Institute of Education

UCL

Millennium Cohort Study

Sixth Survey 2015-2016

User Guide (Second Edition)

Edited by Emla Fitzsimons

(With contributions from Vilma Agalioti-Sgompou, Lisa Calderwood, Emily Gilbert, Lucy Haselden, Jon Johnson, Kate Smith and the Millennium Cohort Team)

August 2020

CENTRE FOR LONGITUDINAL STUDIES

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Institute of Education

UCL

CENTRE FOR LONGITUDINAL STUDIES

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Contents

Important note about figures of this document

Figures that are presented in this document vary compared to the totals of the datasets. This happens due to various reasons: resolution of duplicate cases or whether the data are available under End User Licence (for example, the cases that include triplets are available under Secure Access). The DATA_AVAILABILITY variable of the mcs_longitudinal_family_file marks which cases have available data under End User Licence and helps users to estimate the final sample size that can be used for research purposes under End User Licence.

The data under Secure Access Licence can be requested by applying for Data Access : https://cls.ucl.ac.uk/data-access-training/data-enhancements/ or by contacting clsfeedback@ucl.ac.uk . The mcs_longitudinal_family_file is available here: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8172 .

1. Background

1.1 The Millennium Cohort Study

The Millennium Cohort Study (MCS) is a multi-disciplinary research project following the lives of an original 18,818 children born in the UK in 2000-01. The sample was augmented in early childhood with a further 701 children born in the same period who had been missed previously, taking the total sample to 19,519 (note, there is no sample refreshment by immigrants). It is the most recent of Britain's world-renowned national longitudinal birth cohort studies. The

study has been tracking the Millennium children through their early childhood years and plans to follow them into adulthood. It collects information directly from the children, their resident parents and, in two of its sweeps, older siblings. The MCS covers such diverse topics as parenting; childcare; schooling and education; daily activities and behaviour; cognitive development; child and parent mental and physical health; employment and education; income and poverty; housing, neighbourhood and residential mobility; and social capital, ethnicity and identity.

The six surveys of MCS cohort members carried out so far have built up a uniquely detailed portrait of the children of the new century. The sixth, Age 14, survey, which is the subject of this user guide, was a major initiative, with important new additions to the study including: the collection of saliva samples from cohort members and biological parents, enabling studies of genetic influences on life course events and trajectories; physical activity and sedentary behaviour, through wrist-worn accelerometers; and detailed time-use information, through mobile phone apps/online completion. These new additions were alongside a much extended cohort member questionnaire, and continuing parent questionnaires.

To date there have been six surveys:

- MCS1 – the first sweep took place when cohort members were around 9 months old, between June 2001 and January 2003.
- MCS2 – the second sweep took place when cohort members were 3 years of age, between September 2003 and April 2005.
- MCS3 – the third sweep took place when cohort members were 5 years old; between February 2006 and January 2007.
- MCS4 – the fourth sweep took place when cohort members were 7 years old; between late January 2008 and February 2009.
- MCS5 – the fifth sweep took place when cohort members were 11 years old, between January 2012 and February 2013.
- MCS6 – the sixth, most recent, sweep took place when cohort members were 14 years old, between January 2015 and April 2016.

Funding of MCS6

The sixth sweep of the Millennium Cohort Study was core-funded by the Economic and Social Research Council (ESRC), and co-funded by the following consortium of government departments: Department for Education, Department of Health, Ministry of Justice, Home Office, Department for Transport, Department of Work and Pensions, Welsh Government and Department for Employment and Learning (Northern Ireland).

1.2 Overview of MCS6

The sixth sweep of the Millennium Cohort Study was carried out when the cohort members were 14 years old. As 14 is a key transitional age, the sweep was purposefully ambitious in the breadth and scope of its contents. It included:

- an interview (CAPI [computer-assisted personal interview] and CASI [computer-assisted self-interview]) with the main parent and partner (where relevant)
- a self-completion interview with the cohort members
- cognitive assessments for the main parent, the partner and the cohort member
- DNA collection of the cohort member and natural parents in the household
- physical measurements of the cohort member
- placement of a time-use diary with the cohort member

- placement of an accelerometer with the cohort member.

1.3 The sample

The original MCS sample covered children from all four countries of the UK who were eligible for child benefit[1] and were 9 months old at the time of the first sweep. It used a stratified, clustered random sample design and oversampled from areas that were disadvantaged or had high ethnic minority populations. This was to facilitate robust study of the effects of disadvantage on children, as well as analysis of different ethnic groups.

1.3.1 Birth dates

The cohort members were sampled from a population born across a 16-month period. This not only allowed for season of birth to be taken into account in analysis, but also had the practical advantage of allowing for a longer, less intense and more manageable fieldwork period.

- In **England and Wales** – the sample was drawn from the population of children born between **1 September 2000 and 31 August 2001.**
- In **Scotland and Northern Ireland** – the sample was drawn from the population of children born between **24 November 2000 and 11 January 2002.**

1.3.2 Stratification

In England and Wales, the population was divided into three strata:

- **The ethnic minority stratum** was comprised of children living in wards where the proportion of ethnic minorities in that ward in the 1991 Census was at least 30 per cent.
- **The disadvantaged stratum** was comprised of children living in wards, other than those falling into the ethnic minority stratum, which fell into the poorest 25 per cent of wards according to the Child Poverty Index for England and Wales.
- **The advantaged stratum c**omprised children living in wards other than the two described above.

In Wales, Scotland and Northern Ireland there were only two strata (because of the low percentages of ethnic minority groups, at around 1 per cent of the population):

- **The disadvantaged stratum** was composed of children living in wards (known as 'Electoral Divisions' in Wales) that fell into the poorest 25 per cent of wards according to the Child Poverty Index.
- **The advantaged stratum** was made up of children living in other wards in these countries.

It is important to bear in mind that both the ethnic minority indicator and the Child Poverty Index are area-level measures. That means the design will be useful for identifying those who are disadvantaged or from an ethnic minority background – for those who live in areas with others from a similar background – but will be less well placed to identify those who are likely to be part of these groups but do not live in areas with similar people. Indeed, focusing on families in poverty, Plewis (2007) found that in England in 1998, about 37 per cent of

[1] Child Benefit claims covered virtually all of the child population except those ineligible due to recent or temporary immigration status.

disadvantaged families with children under 16 were living in advantaged wards; 54 per cent were in disadvantaged wards; and 10 per cent were in ethnic minority wards.[2]

1.3.3 Clustering

The sample was clustered by characteristics of electoral wards. Clustering is efficient, and it is more cost-effective to draw a cluster sample of specific areas than to sample the whole UK. It also helps in keeping fieldwork costs down as it enables interviewer workloads to be concentrated, thereby reducing travel costs. Moreover, from an analysis perspective, clustering brings the neighbourhood context into the picture, as having multiple respondents in the same areas allows researchers to better understand area effects. Another advantage of the cluster design is that data from the census and other sources can be matched at the electoral ward level. However, a drawback of cluster sampling is that estimates are less precise than those obtained from a simple random sample.

1.3.4 Drawing the sample

The sample was randomly selected within each of three strata in each country, producing a disproportionately stratified cluster sample. This means that the sample is not self-weighting, and so weighted estimates of means, variance etc. are required (Plewis 2007).

Once the sample wards were selected, a list of all children turning 9 months old during the 16-month survey window and living in those wards was generated from the Child Benefit (CB) register provided by the then Department for Social Security (DSS), now the Department for Work and Pensions (DWP). At that time, CB was a universal provision, payable (usually to the mother) from birth. The DWP wrote to all eligible families asking the CB recipient to opt out if they did not want to be included in the survey. An opt-out procedure tends to be more inclusive of marginal and low literacy respondents than an opt-in procedure, and also results in higher response rates. The DWP withdrew sensitive cases from the issued sample. These included families where children had died or had been taken into local authority care by that point, or where there was an investigation into benefit fraud within the family. In addition, if families had already taken part in the DWP's Families and Children Survey (FACS), they were excluded from the sample.[3]

Because the CB records did not include all families who had moved into the sample wards as the child approached 9 months, an additional sample was drawn using health visitors to find eligible families who had moved into the selected areas and who had eligible children. Fifty-six families were found in this way.

1.3.5 The original sample size

The MCS1 survey reached 18,552 families which, after allowing for 256 sets of twins and 10 sets of triplets, amounted to 18,818 cohort children. Six families have two singletons in the sample. The table below shows how these respondents are distributed across the four countries of the UK. Further details by stratum appear in the Technical Report on Sampling (4th edition) (Plewis 2007).

	Number of sample 'wards'*	Target sample as boosted	Achieved responses**	
			Children	Families
England	200	13,146	11,695	11,533
Wales	73	3,000	2,798	2,760

[2] Percentages do not add to 100% due to rounding.
[3] This affected only 40 cases.

Scotland	62	2,500	2,370	2,336
N. Ireland	63	2,000	1,955	1,923
Total UK	398	20,646	18,818	18,552

*Counting amalgamations in 'superwards' as a single unit
**All productive contacts

1.3.6 MCS sample Sweeps 2 to 5

- **MCS2** – The sample issued for MCS2 consisted of productive families at MCS1 and new families that, although eligible, had not participated in MCS1. The total issued sample was 19,870; 18,481 were productive families at MCS1 and 1,389 were new families.
- **MCS3** – The sample issued for MCS3 comprised all those who had responded to the survey at least once, i.e., to MCS1 (18,522) or to MCS2 (including 692 additional cases who had responded to MCS2 as new families). There were 19,244 families potentially eligible for inclusion in the issued sample; however, 718 families were not issued to the field due to ineligibility (death or emigration), permanent refusal or sensitive family situations.
- **MCS4** – The sample for MCS4 was the same as for MCS3 (i.e., those who had responded at least once to MCS1 and MCS2). There were 19,244 families potentially eligible for the survey. However 2,213 cases were not issued to the field due to ineligibility from death or emigration, permanent refusal or sensitive family situations.
- **MCS5** – The sample for MCS5 was the same as for MCS3 and MCS4 (i.e., those who had responded at least once to MCS1 and MCS2). There were 19,244 families potentially eligible for the survey. However, 2,581 were not issued to the field due to ineligibility from death or emigration, permanent refusal or sensitive family situations.

Full details on the samples and responses for each of these sweeps can be found in their respective user guides.

2. The MCS 6 sample and response

2.1 MCS6 sample

Potentially eligible families – There were 19,243 families potentially eligible for MCS6 (one less than in previous waves as one family was identified as having had duplicate records in previous waves).

Not issued families – 3,828 families were not issued into the field (due to death or emigration; permanent refusal; untraceability; sensitive situations).

Issued cases – 15,415 cases were issued into the field.

