

Millennium Cohort Study: Core dataset

User Guide (Version 1)

May 2026

Contact

Data queries: help@ukdataservice.ac.uk

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

Authors

Nicolás Libuy, Dominic Kelly, Emla Fitzsimons

How to cite this guide

Libuy N., Kelly, D. P., Fitzsimons E. (2025) *Millennium Cohort Study: Core dataset User Guide (version 1)*. UCL Centre for Longitudinal Studies.

You should cite the data and acknowledge CLS following [guidance from the CLS website](#).

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 contact the Centre for Longitudinal Studies at clsdata@ucl.ac.uk.

Acknowledgements

CLS would like to thank the MCS User Panel for their expert our contributions to the development of this dataset: Megan Arnot (Dept. for Education), Gergo Baranyi (CLS, UCL), Uta Bolt (University of Bristol), Marta Francesconi (UCL), Larissa Pople (CLS, UCL), Georgia Turner (Cambridge University). Thanks to Tom Metherell (CLS, UCL) and Ollie Cassagneau-Francis (UCL) for valuable feedback on the R code.

Contents

About the Millennium Cohort Study	5
1. Introduction to the MCS Core dataset	6
1.1 Overview of the CLS longitudinal datasets.....	6
1.2. Overview of the MCS Core dataset.....	6
2. Structure of the MCS Core dataset	9
2.1 Overview of MCS Data Structures and Family and Cohort Member Identifiers.	9
2.2 Description of the MCS Core dataset.....	11
2.3 Derived variables	16
3. Usability	18
3.1 Licensing and data access.....	18
3.2 Usability of the MCS Core dataset	18
3.3 Instructions to create the dataset using the accompanying codes	19
4. Content of the MCS Core dataset	22

About the Millennium Cohort Study

The Millennium Cohort Study (MCS) is a longitudinal birth cohort study, following a nationally representative sample of approximately 19,000 people born in the UK at the turn of the century.

The study has captured rich information about the different aspects of cohort members' lives, from birth and childhood to adolescence and into early adulthood, as part of this longitudinal study.

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.

MCS has thus far consisted of eight sweeps:

- **MCS 1** – age 9 months (fieldwork began in 2001)
- **MCS 2** – age 3 years (fieldwork began in 2004)
- **MCS 3** – age 5 years (fieldwork began in 2006)
- **MCS 4** – age 7 years (fieldwork began in 2008)
- **MCS 5** – age 11 years (fieldwork began in 2012)
- **MCS 6** – age 14 years (fieldwork began in 2015)
- **MCS 7** – age 17 years (fieldwork began in 2018)
- **MCS 8** – age 23 years (fieldwork began in 2023).

1. Introduction to the MCS Core dataset

1.1 Overview of the CLS longitudinal datasets

CLS is committed to making our rich longitudinal data easier to use. As part of this, for all our established adult cohorts, we have created core longitudinal datasets and accompanying code. The aim is twofold: first, to provide a research-ready dataset with all major / core variables across several sweeps; second, to provide the code so that researchers can straightforwardly and flexibly add more variables to this dataset / amend it as they see fit. We envisage that these datasets will be useful for both new and established users, as well as analysis datasets for teaching purposes.

These datasets are optimised for within-cohort comparisons; users who wish to conduct comparative analyses across multiple cohorts should ensure that items used are harmonised. We plan to update these datasets as new sweeps of data are collected.

This is the first version of this file. We welcome user feedback on this file including details of any improvements or errors, please [see our contact details on page 2](#).

1.2. Overview of the MCS Core dataset

This user guide provides information about the core longitudinal file for the Millennium Cohort Study MCS core dataset.

The guide aims at helping researchers to understand the data structure and contents of this dataset. It focuses on explaining the data and providing important considerations for researchers interested in using this data resource. Its objectives are:

1. To streamline **access to a core set of variables** spanning key domains across life, including socio-demographic information for cohort members and their families, and indicators of physical and mental health. Those domains include:
 - Demographics, e.g. sex, gender, ethnicity

- Physical and mental health, e.g. self-rated health and wellbeing
- Anthropometrics, e.g. height and weight
- Socioeconomic circumstances, e.g. social class
- Educational attainment
- Behaviours, e.g. physical activity, alcohol intake.

2. To provide a set of **code** allowing researchers to readily **modify the derived variables** in ways more suited to their own research needs, and / or to **adapt in order to create new derived variables** to enhance the core set.

The MCS Core dataset is generated from MCS datasets deposited at the UK Data Service ([UKDS; SN: 2000031: Millennium Cohort Study \[data series\]](#)). The creation process involves applying a set of codes to these data files, in order to harmonise and derive variables in longitudinal format. A single row in the dataset corresponds to each participant-sweep.

The codes to replicate and extend this work are [accessible through the MCS GitHub repository](#), providing researchers the flexibility to utilise, modify, and add to them based on their specific needs. R code is already available; Stata code will be made available shortly. Alongside the code, users must request access to the following MCS datasets deposited at UKDS:

Table 1: MCS Datasets

UKDS SN	Study description
8172	Millennium Cohort Study: Sweeps 1-8, 2001-2025: Longitudinal Files
4683	Millennium Cohort Study: Age 9 months, Sweep 1, 2001
5350	Millennium Cohort Study: Age 3, Sweep 2, 2004
5795	Millennium Cohort Study: Age 5, Sweep 3, 2006
6411	Millennium Cohort Study: Age 7, Sweep 4, 2008
7464	Millennium Cohort Study: Age 11, Sweep 5, 2012
8156	Millennium Cohort Study: Age 14, Sweep 6, 2015
8682	Millennium Cohort Study: Age 17, Sweep 7, 2018

UKDS SN	Study description
9509	Millennium Cohort Study: Sweep 8, Age 23, 2023-2025
8550	Harmonised Height, Weight and BMI in Five Longitudinal Cohort Studies: Millennium Cohort Study

2. Structure of the MCS Core dataset

In this section, we outline the MCS data structures, provide a description of the MCS Core file, including its purpose and associated codes.

