Institute of Education

UCL

Millennium Cohort Study:

Longitudinal Family and Cohort Member Files

User Guide (Version 3)

February 2026

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

Contact

Data queries: help@ukdataservice.ac.uk

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

Authors

Agalioti-Sgompou Vilma and Johnson Jon (2020), Rachel Rosenberg and Aida Sanchez-Galvez (2026)

How to cite this guide

Agalioti-Sgompou V., Johnson J., Rosenberg R., Sanchez-Galvez A. (2026) *Millennium Cohort Study: Longitudinal Family and Cohort Member Files (Version 3).* UCL Centre for Longitudinal Studies.

Data citation and acknowledgement

You should also acknowledge CLS following the guidance from Citing our data page on the CLS website.

This version of the guide was published in February 2026 by the UCL Centre for Longitudinal Studies (CLS).

Centre for Longitudinal Studies

Centre for Longitudinal Studies (CLS)

UCL Social Research Institute

University College London

20 Bedford Way, London WC1H 0AL

www.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 UCL Institute of Education.

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

Contents

About the Millennium Cohort Study 4

1. Introduction to the MCS longitudinal files 5

2. The MCS longitudinal family file (1-7) 6

2.1 Structure and level of the dataset 6

2.2 Usability of the mcs_longitudinal_family_file_1-7 6

3. The MCS longitudinal cohort member file (1-8) 7

3.1 Structure and level of the dataset 7

3.2 Usability of the mcs_longitudinal_cm_file 7

4. Useful information on specific variables 8

4.1 MCS identifiers MCSID, CNUM, PNUM 8

4.2 Cohort member identifier MCSIDCMH 8

4.3 Dual baby families: DUALBABYFAMILY 9

4.4 Sweep of study entry: SENTRY 9

4.5 Access licence: DATA_LICENCE 10

4.6 Cases available for research use: DATA_AVAILABILITY 10

4.7 Case response type: HCASETYPE 10

About the Millennium Cohort Study

The Millennium Cohort Study or MCS (also known as Child of the New Century to participants) is a longitudinal cohort study, following a nationally representative group of over 19,000 people born in the UK between September 2000 and January 2002.

Over the years, the study has built a uniquely rich record of cohort members' lives - from their early health and development, through childhood, into adolescence and now into adulthood. This long-term evidence allows researchers to understand how experiences at different stages of life shape later outcomes.

As a multidisciplinary study, MCS is used by researchers working in a wide range of fields. Findings from MCS have influenced policy at the highest level, and today the study remains a vital source of evidence on the major issues affecting young people's lives.

1. Introduction to the MCS longitudinal files

This guide provides information about the longitudinal files, which are important datasets from the Millennium Cohort Study (MCS).

The MCS longitudinal family file was created at family level to follow MCS families across sweeps 1 to 7 and is widely used by researchers. Until 2025 the dataset name was mcs_longitudinal_family_file, however this has now been renamed with the more descriptive title *mcs_longitudinal_family_file_1-7*. This longitudinal family level file has been retained for researchers who may be using data from older sweeps only and/or requiring family (MCSID) level outcomes/weights.

A new longitudinal cohort member file, named *mcs_longitudinal_cm_file*, has been constructed at individual cohort member level following the MCS data collection at age 23 (sweep 8). This dataset includes cohort member outcomes and survey weights from MCS8.

Both longitudinal datasets are available via the UK Data Service as safeguarded data (End User Licence).

2. The MCS longitudinal family file (1-7)

2.1 Structure and level of the dataset

The longitudinal family file that covers MCS sweeps 1 to 7, originally named as mcs_longitudinal_family_file and now renamed as *mcs_longitudinal_family_file_1-7*, contains one row for each original MCS family unit that participated in the Millennium Cohort Study at any sweep (excluding MCS8) since entry to the study at either Sweep 1 or 2.

Due to sample attrition some families may not have participated in certain sweeps, and datasets relating to a specific sweep may contain productive or partially productive cases.

2.2 Usability of the mcs_longitudinal_family_file_1-7

The mcs_longitudinal_family_file_1-7 can be used for research that focuses on a single sweep between MCS1 and MCS7, or for a project that requires more than one sweep. The file contains information about the outcomes of each family (MCSID) in each sweep. This means that it can be used to follow families longitudinally.

It also includes the family-level weights that have been calculated for sweeps 1 to 7 and the user can use them either for research using one sweep or when using more than one sweep.

Specific information on how the weight variables have been constructed for each sweep is provided in the User Guide for the respective sweep.

3. The MCS longitudinal cohort member file (1-8)

3.1 Structure and level of the dataset

In the MCS data collection at sweep 8 (age 23), rather than following the original MCS family units, cohort members from twin and triplet families are treated as having their own unique households, given that many will no longer live together. To account for this, a new longitudinal dataset named *mcs_longitudinal_cm_file* has been constructed with outcomes and weights produced at cohort member level instead of at family level. Family level outcomes and weights from MCS1-7 have been merged onto this cohort member level data and duplicated/triplicated in cases of twins/triplets.

Therefore, the mcs_longitudinal_cm_file contains one row for each individual cohort member who has (or their family has) participated in the Millennium Cohort Study at any sweep since entry to the study at either Sweep 1 or 2. Due to sample attrition some cohort members may not have participated in certain sweeps, and datasets relating to a specific sweep may contain productive or partially productive cases.