2.2 MCS6 response

MCS6 overall response is shown in the table below:

Outcome code	Number of families	Percent
Productive	11,726*	60.9
Refusal	6,109	31.7
Other unproductive	115	0.6
Ineligible	668	3.5
Untraced	550	2.9
No contact	75	0.4

*Note that 12 productive families had lost information on the household grid. As a consequence they have been removed from the deposited data as we have no information on which parental figures are eligible or which responded. Thus the number of families available for analysis is 11,714.

3. Survey development and contents

3.1 Development and piloting of MCS6

Development work on the MCS was extensive and covered the elaboration of survey contents, instruments and materials as well as study engagement and branding. It also included a pilot and a dress rehearsal which tested all aspects of the survey. Details of all the development phases are provided below:

- **Participation and engagement** – a number of qualitative studies examined and tested participant engagement approaches, the dynamics of family decisions about participation, experiences of taking part and preferences for mode of communication.
- **Rebranding** – the study underwent a major rebranding, following extensive focus group testing, in order to make the study materials (website, mailings) more relevant and appealing to 14-year-olds.
- **Qualitative pre-testing** – Qualitative research was carried out with young people in their third year of secondary school (aged 13-14) and their parents. Eight single-sex focus groups took place in schools across England and Scotland. A range of schools was included according to the proportion of pupils receiving free school meals (FSM), attainment levels and whether they were located in urban or rural areas. Twelve in-depth interviews were conducted at home with a parent and their child from a range of ethnic backgrounds and socio-economic groups. More information about this can be found in the MCS6 technical report and related published material.
- **Cognitive testing** – Selected sections of the young person questionnaire were cognitively tested in October and November 2013. Specific objectives were to test question wording to ensure comprehension by 14-year-olds; to explore how young people understood and interpreted the meaning of specific terms and words in the questions; to understand the cognitive processes young people went through to provide answers (for example, how they retrieved, derived and reported their answers); and to provide recommendations to change the wordings of questions to improve reliability.
- **Development of the time-use record –** As this was a new component at MCS6, the development of the time-use record instruments involved extensive development work. It was led by CLS in collaboration with Ipsos Mori (IM) and the Centre for Time Use Research (CTUR) at the University of Oxford. CLS oversaw and contributed to all aspects of the development. IM produced the time-use record instruments and leaflets

and carried out the different testing phases. CTUR made a major contribution to the instrument development, regularly advising on key research design and implementation decisions. This covered both cognitive testing of the activity codes and usability testing of the survey instruments (see MCS6 Technical Report). More details on the survey instruments can be found in the time-use diary section.

- **Development of survey materials –** In July 2014, IM conducted interviews with young people to test the Young Person Engagement Materials developed for the dress rehearsal. The objectives of the materials testing were to explore young people's understanding of the language used, particularly in more complex sections (such as data linkage [which was subsequently dropped for the mainstage] and saliva); examine their understanding of the images and associated connotations; and gauge overall reactions to the materials (e.g., whether they liked them, length, etc.). Full details of the survey materials testing can be found in the technical report.
- **First pilot** – The first pilot survey took place between 7 February and 2 March 2014 in five locations in England, Scotland and Wales using a quota sample to ensure that a representative cross-section of families was included. An external agency recruited families with a child in Year 9 in England and Wales and Secondary 3 in Scotland, aged 13 to 14. Fifty families were interviewed, ten in each area. The pilot aimed to test the approaches to MCS6. This included testing the length of the questionnaire, the ethical considerations and consent; testing the implementation of each study element; and assessing the training approach and the materials used as well as the office procedures. Further details of the first pilot can be found in the technical report.
- **Dress rehearsal –** The dress rehearsal fieldwork took place between 4 July and 20 August 2014 in 13 locations across England, Scotland, Wales and Northern Ireland. The sample comprised a longitudinal sample previously recruited by CLS and used for the dress rehearsal piloting of previous waves of the study as well as a top-up sample sourced from the National Pupil Database (NPD) in England and via schools in Wales, Scotland and Northern Ireland. The sample was located in 13 areas. In total, 200 addresses were issued. Of these, 152 were longitudinal samples and 48 were new families. Because the dress rehearsal was designed to fully assess every aspect of the survey design and implementation, it mimicked the main stage procedures and content as closely as possible.

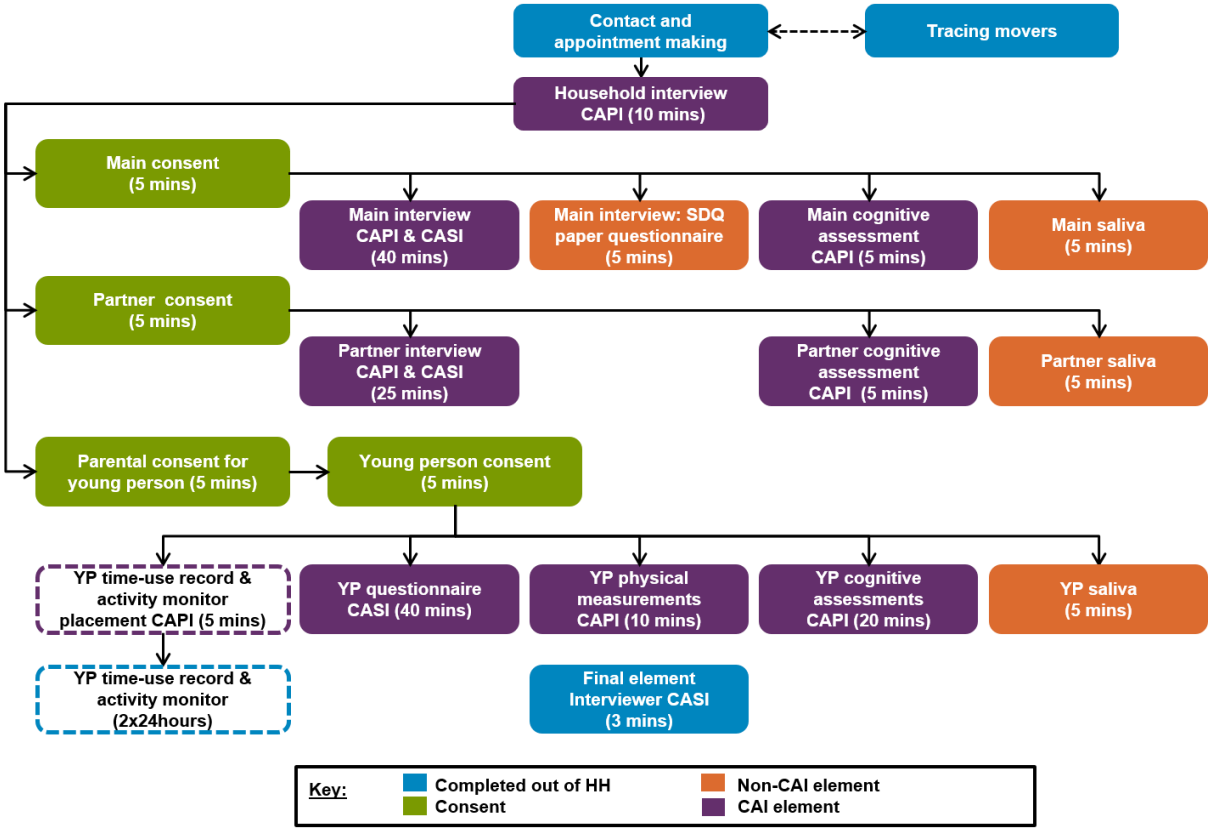
3.2 Content

The survey contained the following key elements:

1. The household questionnaire
2. The main parent questionnaire
3. The partner questionnaire (or proxy partner questionnaire), where present
4. The cohort member (referred to interchangeably as 'young person') self-completion questionnaire
5. Cognitive assessments:
   a. A word activity for the main, partner (if present) and cohort member
   b. Cambridge Neuropsychological Test Automated Battery (CANTAB) Cambridge cognition gambling task
6. Physical measurements (height, weight and body fat) of the cohort member
7. Saliva samples from the cohort member and natural parents (if present)
8. Time-use diary – Placed at the time of interview, to be completed by cohort member for two specified days (one weekday and one weekend) after the interview

9. Accelerometry – Placed at the time of interview, to be worn by cohort member on two specified days (one weekday and one weekend) after the interview (to match with time-use record days).

The diagram below provides an overview of the survey elements. It also indicates average timings for each element, mode of administration, which consents were required (and when), and whether the element was completed during or outside of the household visit. This chart was used in the interviewer briefings to help interviewers to understand how each of the different household elements fitted together and to ensure that the visit was conducted as efficiently as possible.



The contents of each of the elements can be found in the respective sections of the user guide.

4. Fieldwork

Following a competitive tender process, IM was appointed to carry out the fieldwork for MCS6. The first wave of the mainstage fieldwork began in all countries in January 2015.

4.1 Briefings

All interviewers attended a three-day briefing before working on the survey. The briefings were run by researchers from IM and CLS, members of the IM internal field team and region managers or region co-ordinators from IM's field force. In total, 291 interviewers completed all three days of the briefing. The size of the briefings varied between regions and attendance ranged between 7 and 45 interviewers.

4.2 Fieldwork timetable

Fieldwork was conducted between 15 January 2015 and 30 March 2016. The fieldwork timetable for MCS6 was driven by the requirement to interview the family during Year 9 (Year

S3 in Scotland and Year 10 in Northern Ireland). As at previous sweeps, the fieldwork was compressed into school years. In England and Wales, the cohort birth dates span a single school year. However, in Scotland and Northern Ireland the birthdates are spread over more than one school year. In England, Wales and Northern Ireland, school year is normally determined by date of birth. In Scotland, school year is determined by parental choice in addition to date of birth. It is worth noting that because of fieldwork overrunning for this sweep of MCS, 14 per cent of families were interviewed in a different school year.

Fieldwork was split into three phases and nine waves, as shown in the following table:

	Wave	Fieldwork dates	Countries	Date of birth	Date due to start Year 9 (England & Wales) / Year S3 (Scotland) / Year 10 (Northern Ireland)
Phase 1	1	Jan 2015 – Dec 2015	England, Wales	1 Sep 2000 – 31 Aug 2001 in England and Wales 24 Nov 2000 – 28 Feb 2001 in Scotland 1 Sep 2000 – 31 Aug 2001 in England & Wales 24 Nov 2000 – 28 Feb 2001 in Scotland	Sep 2014 in England and Wales Aug 2014 in Scotland
	2	Feb 2015 – Feb 2016	England, Wales, Scotland, Northern Ireland		
	3	Mar 2015 – Dec 2015	England, Wales, Scotland, Northern Ireland		
Phase 2	4	Apr 2015 – Feb 2016	England, Wales, Scotland, Northern Ireland		
	5	May 2015 – Dec 2015	England, Wales		
Phase 3	6	Aug 2015 – Mar 2016	Scotland	24 Nov 2000 – 11 Jan 2002 in Scotland 2 Jul 2001 – 11 Jan 2002 in Northern Ireland	August 2015 in Scotland September 2015 in Northern Ireland
	7	Sep 2015 – Mar 2015	Scotland, Northern Ireland		
	8	Oct 2015 – Mar 2016	Scotland, Northern Ireland		
	9	Nov 2015 – Mar 2016	Scotland		

4.3 Languages

A breakdown of the interviews by 'language interviewed in' is provided in the technical report.[4] Respondents in Wales were provided with all main communication materials in both languages and were also able to choose which language they participated in. At the appointment-making stage, families were asked if they would like any of the parent or young person elements to be administered in English or Welsh. If the family requested the interview to be conducted in Welsh, the address was reallocated to a Welsh-speaking interviewer.