2.1 Overview of MCS Data Structures and Family and Cohort Member Identifiers

MCS datasets available from UKDS ([Table 1](#)) can have one of four structures, or levels, depending on the information they incorporate:

- **Family level** (one row per family): This level encompasses information about the family as a singular entity. Families are uniquely identified by the **mcsid** variable, with the same **mcsid** assigned to families across sweeps.
- **Cohort Member level** (one row per cohort member): This level incorporates information pertaining to questions answered by individual cohort members. A distinctive, permanent cross-sweep identifier for cohort members can be established by combining **mcsid** and **cnum**. The identifier **cnum** represents the Cohort Member number within a family. Specifically, ascending **cnum** values starting from 1 indicate the first cohort member within a family. While most cohort members have **cnum**=1, families with twins and triplets designate **cnum**=2 and **cnum**=3 for the second and third cohort members, respectively.
- **Parent level** (one row for Main and one row for Partner respondent): This level includes information derived from questions addressed to parents about the family or themselves.
- **Parent - Cohort Member level** (one row for each parent-child combination within the household, e.g. one row if a single parent family of a singleton, four rows if a two-parent family with twins etc.): At this level, information is provided about each cohort member from the perspective of the parent(s).

From sweeps 1 to 6, one of the parents of the cohort members served as the primary respondent for information about the cohort member. From sweeps 7 onwards, the cohort members themselves served as the primary source of information about the cohort member. Thus, the datasets which are being used to derive information about the cohort member notably change between sweeps 6 and 7.

Besides the unique identifiers of cohort members across sweeps, i.e., **mcsid** and **cnum**, other key identifiers in MCS can be created by using **mcsid** in combination with **pnum**, **elig** or **resp**:

- **pnum**: The combination with **mcsid** creates a unique, permanent cross-sweep identifier for other individuals residing in the household, distinct from the Cohort Member(s).
- **elig**: This identifier provides information about an individual's eligibility for the role of Main (**elig**=1), Partner (**elig**=2) or Proxy (**elig**=3) respondent. **elig**=4 indicates the individual is not eligible for this role. For instance, the mother of the Cohort Member may have **pnum**=2, and the father **pnum**=3, but only the mother might be eligible for the survey interview (**elig**=1), whereas the father is not eligible (**elig**=4). It is important to note that the Main respondent may be the same person across sweeps or may change, e.g., it could be mother at some sweeps but father (or partner of mother) at others.
- **resp**: This variable indicates whether the person selected at **elig** has participated in the interview or not (e.g., refusal). **resp** serves as the outcome of the **elig** process.

From the eighth wave of the MCS onwards, an additional identifier for each individual, **mcsidcmh**, was included in the datasets. This is to recognise that the cohort members are now young adults and in situations where there are multiple cohort members from the same family, these cohort members may now have their own separate families (even if they still live in the same household as their siblings). Given this, **mcsidcmh** is a new ID for each unique individual which represents a unique “cohort member household” ID going forward.

In creating a core longitudinal dataset for MCS, we worked with the four distinct data structures of MCS to derive variables that aggregate pertinent information at the Cohort Member-Sweep level (`mcsid cnum sweep`).

Further Information and additional resources:

A comprehensive description of MCS data structures is [available on the online CLS Data Handling guide on GitHub](#) and also as a [PDF resource from the UK Data Service website](#).

For a more comprehensive understanding of the MCS datasets, additional resources such as webinars, codes, and tutorials are [available on the CLS website](#) and [on the MCS GitHub repository](#).

2.2 Description of the MCS Core dataset

The MCS Core dataset has a single row for each Cohort Member-Sweep combination (`mcsid cnum sweep`). It mirrors the number of cohort members found in the `mcs_longitudinal_cm_file`¹, allocating one row for every cohort member in the MCS.

The MCS Core dataset is created from a set of codes. The `1_mcs_core` code uses as input the codes `2_mcs_long`, `3_mcs_linkage` and a set of *Variable-Specific Codes* to create the core dataset. This is described below:

a) Main code file

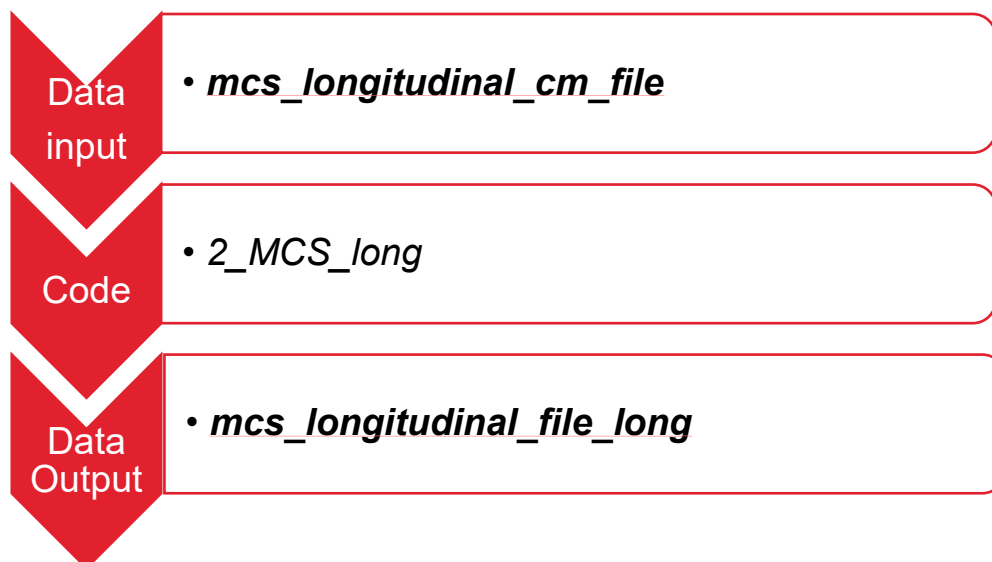
The `1_mcs_core` code file is the main file for creating a dataset, and the only one that needs to be opened and amended by the user. At the start of the file, the user will see spaces where the paths for the core directory (`core_dir`) must be defined and then additional spaces where the paths for each of the 10 folders downloaded from UKDS must be specified. The remainder of the code is already set for the user and