3.2 Usability of the mcs_longitudinal_cm_file

The mcs_longitudinal_cm_file can be used for research using data from a single sweep between MCS1 and MCS8, or for a project that requires data from more than one sweep. The file contains information about the outcomes and weights for each cohort member (MCSIDCMH) at MCS8 and for their original MCS family/household (MCSID) at each earlier sweep (1-7). Therefore, this dataset can be used to follow longitudinally the cohort members' family outcomes from MCS1-7, as well as individual outcomes at cohort member level from MCS8 onwards.

Specific information on how the weight variables have been constructed for each sweep is provided in the User Guide for the respective sweep.

4. Useful information on specific variables

4.1 MCS identifiers MCSID, CNUM, PNUM

MCSID is the anonymised identifier for each original MCS family/household. A family holds the same MCSID across sweeps.

In MCS datasets, individuals within a family can be identified by using the cohort member number CNUM (some families have more than one, e.g. in the case of twins or triplets), and the person number PNUM, which is used to identify any individual other than the actual cohort members.

Individuals have the same PNUM across sweeps. However, from MCS8 the PNUM of new household members is hierarchical within MCSIDCMH due to twin/triplet cohort member "households" diverging, as explained in the next section (4.2). CNUM follows each cohort member across all sweeps.

The original family identifier MCSID is present in both the longitudinal family file (mcs_longitudinal_family_file_1-7) and the longitudinal cohort member file (mcs_longitudinal_cm_file), and new identifier MCSIDCMH (described below) is included in the latter along with CNUM.

4.2 Cohort member identifier MCSIDCMH

In MCS8, the anonymised identifier for each cohort member is MCSIDCMH, which is a combination of the original family identifier MCSID and the cohort member number CNUM (both of which are also included in the cohort-member longitudinal dataset).

The MCSIDCMH identifies a cohort member's unique household from MCS8 onwards. This change from MCSID is due to twin/triplet cohort members now living in separate household units from their siblings, or being interviewed on different dates when household composition may differ for each sibling even if living in the same physical house.

Individuals within a cohort member's family/household (MCSIDCMH) can be followed by using PNUM (Person number). At MCS8, cohort member twin/triplet siblings are identified in the cohort member's household by PNUM with value CNUM*100, so twin1 has PNUM=100, twin/triplet 2 has PNUM=200, triplet 3 has PNUM 300.

Due to divergence of twin/triplet households at MCS8, the original MCSID/PNUM combination may no longer be unique **from MCS8 onwards** and so **MCSIDCMH/PNUM** must be used instead for merging person-level data**.** However, person level data from earlier sweeps (MCS1-7) will not contain duplicates on MCSID/PNUM and so this can still be used for merges from older datasets into MCS8 data.

Merges involving cohort member level data from earlier sweeps can be done using the two keys MCSID/CNUM or the new concatenation MCSIDCMH, which can easily be constructed in datasets from older sweeps.

4.3 Dual baby families: DUALBABYFAMILY

The DUALBABYFAMILY variable indicates families where a mother gave birth to two Cohort Members within the sample selection period. Both of the children are Cohort Members, however they were not part of a multiple birth (like twins and triplets) and therefore do not share the same MCSID.

4.4 Sweep of study entry: SENTRY

Not all families were interviewed in MCS sweep 1. Some “new families” were interviewed for the first time in sweep 2 and were asked similar questions about the birth of the cohort child as were asked to the families that entered the survey in sweep 1. Therefore, if the research project requires information about birth and the first days of life of the Cohort Member, this can be obtained for families that entered the survey in sweep 2 from the MCS2 data.

Variable SENTRY indicates the Sweep in which an MCS family entered the study. Its value labels are

- -1 = Not applicable
- 1 = Sweep 1
- 2 = Sweep 2

4.5 Access licence: DATA_LICENCE

The DATA_LICENCE variable indicates which cases in the sample are available under safeguarded data through the UKDS (End User Licence, EUL). A few cases are available as controlled data under Secure Access (SA). For example, the survey data of families that contain triplets are available under Secure Access due to its potential disclosivity.

More information on licences is available in the CLS Data Classification Policy on the CLS website.

The DATA_LICENCE value labels are

- 0 = Under EUL
- 1 = Under SA

4.6 Cases available for research use: DATA_AVAILABILITY

In addition to the DATA_LICENCE variable, the DATA_AVAILABILITY variable indicates cases that are currently available for research use, irrespective of licence type. Some cases that have been removed temporarily or permanently from sharing are flagged with DATA_AVAILABILITY = 0. This status may change for individual cases at future data releases and will be updated accordingly.

The DATA_AVAILABILITY value labels are

- 0 = No
- 1 = Yes

4.7 Case response type: HCASETYPE

For MCS8, the variable HCASETYPE shows which MCS8 cohort members (cases) responded to the Soft Launch or Mainstage interviews, which responded to the shorter subset of questions in the Mopup interview, and which responded to a brief interview by proxy. Unproductive and non-issued cases are also flagged by this variable.

The value labels are:

- 1 = Soft Launch or Mainstage
- 2 = Mopup
- 3 = Proxy (data unavailable)
- 4 = Not Productive
- 5 = Not issued