To support participation of parents with limited English, other language materials were provided. (They were not provided or required for young people because all cohort members

[4] The MCS6 Technical Report on fieldwork is available on the CLS website: http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=1109&sitesectiontitle=MCS6+(2015)~~The MCS6 Technical Report on fieldwork is available on the CLS website: http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=1109&sitesectiontitle=MCS6+(2015)~~

were born in the UK and therefore have good spoken English). Parents' materials were provided in the seven languages most commonly required at previous sweeps of the study: Arabic, Bengali, Gujarati, Hindi, Punjabi (Gurmukhi script), Punjabi (Urdu script) and Urdu. Occasionally the main and partner respondents were unable to speak English or were uncomfortable with completing the interview in that language. In such cases, interviewers were instructed to find a 'household interpreter' or other informal interpreter to translate some of the elements (the consent process and parent questionnaire only). If a household interpreter was not available, the address was reallocated to a bilingual interviewer to conduct the interview. The nature of any language support given to respondents was recorded in the main parent questionnaire CAPI section – specifically, whether either of the parent interviews were translated and if so, which language and who translated (including any interviews in Welsh); and whether any translated materials were used by the main and partner respondents and if so, which language. Further details about translations can be found in the technical report.

5. Data conventions

5.1 Variable names

Each question name in the instrumentation is made up of four letters. Each variable name in the data is eight characters long – made up of the four-letter question name (e.g., ETHE), two single-letter prefixes and two single-character suffixes as follows:

[prefix1][prefix2][question name][suffix1][suffix2] where:

- **Prefix1**: indicates the sweep; a = MCS1; b = MCS2; c = MCS3; d = MCS4; e = MCS5; f = MCS6
- **Prefix2**: Identifies the instrument/respondent main and partner respondents with the prefixes m and p respectively and proxy partner interviews with x. The full list of potential variants for prefix 2 is:

Prefix2	Instrument/respondent
P	Parental respondent
X	Proxy interview
H	Household module completed by main or partner respondent
D	Derived
C	Cohort member 'level' data

- **Question name**: the four-letter question name in the questionnaire
- **Suffix1**: identifies the iteration, i.e., where the same question is repeated for different events/individuals:

Suffix1	Iteration
0	no iteration
1	1st iteration
2	2nd iteration
3	3rd iteration
…..	and so on

- **Suffix 2**: identifies a multi-coded variable, ie, where a single question produces more than one answer:

Suffix2	Iteration
0	no multi-code answer
A	1st iteration
B	2nd iteration
C	3rd iteration
…..	and so on

Taking this together, the variable names on the dataset have the following form:

[Sweep][Instrument][Question name][iteration][multi-coding]

5.2 Variable labels

Variables are labelled in a consistent manner to aid navigation within the datasets. Labels have abbreviated descriptions to indicate sweep, instrument and position in loops, as follows:

Abbreviation	Description
S6	Sweep 6 (NB similar abbreviations are used for Sweeps 1-5)
DV	derived variable
COG	cognitive assessments, e.g., word activity and Cambridge Gambling Task
PHYS	physical measurements, e.g., height and weight
MC	These appear at the end of labels and indicate a multi-coded question
R	These appear at the end of labels and indicate an event loop
IWR	These indicate the capture of an interviewer response
IWInf	These indicate the capture of interviewer information

6. The household grid and the household questionnaire

6.1 Background and introduction

6.1.1 What is the household grid?

The household grid is part of an initial household module which is administered before any other module in the interview. It contains information about every person in the household and includes two types of information: individual identifiers and identifying characteristics (number, sex and date of birth), and cross-sectional variables (e.g., relationships between household members).

The household grid contains one record for each person who has ever appeared in the household, for each family who participated in that sweep. Each household has a unique number (MSCID).

There is a variable which indicates for each person whether or not they were present at any particular sweep: AHCPRS00, BHCPRS00, CHCPRS00, DHCPRS00, EHCPRS00, FHCPRS00 for cohort members in MCS1, MCS2, MCS3, MCS4, MCS5 and MCS6 respectively, and AHPRES00, BHPRES00, CHPRES00, DHPRES00, EHPRES00 and FHPRES00 respectively for other people in the household. These can be used to identify people moving into, out of or back into the household by merging the household grid files from each sweep. Details about the household grid for previous sweeps can also be found in the respective user guides.

6.1.2 How is the household grid information collected?

At MCS6 the household grid was collected as part of the household module. It can be completed by any adult in the household (although in practice it is usually the main or the partner). It was collected at the start of the household visit, as its contents determined who was eligible for other elements (especially the main and the partner elements). The household grid used data fed forward from previous waves. In this way, it was possible to check whether each person identified as being present at Sweep 6 had been present at any of the previous sweeps: the person completing the grid was asked to list all of the people currently present in the household and, for each person, was asked if that person was someone whom we had listed as living in the household previously so that they could be assigned the same person number. If they had never been listed as living in the household before, they were assigned a new person number.

6.2 Contents of the household grid and household questionnaire

6.2.1 What information is collected in the household grid?

The household grid collected (or confirmed) the following information for each person in the household:

- who is living in the household currently, preserving the person number of those who have appeared previously
- what happened to people who were household members at the last sweep interviewed but who are not currently present in the house
- name, sex and date of birth of new people in the household (and confirmation of these details for previously listed people)
- whether each household member is a full-time or a part-time member of the household
- the working status of adults (aged 16 and over)

- relationship of each household member to the cohort member and to each other
- who has the main responsibility for caring for the cohort member.

This data is stored in the file mcs6_hhgrid.

6.2.2 What other information was collected in the household questionnaire and where is it stored?

The household questionnaire covers a number of other topics. As this information covers the household, it appears in the parent interview file:

- whether the cohort member is in a care home
- the country in which the interview is taking place
- whether the address is the same as at the last interview and dates of any moves
- repetition of some household grid variable for the main and partner respondent for ease of use
- selection of main and partner
- establishment of legal parental responsibility (for consents)
- confirmation of key identification and contact details given in the live sample (for cohort members as well as main and partner, where applicable)
- consent information.

6.3 Data format

The data is available as one row per person (including cohort members) ever in the household for productive families.

6.3.1 The household grid in previous sweeps

6.3.2 MCS1 and MCS2

At MCS2, the household grid was collected independently from MCS1; i.e., the MCS1 grid was not 'fed forward'. In subsequent sweeps the household grid was fed forward and soft checks were applied for basic identification such as date of birth, name and sex.

6.3.3 Merging household grids

The household grids can be combined across sweeps, using the family identifier (MCSID) and the person number of non-cohort members using [X]PNUM00 (where X is A,B,C,D,E or F). The cohort members can also be added using [X]CNUM00 (where X is A,B,C,D,E or F).

6.3.4 Known issues and data cleaning

While every attempt has been made to ensure consistency across the sweeps, reporting of relationships in particular is sometimes problematic. For instance, the relationship between parental figures themselves will change due to changes in cohabitation, marriage and divorce and this will have an effect on the relationship to their children and the relationship between siblings. This means that children will shift between 'step' and 'adopted' for instance, so care should be taken with the use of adoption as a category. Also, there are a few families where there are more than two natural parents. We have concentrated in the data cleaning on trying to ensure consistency between parental figures and their relationship to cohort members.

6.3.5 Derived variables

Household composition variables have been produced that are comparable to previous sweeps. These are available in the mcs6_family_derived file.

Variable name	Description
FDRSP000	S6 DV Parent interview response summary
FDMINT00	S6 DV Main interview outcome
FDPINT00	S6 DV Partner interview outcome
FDHTYP00	S6 DV Parents/carers in household
FDHTYS00	S6 DV Summary of parents/carers in household
FDRELP00	S6 DV Relationship between parents/carers in household
FDNATM00	S6 DV Natural mother status
FDMINH00	S6 DV Natural mother in household
FDNATF00	S6 DV Natural father status
FDFINH00	S6 DV Natural father in household
FDOTHS00	S6 DV Number of siblings of cohort member in household
FDNOCM00	S6 DV Number of cohort members in household
FDTOTS00	S6 DV Number of siblings in household plus cohort members
FDNSIB00	S6 DV Natural siblings of cohort member in household
FDHSIB00	S6 DV Half-siblings of cohort member in household
FDSSIB00	S6 DV Step-siblings of cohort member in household
FDASIB00	S6 DV Adopted siblings of cohort member in household
FDFSIB00	S6 DV Fostered siblings of cohort member in household
FDGPAR00	S6 DV Grandparent of cohort member in household
FDOTHA00	S6 DV Other adult in household
FDNUMH00	S6 DV Number of people in household excluding cohort member
FDTOTP00	S6 DV Number of people in household including cohort member

7. Overview of parent questionnaires

7.1 Background and introduction

7.1.1 What are the main and partner interviews?

At MCS6 (as at MCS 1, 2, 3, 4 and 5), there were three possible parent interviews which could be completed with up to two different people per family: main, partner and proxy. The main questionnaire was given to one parent or carer (usually the mother) of the young person. It used a combination of CAPI (computer-assisted personal interview) and CASI (computer-assisted self-interview) and included questions about the cohort member(s) as well as about the parent/carer him or herself and the household circumstances. The partner questionnaire was given to their partner (if resident in the household). The proxy partner interview was asked of the main respondent about the partner in certain circumstances. Interviewers were instructed to only conduct a proxy partner interview when the partner was not available for interview for the fieldwork period. The questions asked of the partner were a subset of those asked of the main; the questions asked for proxy partners were a subset of those asked of partners.

7.1.2 How were the main and partner identified at MCS6?

The main and partner respondents were established during the household questionnaire using an algorithm within the CAPI questionnaire. The algorithm was based exclusively on relationships between household members. In certain circumstances the interviewer could override this selection and replace the person selected for the main interview with the one selected for the partner interview. This might happen if, for instance, the person selected in CAPI as the partner was the main carer, or if the mother was unwilling to take part but her partner was, and was willing to be interviewed as the main respondent.