¹ The `mcs_longitudinal_cm_file` contains one row per family. A detailed description of the `mcs_longitudinal_cm_file` dataset can be found in *Agalioti-Sgompos Vilma & Johnson Jon (2020) Millennium Cohort Study Longitudinal Family File. Ed2. London: UCL Centre for Longitudinal Studies.*

draws upon each of the smaller code files for each subset of variables. The user only needs to run (or source) this file for the MCS Core file to be created; no other file requires amendments by the user.

b) *2_mcs_long* code

This code uses as input the *mcs_longitudinal_cm_file* to create a dataset – named *mcs_longitudinal_file_long* – which has a longitudinal structure (one row per cohort member-sweep). This dataset retains key variables with information on the MCS research design included in the *mcs_longitudinal_cm_file*. This is illustrated below:

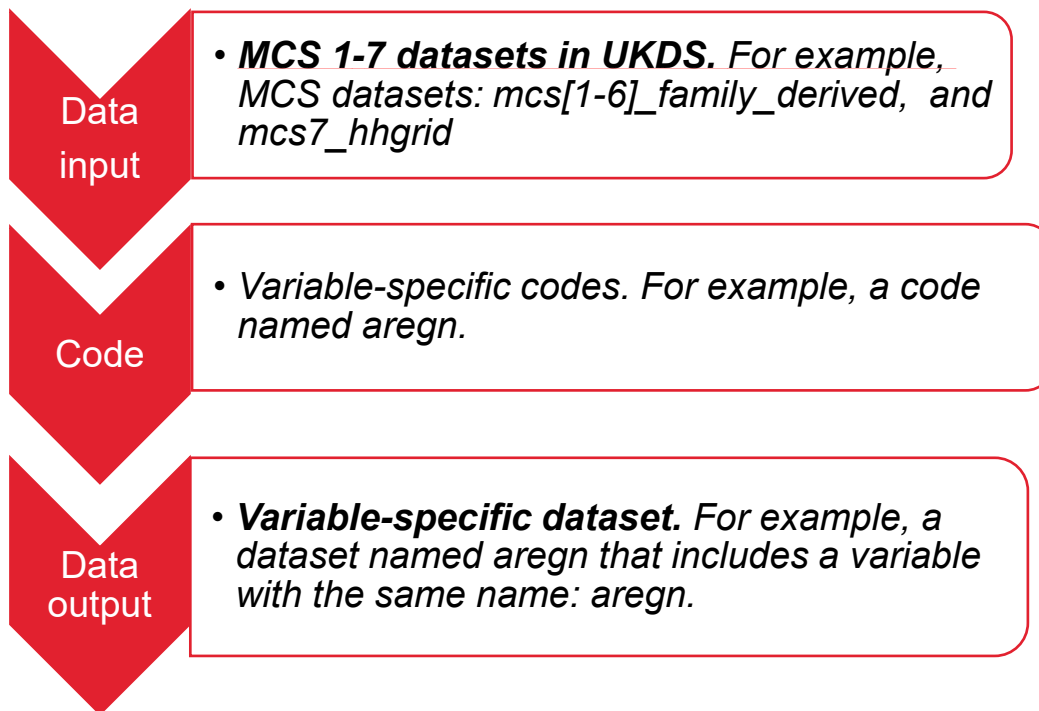


c) Variable-specific codes:

There is one code file for each derived variable or set of derived variables. A specific code generates a dataset with one or multiple derived variable(s), using as input MCS datasets across different sweeps. It generates a dataset per derived variable (or set of derived variables) in a longitudinal structure with the corresponding MCS IDs and derived variables.

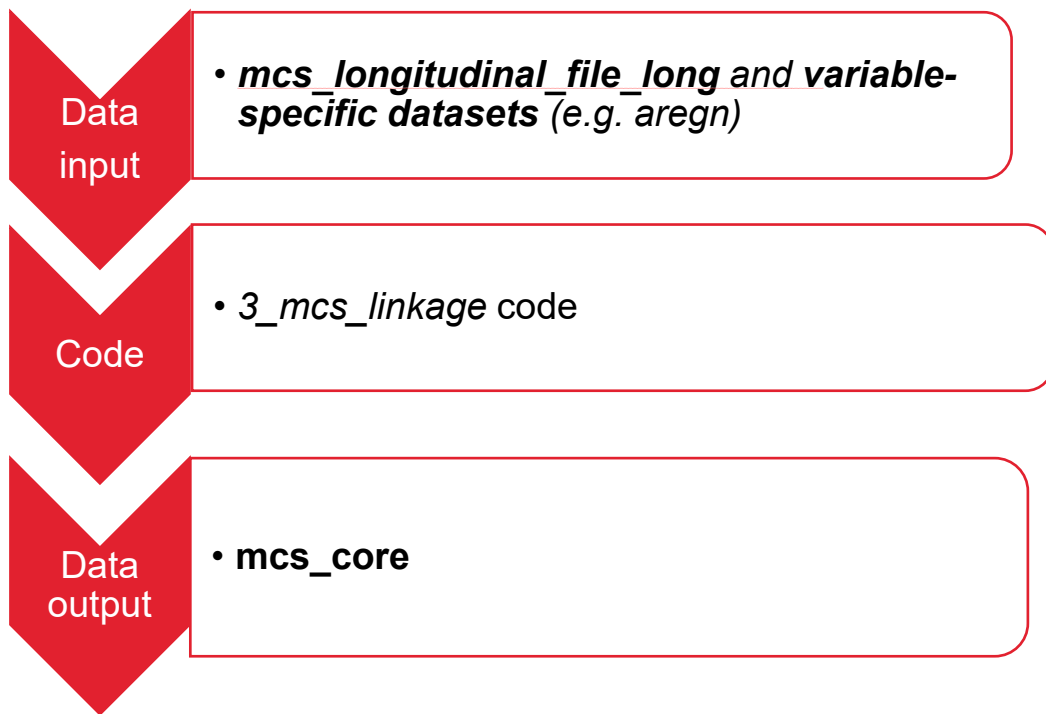
- For example, the code "aregn" uses the MCS datasets *mcs[1-6] family_derived*, and *mcs7_hhgrid* to create the variable with the same name "aregn". This variable is created using the following original variables, which are across

different MCS sweeps datasets: AAREGN00, BAREGN00, CAREGN00, DAREGN00, EAREGN00, FAREGN00 and GAREGN00. This is illustrated below:



d) **3_mcs_linkage code:**

This creates the MCS Core dataset by merging the derived variables created in step c) to the dataset **mcs_longitudinal_cm_file_long** created in step b).



