wp-content/uploads/2020/09/MCS7-user-guide-A https://cls.ucl.ac.uk/wp-content/uploads/2017/11/5545age_25_surve
- Other more advanced methods e.g.multiple imputati
 - MI involves the generation of multiple copies of th values are replaced by imputed values sampled f distribution given the observed

Handling missing data in the National Child Development Study

User guide (Version 2)

July 2021

Institute of Education





and Social



Where to go for more information

CENTRE FOR LONGITUDINAL STUDIES



Economic and Social Research Council



CLOSER Learning Hub: https://learning.closer.ac.uk/

CLOSER provides training and resources for students and early-career researchers to *"maximise the use, value and impact of longitudinal research"*

Learning Hub – demonstration video: <u>https://youtu.be/Z_bFCClq2Dc</u>

A Learning modules	Teaching resources	Research case studies	Explore by topic		I	Q	≣ Glossary	 *
Learning mo	odules	Introduction to la Study design Data harmonisat Understanding r Analysing longit Research comm	ongitudinal studies tion netadata udinal data unication					
	1 more	1/2	1	A Real	1.3	2		



CENTRE FOR LONGITUDINAL STUDIES

https://www.ncrm.ac.uk/resources/

Resources available: UK Data Service learning hub



CENTRE FOR LONGITUDINAL

STUDIES

https://ukdataservice.ac.uk/learning-hub/

Resources available: CLS website

https://cls.ucl.ac.uk/







So we've covered



- An introduction to birth cohorts
- Some of the content by subject areas
- Examples of the types of analysis
- Getting started with the data
- Where to go for more information



Thank you Any questions?





Economic and Social Research Council