In most cases, the CAPI selected the mother figure to complete the main questionnaire. However, there were notable exceptions: If the father was the only natural parent in the household, he was chosen; if there were no parents (including step, foster and adoptive) living with the young person, the CAPI selected the main carer and his or her partner for interview. If the person selected as the main had a partner living in the household, that person would be eligible to complete the partner interview.

For details of identification of main and partner in previous sweeps see Hansen (2008).

7.1.3 Are the main and partner the same people in all sweeps?

Not necessarily, although the criteria for the selection of main and partner were the same (or very similar) in each sweep, with main generally being the natural mother where possible. However, in some cases the household composition changed over time; in others, the CAPI selection was overridden by the interviewers. The table below shows the proportion of households with the same main respondent, as at their previous interview.

Sweep	**Same main respondent**
MCS2	98.7%
MCS3	96.7%
MCS4	96.5%
MCS5	96.1%
MCS6	95.5%

It is possible to check the person number of the main for MCS6 in FPNUM00. The person number for the main has been stored in similar variables for previous sweeps. The person number assigned to a household member is consistent throughout all prior sweeps.

7.2 Baseline numbers

MCS6 Proportion of households with main and partner interviews

	Frequency	Percent
1. Main respondent in person (no-one eligible for partner)	2,853	24.4
2. Main and partner respondent interviewed in person	7,230	61.7
3. Main respondent interviewed in person; partner by proxy	334	2.9
4. Main respondent interviewed in person; Partner eligible but no response	1,170	10.0
5. No main interview; partner interviewed in person	33	0.3
6. No main interview; no eligible partner	28	0.2
7. No parent interview	66	0.6
All productive households	**11,714**	**100**

Note: Percentages may not sum to 100 due to rounding.

MCS6 Sex and relationship of main respondent to cohort member

	Frequency	Percent
Female		
Natural mother	10,782	98
Other	225	2
Male		
Natural father	664	94
Other	43	6
All main respondents	**11,714**	**100**

7.3 Contents of the main, partner and proxy partner interviews

7.3.1 Topics covered in main and partner questionnaires

The table below shows the main topics covered in the MCS6 main, partner and proxy interviews. All were through interview (CAPI), with the exception of those under 'self-completion', which were done on their own (without interviewer) through CASI.

Broad topic	Sub-topic	Main	Partner	Proxy
Family context	Marital status	✓	✓	✓
	Languages spoken at home	✓		
	Ethnic group	✓	✓	✓
	Reasons for separation from previous partner	✓		
	Relationship/contact with absent parent of cohort member	✓		
	Cohort member contact with natural parents	✓		
	Own parents	✓	✓	✓
	Own childhood circumstances	✓	✓	
	Periods when cohort member was not living with main	✓		
Education and schooling	School year	✓		
	Details of the school the cohort member currently attends (or reasons for not currently attending school)	✓		
	Reasons for attending a fee-paying or faith school	✓		
	Language taught in	✓		
	Details of other schools attended and reasons for leaving	✓		
	Periods of absence from school	✓		
	Details of any special needs	✓		
	Parental aspirations for cohort member after leaving school	✓	✓	
	Parent–school communication	✓		
	Details of any extra tuition	✓		
	Travel to school	✓		
	Free school meals	✓		
Parenting activities	Eating together	✓		
	Cohort member chores	✓		
	Parental knowledge of cohort member going out	✓		
Young person's health	Details of cohort member's longstanding illnesses	✓		
	Eyesight problems	✓		
	Speech problems	✓		
	Details of atopic conditions	✓		
	Details of certain communicable diseases	✓		
	Behavioural disorders	✓		
	Accidents and injuries	✓		
	Admissions to hospital	✓		
	Vaccinations	✓		
Parent's health	Parent's general health	✓	✓	✓
	Details of any longstanding illnesses	✓	✓	✓
	Smoking	✓	✓	
	Physical health	✓	✓	
Employment and income	Details of current jobs (including second jobs)	✓	✓	✓

	Employment and unemployment history since last interview	✓	✓	✓
	Hours worked	✓	✓	✓
	Paid and unpaid overtime	✓	✓	✓
	Activity if unemployed	✓	✓	✓
	Reasons for absence from employment	✓	✓	✓
	Current pay	✓	✓	✓
	Partner's current pay	✓		
	Benefits received	✓	✓	✓
	Pensions received	✓	✓	✓
	Other income	✓	✓	
	Assets	✓	✓	
	Debts	✓	✓	
	Financial wellbeing	✓	✓	
	Qualifications since last interview	✓	✓	
	Literacy and numeracy	✓	✓	
Housing	Details of current accommodation	✓		
	Cohort member's bedroom	✓		
	Tenure	✓		
	Rent	✓		
	Housing benefit	✓		
	Mortgage	✓		
	Reasons for moving to this address	✓		
	Housing history since last interview	✓		
	Periods of homelessness	✓		
	Car ownership	✓		
	Local area	✓		
Other matters	Religion	✓	✓	
	Time spent with children	✓	✓	
Self-completion	Personality traits (Big Five)	✓	✓	
	Relationship with cohort member	✓	✓	
	Cohort member's ability to control emotions	✓	✓	
	Own mental health	✓	✓	
	Own alcohol consumption	✓	✓	
	Cohort member's alcohol consumption	✓		
	Own drug consumption	✓	✓	
	Own happiness	✓	✓	
	Violent relationships	✓	✓	
	Partner living outside the house	✓		
	Life satisfaction	✓	✓	
Contact information	Address and contact details used by the MCS team to maintain contact for future waves	✓	✓	
Strengths and Difficulties Questionnaire (SDQ)	A strengths and difficulties questionnaire on paper	✓		

7.3.1.1 Income

Income has been collected at each wave of the MCS through two banded questions administered to two-parent and single-parent families respectively. This section

describes the collection of income measures in the survey, and the derivation of the income-derived variables at MCS6 and poverty indicator.

7.3.1.2 Banded data

Respondents were shown a card with weekly, monthly and annual bands of total take-home income from all sources and earnings after tax and other deductions. These sources implicitly included state benefits, which had been the subject of more detailed previous questions. Note that, unlike other state benefits, there was no attempt to ascertain the amounts of Housing Benefit and Council Tax Benefit as separate components. Therefore, these may well have been omitted from estimates of total net income. Bands of different sizes were used for single- and two-parent families. In the latter, both main respondents and partners answered the banded income question. There are, therefore, two reported values for each household in the case of two-parent families. In 54 per cent of the households with two parents/carers, main respondents and partners reported a different value.

7.3.1.3 Missing income data (item non-response)

Some families did not report income: the table below shows that 1,689 of MCS families in wave 6 did not provide banded income data, and shows why.

Missing data on the banded income questions (number of families)

	Number of families
Missing income data (refusal)	368
Missing income data (don't now)	1,179
Other missing	142
Total missing	1,689
Observed number of families	11,714

7.3.1.4 Imputation of missing and continuous income from banded data

Income was imputed for the main respondents and for partners in two-parent families, where missing, using interval regression (Stewart 1983). This method allowed us to impute a continuous value within a band, rather than assuming that all cases in a band had the same midpoint income. This was achieved using Stata's INTREG command (StataCorp 2007; Conroy 2005). INTREG fits a model of y = [dependent variable 1, dependent variable 2] on independent variables where the dependent variable 1 was the log lower income band and dependent variable 2 was log upper income band. Note that the left-hand-side bound for the lowest band is 0 and the right-hand-side bound for the top band is the 100th income percentile in the UK. The predictors are shown in the table below.

Predictors of income in MCS6

Variable	**Categories**
Main respondent's age at interview	Continuous
Housing tenure	Own Private renting Renting from local authority or housing association

	Other
Person currently in work	Yes No
Sampling point type	Advantaged Disadvantaged Ethnic
DV interview government region	North East North West Yorkshire and the Humber East Midlands West Midlands East of England London South East South West Wales Scotland Northern Ireland
In receipt of state benefit?	No
	Yes
Main respondent's ethnic background	White Mixed Indian Pakistani and Bangladeshi Black or Black British Other ethnic group (including Chinese and other Asian)
DV combined education Highest level of either respondent	NVQ level 1 NVQ level 2 NVQ level 3 NVQ level 4 NVQ level 5 Overseas quallfication only Other including NA
Main type of accommodation	A house or bungalow A flat or maisonette A studio flat
Number of children including cohort member	1; 2; 3; 4 or more
Derived variable number of parents /carers in household	Two parents/carers One parent/carer

7.3.1.5 Income-derived variables

Variable name	**Description**
FOEDE000	S6 DV OECD Equivalised weekly family Income

FOEDP000	S6 DV OECD Below 60% mean indicator
FOECDUK0	S6 DV OECD Equivalised income quintiles – UK whole
FOECDSC0	S6 DV OECD Equivalised income quintiles – by country

7.3.1.6 Equivalisation

After imputation, the values for main respondents and partners were averaged for families with two parents. This yielded a continuous income measure for each family in MCS6. Modified OECD scales were used for equivalisation. Each scale sets the family's needs relative to those of a couple with no children, whose scale is set equal to 1. In the modified OECD scale, a family of one parent and one child under 14 has a scale of 0.87; one parent and two children 1.07; and so on. This is shown in the table below.

OECD household equivalence scales

Equivalence scales before housing cost	**OECD scale used**
First adult (main respondent)	0.67
Spouse	0.33
Dependent child aged between 14 and 18 years	0.33
Child aged under 14 years	0.2

7.3.1.7 Construction of the poverty indicator

A binary poverty indictor was constructed based on the OECD equivalised income. The indicator takes the value 1 if the OECD equivalised income of the household is below 60 per cent of the DWP national HBAI weekly median income, which was equal to £284 in 2014-15 (DWP 2016). Moreover, two variables containing income quintiles were constructed (one for the UK as a whole and one for each of the four countries of the UK: England, Wales, Scotland and Northern Ireland). The quintiles were weighted using the overall weights in MCS6 for the UK as a whole and for 'by-country' analyses.

7.4 Scales

7.4.1 Big Five personality traits

The **Big Five personality traits**, also known as the **five factor model** (**FFM**), is a model based on common language descriptors of personality. The five factors have been defined as openness to experience, conscientiousness, extraversion, agreeableness and neuroticism, often listed under the acronyms OCEAN or CANOE.

The main and partner were each asked to rate how much each of the following 15 statements applied to them using a scale of 1 to 7, where 1 is 'does not apply to me at all' and 7 is ' applies to me perfectly'.