The longitudinal structure of the data follows the structure of **mcs_longitudinal_cm_file**. This means that in the final dataset, each cohort member has 8 rows, even if they did not participate in the specific sweep.

Table 2: Example of longitudinal structure of MCS Core

mcsid (Family number)	cnum (Cohort Member number)	sweep (MCS sweep)	aoutc (Survey Outcome Code)	Derived Variables (...)
M40708V	1	Sweep 1	Productive	Value
M40708V	1	Sweep 2	Productive	Missing
M40708V	1	Sweep 3	Productive	Value
M40708V	1	Sweep 4	Untraced	.
M40708V	1	Sweep 5	Productive	Value
M40708V	1	Sweep 6	No contact	.
M40708V	1	Sweep 7	Refusal	.
M40708V	1	Sweep 8	Productive	Value

All MCS variables with essential information about the MCS Survey Design in **mcs_longitudinal_cm_file** were included in **MCS Core** (Table 3). Note the values of variables in Table 3 are fixed across sweeps (time-invariant).

Table 3: MCS Survey Design variables (time-invariant) in MCS Core

Variable name	Variable label
mcsidcmh	Cohort member household ID
mcsid	MCS Research ID - Anonymised Family/Household Identifier
cnum	Cohort Member number within an MCS family
nocmhh	Number of cohort children in household at entry to survey
sentry	Sweep in which family entered study (1,2)
country	Country as at MCS sampling
ptty	Point type (advantaged, disadvantaged, ethnic)
pttype2	Stratum within Country (e.g. England – advantaged, England – disadvantaged, England – ethnic etc)
sptn	Fieldwork point number incorporating superwards (updated)
nh2	Population Correction Factor
weight1	MCS sample design weights by stratum and country - to use in single country analyses
weight2	MCS sample design weights by stratum - to use in whole UK analyses
weightgb	MCS sample design weights by stratum - to use in GB analyses

We also included from the **mcs_longitudinal_cm_file** variables encompassing each survey sweep's fieldwork Non-Response and MCS sample design weights fieldwork ([Table 4](#)). These variables can change across sweeps in **mcs_longitudinal_cm_file** ([see, for instance, aoutc in Table 2](#)) as fieldwork non-response codes and overall non-response weights are sweep-specific. Comprehensive information on utilizing these variables is available in the User Guide of the respective sweep and in:

- Millennium Cohort Study User Guide (Surveys 1 – 5), 9th Edition 2020 London: UCL Centre for Longitudinal Studies.
- Agalioti-Sgompou & Johnson. (2020) Millennium Cohort Study Data Handling Guide with syntax in R, Stata and SPSS. London: UCL Centre for Longitudinal Studies.

Table 4: Sweep fieldwork non-response (aoutc) and Sweep-specific overall weights (time-varying) in MCS Core

Variable name	Variable label
aoutc	Survey Outcome Code (e.g. productive, ineligible, untraced, refusal, non-contact etc)
issued	Family Issued at MCS survey (e.g. Yes, No)
nrespwt	Non-Response Weight
ovwt1	Overall Weight (including Non-Response adjustment) single country analysis
ovwt2	Overall Weight (including Non-Response adjustment) whole UK analyses
ovwtgb	Overall Weight (including Non-Response adjustment) for use in GB analysis

2.3 Derived variables

The MCS datasets deposited at UKDS for each sweep already include MCS derived variables. We also created some new derived variables. Each derived variable, whether newly created or previously deposited at UKDS, was generated using customised codes. The codes take the relevant MCS datasets listed in Table 1 as input. Specific details and assumptions used to generate the variables available in **MCS Core** can be found in the specific code files.

Below, we provide three examples of derived variables to explain the rationale behind the creation of such variables in the **MCS Core**:

- Cohort member's birth weight in kilograms: The code combines information from two datasets, **mcs1_cm_derived** and **mcs2_parent_cm_interview**, to generate the cohort member's birth weight in kilograms. As this variable

remains constant across sweeps, its information is duplicated in each cohort member-sweep row within the **MCS Core**.

- Number of people in household at each sweep including cohort member(s) (dnumh): In the first seven MCS sweeps, CLS has previously created and deposited this derived variable in the datasets mcs[1-7] family_derived. The variable name changes across MCS sweeps, adhering to the naming convention of MCS variables (MCS1: ADNUMH00, MCS2: BDNUMH00...MCS7: GDNUMH00). A code generates the variable dnumh in longitudinal structure (one row per MCS family-sweep) from various datasets and embeds it within **MCS Core**.
- Life satisfaction of the Main respondent (elig=1) at each sweep: The Main respondent (elig=1) and Partner respondent (elig=2) were surveyed from MCS1 through 5 regarding their satisfaction with how their life has unfolded so far. The variables are included in the mcs[1-5] parent_interview datasets, and their names change across MCS sweeps following the naming convention of MCS variables (MCS1: APWALI00, MCS2: BPWALI00... MCS5: EPWALI00). A new derived variable, called wali, is generated and linked to **MCS Core** through a code that retains responses solely from the Main respondent (elig=1) in each survey. Since there is only one Main respondent per family and sweep, wali is linked to **MCS Core** using the mcsid sweep.

3. Usability

3.1 Licensing and data access

The MCS Core dataset is available from the UK Data Service (UKDS).

All users of the data need to be registered with the UKDS. Details of how to do this are [available on the UK Data Service website](#).

The MCS Core dataset is available as safeguarded data at the UKDS. These datasets can be downloaded once the End User Licence access conditions have been accepted by the user.

3.2 Usability of the MCS Core dataset

The **MCS Core** dataset is intended as a longitudinal resource for obtaining essential core information about cohort members and their families across the whole of childhood, adolescence and early adulthood (sweeps 1-8). Researchers can efficiently access a set of core time-invariant and time-varying derived variables without needing to navigate the entire process of combining and generating various MCS-specific datasets. However, considering the complexities inherent in the MCS survey design, data structures, and questionnaires, we recommend researchers utilize **MCS Core** in conjunction with the detailed documentation provided by CLS for each specific MCS sweep.