Question name	Question	Variable name
BIGA	I see myself as someone who is sometimes rude to others	FPBIGA00
BIGB	I see myself as someone who does a thorough job	FPBIGB00
BIGC	I see myself as someone who is talkative	FPBIGC00
BIGD	I see myself as someone who worries a lot	FPBIGD00
BIGE	I see myself as someone who is original, coming up with new ideas	FPBIGE00
BIGF	I see myself as someone who has a forgiving nature	FPBIGF00
BIGG	I see myself as someone who tends to be lazy	FPBIGG00

BIGH	I see myself as someone who is outgoing, sociable	FPBIGH00
BIGI	I see myself as someone who gets nervous easily	FPBIGI00
BIGJ	I see myself as someone who values artistic, aesthetic experiences	FPBIGJ00
BIGK	I see myself as someone who is considerate and kind to almost everyone	FPBIGK00
BIGL	I see myself as someone who does things efficiently	FPBIGL00
BIGM	I see myself as someone who is reserved	FPBIGM00
BIGN	I see myself as someone who is relaxed, handles stress well	FPBIGN00
BIGO	I see myself as someone who has a big imagination	FPBIGO00

Note, parental personality type was assessed at MCS4 using a different set of questions. These and other scales can be found in Johnson *et al.* (2015).

The Big Five derived variables

Description	Sub-scale	Variable name
FDOPEN	Openness	S6 DV OCEAN
FDCONSC	Conscientiousness	S6 DV OCEAN
FDAGREE	Agreeableness	S6 DV OCEAN
FDNEUROT	Neuroticism	S6 DV OCEAN
FDEXTRAV	Extroversion	S6 DV OCEAN

Derived variables were constructed as follows:

FDOPEN = FPBIGE00 + FPBIGJ00 + FPBIGO00
FDCONSC = FPBIGB00 + FPBIGG00* + FPBIGL00
FDAGREE = FPBIGA00* + FPBIGF00 + FPBIGK00
FDNEUROT = FPBIGD00 + FPBIGI00 + FPBIGN00*
FDEXTRAV = FPBIGC00 + FPBIGH00 + FPBIGM00*

The sub-scales are only computed if all the component items are completed. *Denotes the item was reversed.

7.4.2 Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a behavioural screening questionnaire for 3- to 16-year-olds. It measures 25 items on psychological attributes (Goodman 1997).

At MCS6, the **P4-17** - SDQ and impact supplement for the parents of 4-17-year-olds version was used.

The respondent is asked to comment on the following statements with response options: Not true, Somewhat true or Certainly true.

Question name	Question	Variables	Equivalent on SDQ
SDPF	Considerate of others' feelings	FPSDPF00	SDQ item 1
SDRO	Restless, overactive, cannot stay still long	FPSDRO00	SDQ item 2
SDHS	Complains of headaches/stomach-aches/sickness	FPSDHS00	SDQ item 3
SDSR	Shares readily with others	FPSDSR00	SDQ item 4
SDTT	Often has temper tantrums	FPSDTT00	SDQ item 5

SDSP	Tends to play alone	FPSDSP00	SDQ item 6
SDOR	Generally obedient	FPSDOR00	SDQ item 7
SDMW	Often seems worried	FPSDMW00	SDQ item 8
SDHU	Helpful if someone is hurt, upset or ill	FPSDHU00	SDQ item 9
SDFS	Constantly fidgeting	FPSDFS00	SDQ item 10
SDGF	Has at least one good friend	FPSDGF00	SDQ item 11
SDFB	Fights with or bullies other children	FPSDFB00	SDQ item 12
SDUD	Often unhappy	FPSDUD00	SDQ item 13
SDLC	Generally liked by other children	FPSDLC00	SDQ item 14
SDDC	Easily distracted	FPSDDC00	SDQ item 15
SDNC	Nervous or clingy in new situations	FPSDNC00	SDQ item 16
SDKY	Kind to younger children	FPSDKY00	SDQ item 17
SD0A	Often lies or cheats	FPSD0A00	SDQ item 18
SDPB	Picked on or bullied by other children	FPSDPB00	SDQ item 19
SDVH	Often volunteers to help others	FPSDVH00	SDQ item 20
SDST	Can stop and think before acting	FPSDST00	SDQ item 21
SDCS	Steals from home, school or elsewhere	FPSDCS00	SDQ item 22
SDGB	Gets on better with adults	FPSDGB00	SDQ item 23
SDFE	Many fears, easily scared	FPSDFE00	SDQ item 24
SDTE	Sees tasks through to the end, good attention span	FPSDTE00	SDQ item 25

The above 25 items are divided between 5 scales:

1. **Emotional symptoms**
    a. Complains of headaches/stomach aches/sickness
    b. Often seems worried
    c. Often unhappy
    d. Nervous or clingy in new situations
    e. Many fears, easily scared.
2. **Conduct problems**
    a. Often has temper tantrums
    b. Generally obedient*
    c. Fights with or bullies other children
    d. Steals from home, school or elsewhere (In MCS2: Can be spiteful to others)
    e. Often lies or cheats (in MCS2: Often argumentative with adults).
3. **Hyperactivity/inattention**
    a. Restless, overactive, cannot stay still for long
    b. Constantly fidgeting
    c. Easily distracted
    d. Can stop and think before acting*
    e. Sees tasks through to the end*.
4. **Peer relationship problems**
    a. Tends to play alone
    b. Has at least one good friend*
    c. Generally liked by other children*
    d. Picked on or bullied by other children
    e. Gets on better with adults.

5. **Prosocial behaviour**
    a. Considerate of others' feelings
    b. Shares readily with others
    c. Helpful if someone is hurt, upset or ill
    d. Kind to younger children
    e. Often volunteers to help others.

*Denotes items that are reversed when generating sub-scales on behaviour.

Each of the five scales can be used alone or together to create:

- 1-4 when taken together generate a **total difficulties score**
- 1 and 4 create an internalising problems score
- 2 and 3 create an externalising conduct score
- 5 alone measures prosocial behaviour

SDQ derived variables

Variable name	Description
FDEMOT00	S6 DV SDQ Emotional symptoms
FDCOND00	S6 DV SDQ Conduct problems
FDHYPE00	S6 DV SDQ Hyperactivity/inattention
FDPEER00	S5 DV SDQ Peer problems
FDPROS00	S6 DV SDQ Prosocial
FDEBDTAA	S6 DV SDQ Total difficulties

The SDQ was also administered in Sweeps 2, 3 4 and 5. For further details see Johnson *et al.* (2015).

7.4.3 Kessler 6 scale

The Kessler 6 (K6) scale is a quantifier of non-specific psychological distress. It consists of six questions about depressive and anxiety symptoms that a person has experienced in the last 30 days. It was asked as part of the main and partner self-completion questionnaires (Kessler *et al.* 2003).

For each question, respondents were offered a self-report scale of five possible answers plus don't know/don't wish to answer:

1. All of the time
2. Most of the time
3. Some of the time
4. A little of the time
5. None of the time
6. Don't know/Don't wish to answer

The questions are preambled by the statement: 'The next few questions are about how you have felt over the last 30 days'. The six questions are:

Question name	Question	Variable
PHDE	During the last 30 days, about how often did you feel so depressed that nothing could cheer you up?	FPPHDE00
PHHO	During the last 30 days, about how often did you feel hopeless?	FPPHHO00

PHRF	During the last 30 days, about how often did you feel restless or fidgety?	FPPHRF00
PHEE	During the last 30 days, about how often did you feel that everything was an effort?	FPPHEE00
PHWO	During the last 30 days, about how often did you feel worthless?	FPPHWO00
PHNE	During the last 30 days, about how often did you feel nervous?	FPPNHE00

Kessler-derived variables

Variable name	Description
FPKESS00	S6 DV Kessler K6 Scale

Derived variable FPKESS00 was constructed by summing the items: PHDE (reversed), PHHH (reversed), PHRF (reversed), PHEE (reversed), PHHW (reversed), PHNE (reversed) where (1=4) (2=3) (3=2) (4=1) (5=0) (6=missing).

7.4.4 Alcohol questions (AUDIT-PC)

The alcohol questions in the self-completion section of the main parent questionnaire are a shortened, adapted version of the 10-item Alcohol Use Disorders Identification Test Primary Care (AUDIT-PC), developed by the World Health Organisation.

The five questions are:

Question name	Question	Variable
ALDR	How often do you have a drink that contains alcohol?	FPALDR00
AUND	How many standard alcoholic drinks do you have on a typical day when you are drinking?	FPAUND00
AUSD	How often in the last year have you found you were not able to stop drinking once you had started?	FPAUSD00
AUAC	How often in the last year have you failed to do what was expected of you because of drinking?	FPAUAC00
AUCD	Has a relative, friend, doctor or health worker been concerned about your drinking or advised you to cut down?	FPAUCD00

AUDIT-PC derived variables

Variable name	Description
FPAUDIT	S6 DV AUDIT-PC Scale

Derived variable FPAUDIT was constructed by summing the five responses. A total of 5+ indicates increasing or higher risk drinking.[5]

[5] https://www.alcohollearningcentre.org.uk/Topics/Latest/AUDIT-PC/

7.5 Unfolding brackets

Unfolding brackets were used on a number of questions. If respondents were unable or unwilling to give an exact amount as an answer to a particular question (for example exact income), unfolding brackets were used to gain an estimate of that amount. So for instance, if the respondent could not answer a question on income, s/he was asked whether his/her income was above or below a rounded income amount (for example £20,000). S/he could then be asked a series of similarly structured questions in order to narrow down the amount range.

The data is currently in preparation and more details will be available when it is finished.

7.6 Feed forward data

Some information was fed forward from earlier sweeps. The main was identified by a person number which was recorded on the Computer-Assisted Interview (CAI). If the person number for the main was associated with someone who had been the main or partner at a previous sweep, personal information about her or him was fed forward from the last sweep they had participated in. In addition, information that related to the cohort member or the household as a whole could be fed forward from the previous sweep even if the main respondent was different.

7.7 Data structures and data-handling issues

We have endeavoured to make the data available in a format that is conducive to use by as wide a range of researchers as possible. As with MCS5, we have released data at a person level, so that it is possible to combine both individual adult respondents and those of the cohort members themselves. The adult respondents can be joined with other data files in MCS6 and to prior sweeps of data collection using the person number (FPNUM00) variable. In the parent CM file, there is additionally the cohort member number (FCNUM00) variable, which allows joining of the cohort member across data collections. The advantage of this data structure (over that previously available for MCS1 to MCS4) is that the data can be combined longitudinally more straightforwardly, using the person identifier (FPNUM00) and then selecting, for instance, only natural mothers or other characteristics of interest. In addition, because the questions for both respondents are asked identically, the computation for any variable is simplified.

7.7.1 Parent interview file and parent derived file

The parental interviews are available in the data as one row per respondent for those questions asked about the parental figures, e.g. their employment, income, housing, qualifications, and those asked in the self-completion.