Researchers can use any dataset [listed in Table 1](#), combined with adapting the code we have provided, to derive their own variables and subsequently link this information to **MCS Core** using the identifiers of cohort members (mcsid cnum sweep) for cohort member-level information, or their families (mcsid sweep) for family-level information.

The **MCS Core** dataset is provided in a longitudinal structure, with one row per cohort member-sweep. However, researchers have the flexibility to readily transform

the data into wide format. For example, in Stata, to transform the dataset from long to wide, researchers can use the *reshape* command as follows:

```
reshape wide [variable names], i(mcsid cnum) j(sweep)
```

Where [variable names] is a list of time-varying variables in the dataset.

3.3 Instructions to create the dataset using the accompanying codes

The **MCS Core** dataset comes with a set of codes that outline the assumptions and decisions made by CLS during the variable generation process. Researchers can leverage these codes not only to derive their own variables but also to adapt the assumptions and decisions made by CLS to align with their specific research questions or analyses. This flexibility empowers researchers to tailor the dataset to suit the unique requirements of their studies.

The step-by-step instructions to create the Core dataset are in the *1_mcs_core* code.

To replicate the dataset, researchers should follow the next steps:

Step 1) [Go to the CLS GitHub repository](#) and download and open the file 'mcs_core_R.zip'. This file contains a folder named 'mcs_core' with the following sub-folder structure:

- temp_data
 - o cmderived
 - o familyderived
 - o parentderived
- codes
 - o cmderived
 - o familyderived
 - o parentderived

The folder 'temp_data' and its sub-folders are initially empty, and it is where datasets will be stored.

The folder 'codes' includes the codes *2_mcs_long*, *3_mcs_linkage* and a set of *Variable-Specific* codes. The set of *Variable-Specific* codes are in sub-folders 'cmderived', 'familyderived' and 'parentderived'.

The 'cmderived' folder includes codes to create **Variable-Specific datasets** at the Sweep - Cohort Member level (one row per Cohort Member - Sweep).

The 'familyderived' folder includes codes to create **Variable-Specific datasets** at the sweep-family level (one row per Family - Sweep).

The 'parentderived' folder also includes codes to create **Variable-Specific datasets** at the Sweep-Family level (one row per Family - Sweep). However, it is important to notice that these variables are derived from MCS datasets with information at the Parent level² (one row for Main and one row for Partner respondent) or at the Parent - Cohort Member level³ (one row for each parent-child combination within the household, e.g. one row if a single parent family of a singleton, four rows if a two-parent family with twins etc.). Therefore, to aggregate information at the Family Level (one row for each family) for each sweep, a decision was made to retain responses from only one parent in the household. In general, the criteria used was to include information provided by the Main interviewee (elig=1), but for some cases we created separate variables, one with information from the Main interviewee (elig=1) and another with information from the Partner (elig=2). For example, the variable dnvq includes the highest National Vocational Qualification (NVQ) level for the Main Interviewee and the dnvq_p includes the same information for the Partner Interviewee. Details about these decisions are documented in each specific code.

Step 2) Download from the UKDS the MCS datasets listed in [Table 1](#).

Step 3) Open the code *1_mcs_core* and manually change the lines that contain:

² This level includes information derived from questions addressed to parents about the family or themselves.

³ At this level, information is provided about each cohort member from the perspective of the parent(s).

- a. the path which points to the folder 'mcs_core' in your computer, and
- b. the paths which point to the folders of each MCS datasets in Table 1 in your computer.

Step 4) Run the code `1_mcs_core`, which will create the core dataset and save it in the folder 'temp_data'. If following the R code, this dataset will be in the form of a .Rds file.

4. Content of the MCS Core dataset

[Table 5 on the following page](#) provides a summary of variables included in the **MCS Core** dataset at the time of this publication. Additional information about each variable is also available in the Excel file `mcs_core_variables`. Note that variables in Table 5 are those included at the time of publication of this guide; however, as more variables are added to the dataset by CLS, updated information will be maintained in the Excel file. Researcher should review the last version of the Excel file for updated information of the content of the **MCS Core** dataset. The dataset includes the latest wave of the MCS study collected when participants were approximately age 23. Data from this wave is already included in some of the variable contained in the core files – future updates will include more data from this wave.

The Excel file includes the variable name and its label, the title of the code that creates the variable, and the primary topic or area that the variable addresses (demographic, physical health, mental health, survey structure, etc.). Additionally, the file contains information about the variables in the datasets in Table 1 that were utilized to derive each variable in **MCS Core** and specifies whether the derived variable remains constant across MCS sweeps (i.e., time-invariant variables) or its values change over time (i.e., time-varying variables). The Excel file also specifies which MCS sweeps, and the number of sweeps used to create the derived variable (e.g., 1, 2, 3..., 8 sweeps). The original variable name in each MCS dataset used for generating derived variables is also provided in the Excel file.

Table 5: Variables in MCS Core

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
mcsidcmh	MCS8 CM Household ID, concatenation of mcsid and cnum	mcs_longitudinal_cm_file	TI	N/A
mcsid	MCS Research ID - Anonymised Family/Household Identifier	mcs_longitudinal_cm_file	TI	N/A
cnum	Cohort Member number within an MCS family	mcs_longitudinal_cm_file	TI	N/A
sweep	MCS sweep	N/A	TV	8
nocmhh	Number of cohort children in household at entry to survey	mcs_longitudinal_cm_file	TI	N/A
sentry	Sweep in which Family entered study	mcs_longitudinal_cm_file	TI	N/A
country	Country as at MCS sampling	mcs_longitudinal_cm_file	TI	N/A
ptty	Point type	mcs_longitudinal_cm_file	TI	N/A
pttype2	Stratum within Country	mcs_longitudinal_cm_file	TI	N/A
sptn	Fieldwork point number incorporating superwards (updated)	mcs_longitudinal_cm_file	TI	N/A
nh2	Population Correction Factor	mcs_longitudinal_cm_file	TI	N/A
weight1	MCS Weight for use on single country analyses	mcs_longitudinal_cm_file	TI	N/A
weight2	MCS Weight to use on whole UK analyses	mcs_longitudinal_cm_file	TI	N/A

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
weightgb	Overall weight for use on GB analysis	mcs_longitudinal_cm_file	TI	N/A
issued	MCS survey issued	mcs_longitudinal_cm_file	TV	8
aoutc	Survey Outcome Code	mcs_longitudinal_cm_file	TV	8
ovwt1	Overall Weight (inc. NR adjustment) single country analysis	mcs_longitudinal_cm_file	TV	8
ovwt2	Overall Weight (inc. NR adjustment) whole UK analyses	mcs_longitudinal_cm_file	TV	8
nrespwt	Non-Response Weight	mcs_longitudinal_cm_file	TV	5
ovwtgb	Overall weight for use on GB analysis	mcs_longitudinal_cm_file	TV	4
actry	ADMIN Country at interview (E,W,S,NI)	mcs[1-6]_family_derived, mcs7_hhgrid	TV	7
aregn	Interview Government Office Region	mcs[1-6]_family_derived, mcs7_hhgrid	TV	7
drspo	DV Parent Interview response summary	mcs[1-6]_family_derived	TV	6
dhtyp	DV Parents/Carers in Household	mcs[1-7]_family_derived	TV	7
dhtys	DV Summary of Parents/Carers in Household	mcs[1-7]_family_derived	TV	7
drelp	DV Relationship between Parents/Carers in Household	mcs[1-7]_family_derived	TV	7
dminh	DV Natural mother in HH	mcs[1-7]_family_derived	TV	7
dfinh	DV Natural father in HH	mcs[2-6]_family_derived	TV	5

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time- invariant	Number of Sweeps
doths	DV Number of siblings of CM in HH	mcs[1-7]_family_derived	TV	7
dnocm	DV Number of CM in household	mcs[1-7]_family_derived	TV	7
dtots	DV Number of sibs in hhold plus CMs	mcs[1-7]_family_derived	TV	7
dnsib	DV Natural siblings of CM in household	mcs[1-7]_family_derived	TV	7
dhsib	DV Half siblings of CM in household	mcs[1-7]_family_derived	TV	7
dssib	DV Step siblings of CM in household	mcs[1-7]_family_derived	TV	7
dasib	DV Adoptive siblings of CM in HH	mcs[1-7]_family_derived	TV	7
dfsib	DV Foster siblings of CM in household	mcs[1-7]_family_derived	TV	7
dgpar	DV Grandparents of CM in HH	mcs[1-7]_family_derived	TV	7
dotha	DV Other adult in HH	mcs[1-7]_family_derived	TV	7
dnumh	DV Number of people present in HH excluding CMs	mcs[1-7]_family_derived	TV	7
dtotp	DV Number of people in HH including CMs	mcs[1-7]_family_derived	TV	7
dcwrk	DV Combined labour market status of Main and Partner	mcs[1-6]_family_derived	TV	5
droow	DV Housing Tenure	mcs[1-6]_family_derived	TV	6
dmbmi	DV Natural Mothers BMI at Interview	mcs[1-4]_family_derived	TV	4
dhlan	DV Language Spoken in household	mcs[1-6]_family_derived	TV	6

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
doeds	DV Equivalised family income OECD scores	mcs[1-4]_family_derived	TV	4
doede	DV OECD equivalised income	mcs[1-6]_family_derived	TV	6
doedp	DV OECD Below 60% median indicator	mcs[1-6]_family_derived	TV	6
agey	Age at Interview (years)	mcs[1-7]_hhgrid, mcs5_cm_interview, mcs8_23y_cm_derived	TV	8
sweepage	Age at MCS Sweep (years)	mcs[1-7]_hhgrid, mcs5_cm_interview, mcs8_23y_cm_derived	TV	8
emotion	DV SDQ Emotional Symptoms (parent-reported)	mcs[2-7]_cm_derived	TV	6
conduct	DV SDQ Conduct Problems (parent-reported)	mcs[2-7]_cm_derived	TV	6
hyper	DV SDQ Hyperactivity/Inattention (parent-reported)	mcs[2-7]_cm_derived	TV	6
peer	DV SDQ Peer Problems (parent-reported)	mcs[2-7]_cm_derived	TV	6
prosoc	DV SDQ Prosocial (parent-reported)	mcs[2-7]_cm_derived	TV	6
ebdtot	DV SDQ Total Difficulties (parent-reported)	mcs[2-7]_cm_derived	TV	6
dcsbi	DV Child Social Behaviour Questionnaire (Independence-Self Regulation)	mcs[2-4]_cm_derived	TV	3
dcsbe	DV Child Social Behaviour Questionnaire (Emotional-Dysregulation)	mcs[2-4]_cm_derived	TV	3

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
dcsbc	DV Child Social Behaviour Questionnaire (Cooperation)	mcs4_cm_derived	TV	1
emotion_c	DV SDQ Emotional Symptoms (cohort member-reported)	mcs7_cm_derived	TV	1
conduct_c	DV SDQ Conduct Problems (cohort member-reported)	mcs7_cm_derived	TV	1
hyper_c	DV SDQ Hyperactivity/Inattention (cohort member-reported)	mcs7_cm_derived	TV	1
peer_c	DV SDQ Peer Problems (cohort member-reported)	mcs7_cm_derived	TV	1
prosoc_c	DV SDQ Prosocial (cohort member-reported)	mcs7_cm_derived	TV	1
ebdtot_c	DV SDQ Total (cohort member-reported)	mcs7_cm_derived	TV	1
kessler_c	DV K6 Kessler score (cohort member-reported)	mcs7_cm_derived, mcs8_23y_cm_derived	TV	2
selfha	Hurt yourself on purpose during the last year	mcs[6-7]_cm_interview, mcs8_23y_cm_survey	TV	3
smfq	Short Mood and Feelings Questionnaire total score	mcs6_cm_interview	TV	1
sleepq	Self-assessed sleep quality (past 4 weeks/month)	mcs7_cm_interview, mcs8_23y_cm_survey	TV	2
socmedh	Number of weekday hours spent on social media	mcs7_cm_interview, mcs8_23y_cm_survey	TV	2