7.7.2 Parent CM (cohort member) file

Questions which are asked of the parental figures about the cohort members, the CM's education, health and self-completion are available in the data as one row per CM per parental response. This also includes the strength and difficulties questions asked on paper.

7.7.3 Proxy partner interview file

For those questions asked about partners who are either away or incapacitated, the person number (FPNUM00) variable relates to the person being reported upon. The data is available as one row per proxy respondent.

7.8 Known issues and data cleaning

7.8.1 Edits made by interviewers or the fieldwork agency

Interviewers carried out most of the data editing in the field, where inconsistencies were highlighted through 'soft' and 'hard' checks. The former did not allow entries outside a given range (and had to be resolved by the interviewer at the time of the interview); the latter required the interviewers to check and confirm what they had entered. These enabled interviewers to clarify and query data discrepancies directly with the respondent during the interview.

Interviewers recorded in the Final Element module instances where they believed further amendments to the data would be needed, and, in a few instances, interviewers notified Head Office where other amendments to the data were necessary. These reports were reviewed and, if required, an edit was proposed. The proposed edits were then signed off by CLS. If an edit to a variable had routing implications in the questionnaire, the following approach was taken:

- In most cases where, after applying the edit, questions had been answered that should not have been answered, the data contained in those (routed) variables was cleared. However, there were cases where this was not done due to complexity and the risk of introducing further problems. All such cases were agreed with CLS.
- In cases where questions would have been asked but were not, the response was left as '1: not applicable'. However, in some places in the script, it was beneficial to complete these variables. All such cases were agreed with CLS.

Some edits were made to body fat and height measurements based on comments interviewers had made in the physical measurements module. Additionally, edits were made in cases where the body fat value and the weight value were identical; this was deemed to be interviewer error and both values were deleted. All such edits were agreed with CLS.

All edited cases were flagged within the final data sets using a predefined set of variables, along with any cases where data was not edited but where there may have been issues. All such flags are shown in Figure 9.2 of the MCS 6 technical report.[6]

7.8.2 Derived variables

A number of variables, many of which have also been calculated for previous sweeps, are listed below. Those questions asked of the parental respondents appear in the parental derived file, and those relating to the cohort member, e.g., the Strengths and Difficulties battery appear in the CM derived file.

Detailed documentation on their derivation can be found in the MCS6 Derived Variable User Guide.

Parental derived file

Variable name	Description
FDSAM00	S6 DV Respondent same as at Sweep 5
FDLST00	S6 DV Respondent status at Sweep 5
FDRES00	S6 DV Respondent identity and interview status
FDREL00	S6 DV Respondent relationship to CM
FDAGI00	S6 DV Respondent age at interview

[6] http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=1109&sitesectiontitle=The+age+14+survey+of+MCS+(2015))

FDGAI00	S6 DV Respondent age at interview (grouped)
FPRXF00	S6 DV Proxy partner interview flag
FDWRK00	S6 DV Whether respondent is in work or not
FDEMP00	S6 DV Employment status for Standard Occupational Classification (SOC) coding
FDACT00	S6 DV Respondent economic activity status
FD17S00	S6 DV NS-SEC full version (current job)
FD13S00	S6 DV NS-SEC 13 category (current job)
FD07S00	S6 DV NS-SEC 7 category (current job)
FD05S00	S6 DV NS-SEC 5 category (current job)
FDEEA00	S6 DV Respondent's ethnic group (England)
FDEWA00	S6 DV Respondent's ethnic group (Wales)
FDESA00	S6 DV Respondent's ethnic group (Scotland)
FDENA00	S6 DV Respondent's ethnic group (Northern Ireland)
FD06E00	S6 DV Respondent's ethnic group - 6 category census classification (UK)
FD11E00	S6 DV Respondent's ethnic group - 11 category census classification (UK)
FD08E00	S6 DV Respondent's ethnic group - 8 category census classification (UK)
FDKESSL	S6 DV Kessler K6 scale
FDOPEN	S6 DV OCEAN - Openness sub-scale
FDCONSC	S6 DV OCEAN - Conscientiousness sub-scale
FDEXTRAV	S6 DV OCEAN - Extraversion sub-scale
FDAGREE	S6 DV OCEAN - Agreeableness sub-scale
FDNEUROT	S6 DV OCEAN - Neuroticism sub-scale
FPAUDIT	S6 DV AUDIT-PC scale
FDNVQ00	S6 DV Respondent NVQ highest level (all sweeps)
FDACAQ00	S6 DV NVQ equivalent of highest academic level across sweeps
FDRLG00	S6 DV respondent religion - 7 category

Cohort member derived file

Variable name	**Description**
FEMOTION	S6 DV Parent-reported CM SDQ emotional symptoms
FCONDUCT	S6 DV Parent-reported CM SDQ conduct problems
FHYPER	S6 DV Parent-reported CM SDQ hyperactivity/inattention
FPEER	S6 DV Parent-reported CM SDQ peer problems
FPROSOC	S6 DV Parent-reported CM SDQ prosocial
FEBDTOT	S6 DV Parent-reported CM SDQ total difficulties

8. Overview of cohort member questionnaire

8.1 Background and introduction

8.1.1 The cohort members

At MCS6 the cohort members were at a significant age, between childhood and adulthood. Rather than asking their parents, the Age 14 Survey focused much more on the cohort members themselves, as they were doing more in their lives and were able to report on their own activities, thoughts and feelings. The cohort members were asked to complete a 40-minute CASI (self-completion) questionnaire on the interviewer's tablet. They were encouraged to complete the questionnaire in private. In the very few cases where cohort members were unable to complete the questionnaire themselves, the interviewer could read out the questions to them. However, in these cases sensitive questions were skipped. In a small number of cases (8), households have non-twin siblings who are also cohort members (they were born within the time period of the original sample). These can be identified in the longitudinal family file in the DUALBABYFAMILY variable.

8.1.2 Twins and triplets

In households where there was more than one cohort member resident, each cohort member was asked to complete a questionnaire.

8.2 Baseline numbers

Number of cohort members in each household

	Frequency	**Percent**
Singletons	11,714	98.7
Twins	150	1.3
Triplets	8	0.1
All Cohort Members	**11,872**	**100**
Non-twin siblings	8	0.1

Note: Percentages may not sum to 100 due to rounding.

MCS6 Sex and age at interview of cohort members

	Frequency	Percent
Female, age		
13	1,416	23.9
14	2,232	74.8
15	78	1.3
Male, age		
13	1,444	24.3
14	4,420	74.3
15	82	1.4

8.3 Contents of the cohort member interview

8.3.1 Main topics

The age 14 cohort member questionnaire covered a wide range of topics about their lives and comprised the following broad areas.

Things that they do	What activities they do in their free time Watching TV and using internet, computers, games consoles Getting pocket money or money from paid work
Their views	Attitudes to gender roles and consumerism Attitudes to activities like fighting, shoplifting, and spray painting
School and their future	What subjects they study at school Homework How they feel about school Behaviour in school and truancy Their future education and work
About them	Religion and ethnicity
Their family	Relationship with parents and grandparents Parental control and discipline Contact with parents who don't live with them
Their friends	Type of friends, and what their friends are like How much time they spend with friends
Relationships	Support from friends and family Romantic relationships Sexual experiences
Things they may have tried	Smoking Drinking alcohol Illegal drugs Gambling
Things they may have experienced	Being bullied and bullying, including cyber-bullying Being a victim of crime
Things they may have done	Being involved with illegal and anti-social behaviour Contact with the police
Their health	What they eat and drink Sight and hearing problems Dental health Sleeping habits
Their body	Dieting and trying to lose weight Puberty
How they feel	How happy they feel with their lives Moods and feelings
More about them	Attitude to trust, patience and risk

8.3.2 Scales

Questions	Topic	Source
SAFF, TRSS, NCLS	Social support	Three items from Social Provisions Scale – short form
HHND – CONP	Sexual experience	Adapted from Adolescent Sexual Activity Index (ALSPAC)
FRUT, BRED, MILK	Diet	Eating Choices Index
PUHG-AGMN	Puberty	Pubertal Development Scale
SATI-GDSF	Wellbeing/self-esteem	Shortened Rosenberg Self-esteem Scale
MDSA-MDSM	Mental health	Moods and Feelings Scale short form

8.4 Data structures and data handling issues

The data collected in the cohort member interviews is available in the CM interview file and the CM derived file, with one row per cohort member.

8.5 Known issues and data cleaning

8.5.1 Edits made by interviewers or the fieldwork agency

Please refer to section 7.8.1 for this section.

8.5.2 Derived variables

These variables are available in the CM derived file.8.5.1

Variable	Description
FDCE0600	S6 DV CM ethnic group classification - 6 categories
FDCE0800	S6 DV CM ethnic group classification - 8 categories
FDCE1100	S6 DV CM ethnic group classification - 11 categories

9. Cognitive assessments

9.1 Background and introduction

9.2 Cambridge Gambling Task

Young people were asked to complete the Cambridge Gambling Task during the interviewer visit. The Cambridge Gambling Task is taken from the Cambridge Neuropsychological Test Automated Battery (CANTAB). It measures decision-making and risk-taking behaviour.

The assessment was administered using the interviewer's tablet, and the interviewer guided the young person through the assessment using a laminated script.

The young person was presented with a row of ten boxes across the top of the screen, some of which were red and some of which were blue. The young person had to decide whether a 'token' was hidden in a red box or a blue box. The young person started with a number of points displayed on the screen and had to decide what proportion of their points they were willing to risk on their decision. The young person had to try to accumulate as many points as possible.

The same assessment was completed with the young people when they were 11.

9.3 Word activity

Young people and resident parents were asked to complete the word activity during the interviewer visit. It measures respondents' understanding of the meaning of words.

The assessment involved presenting the respondent with a list of target words, each of which had five other words next to them. The respondent had to select, from the five options, the word which meant the same, or nearly the same, as the target word. Each respondent had a list of 20 target words, and the lists of words were different for the young person, main parent and partner. The assessment was carried out on the interviewer's tablet.

The words used in the word activity are subsets of those used in a vocabulary assessment in the 1970 British Cohort Study (BCS70) Age 16 survey, which took place in 1986. The words used in the BCS70 assessment come originally from the standardised vocabulary tests devised by the Applied Psychology Unit at the University of Edinburgh in 1976.

9.4 Consent

As the young people were still children, parents were gatekeepers to their participation. Information about the cognitive assessments was provided to both parents and young people in advance of the interviewer visit in the form of a booklet, which contained information on why the survey included cognitive assessments and briefly explained what each assessment was. Parents were asked to provide written consent for their own participation. Parents were also asked to provide written consent for the interviewer to approach their child to participate, and young people gave verbal consent using a structured consent form, which the interviewer read out, and then signed.