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
socmeda	Feeling addicted to social media	mcs7_cm_interview, mcs8_23y_cm_survey	TV	2
socmedc	More connected and happier online than in real life	mcs7_cm_interview, mcs8_23y_cm_survey	TV	2
wghtp	Perception of weight	mcs[6-7]_cm_interview, mcs8_23y_cm_survey	TV	3
cghe	CM General level of health (Parental report)	mcs[3-5]_parent_cm_interview	TV	3
cghe_sr	CM General level of health (CM self-rated)	mcs[6-7]_cm_interview, mcs_23y_cm_survey	TV	3
clsi	CM has longstanding illness (reported by parent)	mcs[2-6]_parent_cm_interview	TV	5
clsi_sr	CM has longstanding illness (reported by CM)	mcs7_cm_interview, mcs_23y_cm_survey	TV	2
clsl	CM has conditions/illnesses limit everyday activities (reported by parent)	mcs[2-6]_parent_cm_interview	TV	5
clsl_sr	CM has conditions/illnesses limit everyday activities (reported by CM)	mcs7_cm_interview, mcs_23y_cm_survey	TV	2
xage_c	Actual age (CLOSER)	mcs_closer_wp1	TV	1
wt_c	Harmonised weight (CLOSER)	mcs_closer_wp1	TV	1
ht_c	Harmonised height (CLOSER)	mcs_closer_wp1	TV	1
bmi_c	Body mass index (kg/m2) (CLOSER)	mcs_closer_wp1	TV	1

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
bmi_cls	DV Cohort Member BMI according to present weight (calculated by CLS)	mcs[2-7]_cm_interview	TV	6
weight	Weight, measured or self-report	mcs[2-7]_cm_interview	TV	6
height	Height, measured or self-report	mcs[2-7]_cm_interview	TV	6
drosenberg	DV Rosenberg Self-esteem Score (0-15)	mcs[5-7]_cm_interview	TV	3
dwemwbs	DV WEMWBS: Sum of raw mental wellbeing scores transformed to metric scale	mcs7_cm_interview	TV	1
smil	Developmental milestones: smiles	mcs1_parent_cm_interview	TV	1
situ	Developmental milestones: sits up	mcs1_parent_cm_interview	TV	1
stan	Developmental milestones: stands up holding on	mcs1_parent_cm_interview	TV	1
hand	Developmental milestones: hands together	mcs1_parent_cm_interview	TV	1
grab	Developmental milestones: grabs objects	mcs1_parent_cm_interview	TV	1
pick	Developmental milestones: holds small objects	mcs1_parent_cm_interview	TV	1
ptoy	Developmental milestones: passes a toy	mcs1_parent_cm_interview	TV	1
walk	Developmental milestones: take child to park or playground	mcs1_parent_cm_interview	TV	1
give	Developmental milestones: gives toy	mcs1_parent_cm_interview	TV	1
wave	Developmental milestones: waves bye-bye	mcs1_parent_cm_interview	TV	1

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
arms	Developmental milestones: extend arms for being picked up	mcs1_parent_cm_interview	TV	1
nods	Developmental milestones: nods for yes	mcs1_parent_cm_interview	TV	1
dcosc	DV Bracken: Colours Raw Score	mcs2_cm_cognitive_assessment	TV	1
dcmas	DV Bracken: Colours % mastery	mcs2_cm_cognitive_assessment	TV	1
dlesc	DV Bracken: Letters Raw Score	mcs2_cm_cognitive_assessment	TV	1
dlmas	DV Bracken: Letters % mastery	mcs2_cm_cognitive_assessment	TV	1
dnosc	DV Bracken: Numbers Raw Score	mcs2_cm_cognitive_assessment	TV	1
dnmas	DV Bracken: Numbers % mastery	mcs2_cm_cognitive_assessment	TV	1
dszsc	DV Bracken: Sizes Raw Score	mcs2_cm_cognitive_assessment	TV	1
dsmas	DV Bracken: Size % mastery	mcs2_cm_cognitive_assessment	TV	1
dcmsc	DV Bracken: Comparisons Raw Score	mcs2_cm_cognitive_assessment	TV	1
domas	DV Bracken: Comparisons % mastery	mcs2_cm_cognitive_assessment	TV	1
dshsc	DV Bracken: Shapes Raw Score	mcs2_cm_cognitive_assessment	TV	1
dhmas	DV Bracken: Shapes % mastery	mcs2_cm_cognitive_assessment	TV	1
dbsrc	DV Bracken: School Readiness Composite	mcs2_cm_cognitive_assessment	TV	1
dsrcm	DV Bracken: School Readiness Composite % mastery	mcs2_cm_cognitive_assessment	TV	1

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
dsrscs	DV Bracken: School Readiness Composit Standard Score	mcs2_cm_cognitive_assessment	TV	1
dsrscp	DV Bracken: School Readiness Composite Percentile	mcs2_cm_cognitive_assessment	TV	1
dsrscn	DV Bracken: School Readiness Comp Normativ Classificatn	mcs2_cm_cognitive_assessment	TV	1
dnvrs	DV BAS Naming Vocabulary - Raw Scores	mcs[2,3]_cm_cognitive_assessment	TV	2
dnvab	DV BAS Naming Vocabulary - Ability Scores	mcs[2,3]_cm_cognitive_assessment	TV	2
dnvts	DV BAS Naming Vocabulary - T-Scores	mcs[2,3]_cm_cognitive_assessment	TV	2
dpccrs	DV BAS Pattern Construction - Raw Scores	mcs[3,4]_cm_cognitive_assessment	TV	2
dpccab	DV BAS Pattern Construction - Ability Scores	mcs[3,4]_cm_cognitive_assessment	TV	2
dpccs	DV BAS Pattern Construction - T Scores	mcs[3,4]_cm_cognitive_assessment	TV	2
dpccrs	DV BAS Picture Similarities - Raw Scores	mcs3_cm_cognitive_assessment	TV	1
dpccsab	DV BAS Picture Similarities - Ability Scores	mcs3_cm_cognitive_assessment	TV	1
dpccsts	DV BAS Picture Similarities - T Scores	mcs3_cm_cognitive_assessment	TV	1
dwrss	DV BAS Word Reading - Raw Scores	mcs4_cm_cognitive_assessment	TV	1
dwrab	DV BAS Word Reading - Ability Scores	mcs4_cm_cognitive_assessment	TV	1
dwrts	DV BAS Word Reading - T Scores	mcs4_cm_cognitive_assessment	TV	1

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
dvsrs	DV BAS Verbal Similarities - Raw Scores	mcs5_cm_derived	TV	1
dvsab	DV BAS Verbal Similarities - Ability Scores	mcs5_cm_derived	TV	1
dvsts	DV BAS Verbal Similarities - T Scores	mcs5_cm_derived	TV	1
cmtotscor	DV NFER Maths Test (Total Raw Score)	mcs4_cm_cognitive_assessment	TV	1
cmaths7sc	DV NFER Maths Test (Raw score scaled to original test out of 28 marks)	mcs4_cm_cognitive_assessment	TV	1
cmaths7sa	DV NFER Maths Test (Standardised Age Score based on standardisation in 2004)	mcs4_cm_cognitive_assessment	TV	1
cwrds	DV APU Vocabulary Test of CM (Word activity score out of 20)	mcs6_cm_derived	TV	1
alcoev	Ever had an alcoholic drink	mcs[5-7]_cm_interview, mcs8_23y_cm_survey	TV	4
alcofr	Alcoholic drink frequency in the last 4 weeks	mcs[5-7]_cm_interview	TV	3
alcofi	Ever had five or more alcoholic drinks at a time	mcs[5-7]_cm_interview	TV	3
smokci	Ever smoked a cigarette	mcs[5-7]_cm_interview, mcs8_23y_cm_survey	TV	4
smokva	Ever smoked an e-cigarette or used a vape	mcs[6-7]_cm_interview, mcs8_23y_cm_survey	TV	3
weapon	Carried a weapon in the last 12 months	mcs7_cm_interview, mcs8_23y_cm_survey	TV	2

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
polsto	Ever stopped or questioned by police	mcs[6-7]_cm_interview, mcs8_23y_cm_survey	TV	3
polare	Ever cautioned, received warning or arrested by police	mcs[6-7]_cm_interview, mcs8_23y_cm_survey	TV	3
vicphy	Physically assaulted in the last 12 months	mcs[6-7]_cm_interview, mcs8_23y_cm_survey	TV	3
vicwea	Assaulted with a weapon in the last 12 months	mcs[6-7]_cm_interview, mcs8_23y_cm_survey	TV	3
sex	CM Sex at birth	mcs[1-6]_hhgrid	TI	6
dc11e	DV Cohort Member Ethnic Group - 11 category Census class (first recorded)	mcs[1-4]_cm_derived	TI	4
bwgt	Birth weight (Kilograms)	mcs1_cm_derived, mcs2_parent_cm_interview	TI	2
bwgtc	DV Birth weight categories (Low, Normal, High)	mcs1_cm_derived, mcs2_parent_cm_interview	TI	2
lowbw	DV Low birth weight (<2.5 kg)	mcs1_cm_derived, mcs2_parent_cm_interview	TI	2
gestage	Gestational age (days)	mcs1_cm_derived	TI	1
preterm	DV Pre-term Gestational age (days)< 259	mcs1_cm_derived	TI	1
d05s	DV Respondent NS-SEC 5 classes (current job) (Main Interviewee)	mcs[1-2:4-6]_parent_derived	TV	5

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
d07s	DV Respondent NS-SEC 7 classes (current job) (Main Interviewee)	mcs[1-2:4-6]_parent_derived	TV	5
d13s	DV Respondent NS-SEC major categories (current job) (Main Interviewee)	mcs[1-2:4-6]_parent_derived	TV	5
d05c	DV Respondent NS-SEC 5 classes (last known job) (Main Interviewee)	mcs[1-4]_parent_derived	TV	4
d07c	DV Respondent NS-SEC 7 classes (last known job) (Main Interviewee)	mcs[1-4]_parent_derived	TV	4
d13c	DV Respondent NS-SEC major categories (last known job) (Main Interviewee)	mcs[1-4]_parent_derived	TV	4
dacaq	DV NVQ equivalent of highest Academic qualification across previous sweeps (Main Interviewee)	mcs[1-6]_parent_derived	TV	6
dacaq_p	DV NVQ equivalent of highest Academic qualification across previous sweeps (Partner Interviewee)	mcs[1-6]_parent_derived	TV	6
dnvq	DV Respondent NVQ Highest Level (all sweeps-Main Interviewee)	mcs[1-6]_parent_derived	TV	6
dnvq_p	DV Respondent NVQ Highest Level (all sweeps-Partner Interviewee)	mcs[1-6]_parent_derived	TV	6
phde	Kessler item: How often respondent felt depressed in last 30 days (Main Interviewee)	mcs[2-7]_parent_interview	TV	6

Variable Name	Variable label	Original Dataset(s)	TV=Time-varying or TI=Time-invariant	Number of Sweeps
phho	Kessler item: How often felt hopeless in last 30 days (Main Interviewee)	mcs[2-7]_parent_interview	TV	6
phrf	Kessler item: How often felt restless/fidgety in last 30 days (Main Interviewee)	mcs[2-7]_parent_interview	TV	6
phee	Kessler item: How often felt everything an effort in last 30 days (Main Interviewee)	mcs[2-7]_parent_interview	TV	6
phwo	Kessler item: How often felt worthless in last 30 days (Main Interviewee)	mcs[2-7]_parent_interview	TV	6
phne	Kessler item: How often felt nervous in last 30 days (Main Interviewee)	mcs[2-7]_parent_interview	TV	6
kessler	DV Kessler (Main Interviewee)	mcs[2-7]_parent_interview	TV	6
wali	Respondent's Life Satisfaction (Main Interviewee)	mcs[1-5]_parent_interview	TV	5
gehe	Respondent's general health (4 categories, Main Interviewee)	mcs[1-2]_parent_interview	TV	2
gena	Respondent's general health (5 categories, Main Interviewee)	mcs[3-6]_parent_interview	TV	4