9.5 Baseline numbers

Proportion of young people who completed the Cambridge Gambling Task and word activity

	No.	%
Cambridge Gambling Task	10,842	91.3
Word activity	10,921	92.0

Proportion of main parents and partners who completed the word activity

	No.	%
Main parent	11,057	95.4
Partner	6,869	94.6

9.6 Data conventions

For the word activity exercise, different variables have been output for the main (FPMCOG0x) and partner (FPPCOG0x) respondent (as they were given different word lists, the labelling makes this clear).

9.7 Derived variables

Variable	Description
FCGTOUTCM	CGT Test outcome
FCGTTTIME	CGT Test duration (seconds)
FCGTDELAY	CGT Delay aversion

FCGTDTIME	CGT Deliberation time
FCGTOPBET	CGT Overall proportional bet
FCGTQOFDM	CGT Quality of decision-making
FCGTRISKA	CGT Risk adjustment
FCGTRISKT	CGT Risk-taking
FCWRDSC	CM Word activity score out of 20

10. Physical measurements

10.1 Background to physical measurements on MCS

At age 14 height, weight and body fat measurements were taken by the interviewer for each cohort member. Physical measurements have been collected from cohort members since the age of 3. Height and weight measurements have been taken at each survey (ages 3, 5, 7, 11 and 14). In addition, waist measurements were taken at ages 5 and 7, and body fat measurements were taken at ages 7 and 11.

Interviewers were accredited to take the physical measurements at age 14 in order to ensure accurate and consistent measurements. Reasons for not being able to take any measurement and circumstances that applied to measurements were recorded by the interviewer in CAPI.

The data collection instrument limited height to between 120cm and 200cm, and weight to between 20kg and 120kg (even where interviewers confirmed the value outside the range was correct and re-entered it).

10.2 Height

The height measurement was taken by the interviewer using a Leicester height measure. The interviewer used a Frankfort Plane card to check that the cohort member's head was in the correct position. The measurement was taken in metres and centimetres and rounded down to the nearest completed millimetre. Cohort members had to be able to stand unaided in order for the height measurement to be taken.

10.3 Weight and body fat

Weight and body fat measurements were taken together using Tanita™ scales. The body fat measurement was taken by sending a weak electronic current around the body from one foot to the other. The scales measure the amount of resistance encountered by the current as it travels round the body. As muscle and fat have different levels of resistance, the scales use this to calculate body fat percentage. Weight measurements were recorded in kilograms and body fat was recorded as a percent, both to one decimal place. The scales required the cohort member's height and age to be entered before the measurements could be taken, so height had to be measured first.

10.4 Weight only

If the cohort member did not want their body fat measurement to be taken or if the body fat measurement could not be taken (e.g., no height measurement was possible), the Tanita™ scales could be operated to take weight only. This was measured in kilograms to one decimal place.

10.5 Consent

Before taking any of the physical measurements, interviewers sought written consent from the parent/guardian (recorded at **CHIC**) and, if they agreed, verbal consent was obtained for each measurement from the young person and confirmed in CAPI (recorded at **CHAC**). After the measurements were taken the interviewer asked the young person if they would like a record of any of their measurements. If so, these were recorded on a measurement card and given to the young person.

10.6 Baseline numbers

Measurement	Male	Female
Height	5,699	5,701
Weight	5,644	5,506
Body fat	5,614	5,486
Missing	247	225
Total	5,946	5,926

10.7 Data format

Data from the physical measurements module is available in the CM measurement file, and the derived variables are held in the CM derived file. Both of these contain one row per cohort member.

10.8 Derived variables

There are two measures of obesity available, based on the two most widely used reference panels – the International Obesity Task Force (IOTF) (Cole *et al*., 2000) and the UK90 (Cole *et al*., 1990). Cut-offs are based on the age of the cohort member at the time of interview and are shown below.

Variable	Description
FCOVWGT6	Overweight cut-off point for CM's age and sex (IOTF thresholds)
FCOBESE6	Obesity cut-off point for CM's age and sex (IOTF thresholds)
FCUNDWU6	Underweight cut-off point for CM's age and sex (UK90 2nd centile)
FCOVWTU6	Overweight cut-off point for CM's age and sex (UK90 85th centile)
FCOBESU6	Obesity cut-off point for CM's age and sex (UK90 95th centile)
FCMCS6AG	Age at interview to nearest 10th of year
FCBMIN6	MCS6 body mass index calculated (CLS)
FCOBFLG6	MCS6 obesity flag - IOTF thresholds
FCUK90O6	MCS6 obesity flag - UK90 thresholds

10.9 Reference cut-offs

The following cut-offs were used for the construction of the IOTF and UK90 derived variables. For the UK90 derivation cut-off points were generated using the LMSGrowth Microsoft Excel add-in software.

IOTF BMI cut-offs

Age	IOTF BMI cut-offs			
	Boys		Girls	
	Overweight	Obese	Overweight	Obese
13	21.910	26.840	22.580	27.760
13.1	21.982	26.922	22.660	27.848
13.2	22.054	27.004	22.740	27.936
13.3	22.126	27.086	22.820	28.024
13.4	22.198	27.168	22.900	28.112
13.5	22.270	27.250	22.980	28.200
13.6	22.340	27.326	23.052	28.274
13.7	22.410	27.402	23.124	28.348
13.8	22.480	27.478	23.196	28.422
13.9	22.550	27.554	23.268	28.496
14	22.620	27.630	23.340	28.570
14.1	22.688	27.700	23.404	28.630
14.2	22.756	27.770	23.468	28.690
14.3	22.824	27.840	23.532	28.750
14.4	22.892	27.910	23.596	28.810
14.5	22.960	27.980	23.660	28.870
14.6	23.026	28.044	23.716	28.918
14.7	23.092	28.108	23.772	28.966
14.8	23.158	28.172	23.828	29.014
14.9	23.224	28.236	23.884	29.062
15	23.290	28.300	23.940	29.110
15.1	23.352	28.360	23.986	29.146
15.2	23.414	28.420	24.032	29.182
15.3	23.476	28.480	24.078	29.218
15.4	23.538	28.540	24.124	29.254
15.5	23.600	28.600	24.170	29.290
15.6	23.660	28.656	24.210	29.318
15.7	23.720	28.712	24.250	29.346
15.8	23.780	28.768	24.290	29.374
15.9	23.840	28.824	24.330	29.402
16	23.900	28.880	24.370	29.430

UK90 BMI cut-offs

Age	UK90 BMI cut-offs					
	Boys			Girls		
	Underweight	Overweight	Obese	Underweight	Overweight	Obese
13	14.774	20.652	22.768	14.939	21.742	24.056
13.1	14.822	20.726	22.850	14.99	21.815	24.135
13.2	14.87	20.799	22.931	15.041	21.889	24.215
13.3	14.918	20.873	23.012	15.091	21.961	24.293
13.4	14.967	20.947	23.012	15.141	22.033	24.371
13.5	15.017	21.021	23.174	15.192	22.104	24.448
13.6	15.066	21.095	23.255	15.241	22.175	24.524
13.7	15.115	21.169	23.335	15.29	22.244	24.599
13.8	15.166	21.243	23.416	15.339	22.313	24.673
13.9	15.215	21.317	23.416	15.387	22.381	24.745
14	15.266	21.391	23.577	15.436	22.449	24.819
14.1	15.317	21.466	23.657	15.484	22.516	24.89
14.2	15.367	21.539	23.737	15.53	22.581	24.96
14.3	15.417	21.613	23.816	15.577	22.646	25.028
14.4	15.468	21.686	23.895	15.623	22.709	25.097
14.5	15.518	21.759	23.973	15.669	22.773	25.164
14.6	15.569	21.833	24.052	15.714	22.835	25.23
14.7	15.619	21.905	24.128	15.758	22.896	25.295
14.8	15.67	21.978	24.207	15.802	22.956	25.359
14.9	15.72	22.049	24.283	15.846	23.016	25.423
15	15.77	22.121	24.359	15.889	23.075	25.485
15.1	15.82	22.193	24.436	15.931	23.133	25.546
15.2	15.87	22.264	24.511	15.973	23.189	25.605
15.3	15.921	22.335	24.587	16.013	23.244	25.664
15.4	15.969	22.404	24.660	16.054	23.3	25.722
15.5	16.02	22.475	24.735	16.094	23.353	25.778
15.6	16.069	22.544	24.808	16.133	23.406	25.835
15.7	16.118	22.613	24.880	16.171	23.457	25.888
15.8	16.167	22.682	24.953	16.21	23.509	25.943
15.9	16.215	22.749	25.023	16.247	23.56	25.996
16	16.264	22.817	25.094	16.284	23.609	26.048

11. Accelerometry

11.1 Background and introduction

Accelerometry is the objective measurement of physical activity using accelerometers, which are electro-mechanical devices that measure acceleration force. Cohort members were asked to wear wrist-worn accelerometers for two days as part of the Age 14 Survey.

This is the second time accelerometry has been included on MCS; cohort members were asked to wear a hip-worn accelerometer as part of the Age 7 Survey.

At age 14, all cohort members in Scotland, Wales and Northern Ireland were asked to wear one, and an 81 per cent random subsample of respondents in England (due to resource constraints). It was deemed important to have sufficient sample sizes to allow for country-specific analysis. Further information about the implementation of accelerometry in the MCS Age 14 Survey can be found in the MCS6 Technical Report.

11.2 Device use

The device used at the Age 14 Survey was the GENEActiv Original ('GENEActiv'), a wrist-worn triaxial accelerometer. Measurement was obtained at 40Hz, chosen after consulting the relevant literature and experts. This means there are 120 data points for every second the device is recording (40 measurements per second on each of three directional axes). The GENEActiv had a battery capable of lasting for the duration of the period in which measurement would be required, and also had sufficient internal memory to record data for that length of time. The device, which was robust and waterproof, did not provide respondents with any feedback during the time they were wearing it.

Further technical information can be found on the GENEActiv website: http://www.geneactiv.org/.

11.3 Subsampling

All cohort members in Wales, Scotland and Northern Ireland were included in the subsample, along with approximately 81 per cent of cohort members in England.

Numbers and proportions of cohort members in each country included in the subsample

	Number	Percent
England	6,290	80.3
Scotland	1,277	100
Wales	1,636	100
Northern Ireland	1,134	100
TOTAL	10,337	86.9

11.4 Days worn

Cohort members were asked to wear the activity monitor for two randomly selected 24-hour periods in the ten days following the visit (with the day of the visit and the following two days ineligible for selection). One selected day was a weekday; the other, a weekend day. The days were selected by the CAPI programme during the interviewer visit. The two selected days were the same as those for the time-use diaries.

11.5 Consent

As cohort members were still minors, parents were gatekeepers to their participation. Information about the activity monitoring task was provided to both the parents and young

people in advance of the interviewer visit in the form of a leaflet, which contained information on why the survey included an activity-monitoring component and briefly explained that physical activity data would be collected using a wrist-worn monitor. Parents were asked to provide written consent for the interviewer to approach their child to participate, and young people gave verbal consent using a structured consent form, which the interviewer read out, and then signed.

11.6 Baseline numbers

Initial numbers on response are available in the MCS6 Technical Report. Further information on valid data extracted from the devices, wear time and activity levels will be available when the data is processed and released to the UK Data Service (later 2017).

12. Time-use diary

12.1 Background and introduction

The time-use diary aimed to collect information about the daily activities of cohort members, as well as contextual information including where they were, who they were with, and how much they liked each activity. Cohort members were asked to complete the time-use diary for two days as part of the Age 14 Survey. This is the first time MCS cohort members have been asked to complete a time-use diary.

The time-use diary task was paired with the accelerometry element, with cohort members wearing the monitor and completing the time-use diary on the same two days. Due to resource constraints, a random subsample of respondents was invited to wear the accelerometer and complete the time-use diary. Information about the subsample can be found in the accelerometry section of this report (section 11).

The development of the MCS time diary instruments was led by the Centre for Longitudinal Studies (CLS) in collaboration with Ipsos MORI and the Centre for Time Use Research (CTUR) at the University of Oxford. CLS oversaw and contributed to all aspects of the development. IM produced the time diary instruments and leaflets and carried out the different testing phases. CTUR contributed significantly to the instrument development, advising on key research design and implementation decisions.

Further information about the development of the time-use diary for the MCS Age 14 Survey can be found in CLS working paper 2015/5, 'Measuring young people's time-use in the UK Millennium Cohort Study: A mixed-mode time diary approach', available at http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=939&sitesectiontitle=CLS+working+paper+series.

12.2 Days completed

Cohort members were asked to complete the time-use diary on the same two randomly selected days on which they wore the activity monitor. These were two randomly selected 24-hour periods in the ten days following the visit (with the day of the visit and the following two days ineligible for selection). One selected day was a weekday; the other, a weekend day. The days were selected by the CAPI programme during the interviewer visit.

12.3 Modes

The time-use diary was available in three modes – web, app and paper. Young people were encouraged to use the web or app diary and were only offered the paper version if they could not, or declined to, complete the diary using one of the electronic methods. The web and paper versions used the conventional light diary time grid, which asked young people to report what they were doing every 10 minutes across the selected 24 hour period. The app used a question-based approach for reporting activities and contextual information. The following table sets out the similarities and differences between the diaries when administered in the three different modes.

Similarities and differences between the time diary modes

	Web	**App**	**Paper**
Approach	Time-grid	Question-based	Time-grid
Time unit	10-minute slot	User-assigned start and end times	10-minute slot
Diary dimensions	Overlap	Coterminous	Overlap

Soft and hard checks	Yes	Yes	No
Aide-memoire	Yes	Yes	No

12.4 Consent

As the cohort members were minors, parents were gatekeepers to their participation. Information about the time-use task was provided to both the parents and young people in advance of the interviewer visit in the form of a leaflet, which contained information on why the survey included the collection of time-use information and briefly explained how the information would be collected. Parents were asked to provide written consent for the interviewer to approach their child to participate, and young people gave verbal consent using a structured consent form, which the interviewer read out, and then signed.

12.5 Baseline numbers

Initial numbers on response are available in the Technical Report. Further information will be available when the data is processed and released to the UK Data Service

13. Saliva

13.1 Background and introduction

Cohort members and resident biological parents were asked to provide a saliva sample as part of the Age 14 Survey. The purpose of this collection was to extract DNA for later genotyping for future research.

Saliva collection has been included in previous sweeps of MCS, though on a smaller scale and for a different purpose. At the Age 3 Survey, a saliva sample was taken from the children to measure exposure to common childhood infections. The saliva was not used for DNA or genetic testing. At the Age 11 Survey, collection of saliva for DNA extraction and storage was piloted, but not carried out at the main stage due to lack of funding.

13.2 Saliva collection protocols

All cohort members who lived with a parent or guardian who had legal parental responsibility were eligible to provide a saliva sample. This is because a person with legal parental responsibility was required to provide written consent for the cohort member to give a sample. In addition, biological parents who lived with the cohort member were eligible to provide a saliva sample.

Saliva was collected using Oragene 500 DNA self-collection kits. Respondents were required to spit into a container until they had provided 2ml of saliva. The interviewer then closed the lid of the container, releasing preservative into the saliva sample. Samples were posted by interviewers to the laboratory appointed to extract and store DNA, at the University of Bristol.

More information about the collection kit used can be found on the DNA Genotek website: http://www.dnagenotek.com/ROW/products/OG500.html

13.3 Consent

Information about the saliva collection task was provided to both the parents and young people in advance of the interviewer visit in the form of a booklet, which contained information about why the study wanted to collect saliva and what it would be used for. Parents were asked to provide written consent to provide their own sample, and parents with legal parental responsibility provided consent for the cohort member to give a sample. Young people gave verbal consent to provide a sample using a structured consent form, which the interviewer read out, and then signed.

13.4 Storage

The saliva samples were sent to a research laboratory at the University of Bristol, where DNA was extracted and is being stored in freezers. Researchers wishing to use the DNA samples will be required to apply to the METADAC. More information about the METADAC and the application process will be made available on the website when data access arrangements have been established: http://www.metadac.ac.uk/.

13.5 Baseline numbers

The table below gives indicative number of samples expected to be available. The actual number of samples will be lower due to losses in DNA extraction. Final figures will be made available once the samples have been fully processed.

Number of saliva samples obtained

	Number
Main respondent	9,584
Partner respondent	5,169
Cohort member	9,726

14. Response and weights at Sweep 6

As with any longitudinal survey, the MCS is subject to attrition. Attrition takes place when respondents drop out of the survey over time. This leads to two problems:

(i) a reduction in sample size, and
(ii) bias in sample composition.

Sample bias arises when the likelihood of dropping out from the survey is correlated with the socio-demographic characteristics of the respondents. In this case, the survey will lose a particular type of respondent (e.g., disadvantaged families and ethnic minorities) and the sample will no longer be representative of the population it was drawn from. However, there are statistical methods to deal with this so as to ensure the remaining sample recovers (under reasonable assumptions) population parameters, which are the topic of this section.

This section examines attrition in sweep 6 of MCS and presents the procedures used in the construction of MCS6 unit non-response weights. For a full description of attrition in previous sweeps, refer to the MCS Technical Report on Response (3rd edition, 2010) and Technical Report on Response in Sweep 5 (2014). For a description of how to use the weights in Stata and SPSS, refer to the respective guides (Stata, SPSS). For a description of the MCS sample refer to the Technical Report on Sampling (4th edition, 2007). All these documents are on the MCS Survey Design pages of the CLS website[7].

14.1 Response in MCS

In Table 1, the proportions of productive and unproductive cases are presented by category. The proportion of productive cases decreased over time from 96.4 per cent in MCS1 to 60.9 per cent in MCS6. The two categories of non-response which have seen a marked increase over time are 'refusal' and 'not issued'. 'Refusals' consist of respondents who declined to take part in a particular sweep of data collection, and 'not issued' are respondents who have not participated in the survey on two consecutive occasions, and therefore were no longer issued for fieldwork (i.e., the survey agency no longer tries to contact them). Non-contact has declined over time because respondents in this category have either been located and contacted again, or have moved to the not issued category. All other types of non-response are relatively stable over time. Note that 'ineligible' includes child deaths, sensitive cases and temporary and permanent emigrants. The category 'untraced movers' refers to respondents who have changed address and were not located, including possible emigrants. Respondents who were not issued in MCS1 are labelled as 'new families'. These were eligible families who were not contacted in MCS1 because their addresses were not know in time for them to be included in the first wave of data collection.

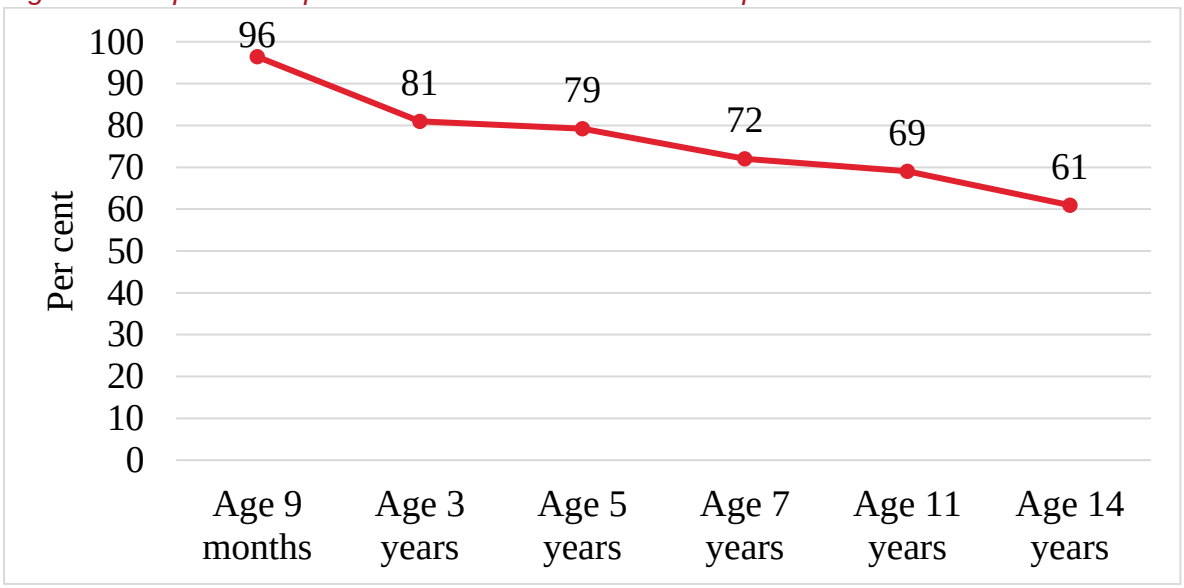
[7] http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=880&sitesectiontitle=Survey+Design

Table 1: Productive and unproductive cases in all MCS sweeps

	MCS1 Age 9 months		MCS2 Age 3 years		MCS3 Age 5 years		MCS4 Age 7 years		MCS5 Age 11 years		MCS6 Age 14 years	
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	%
Productive	18,551	96.4	15,590	81.0	15,246	79.2	13,857	72.0	13,287	69.0	11,726	60.9
Refusal			1,739	9.0	2,315	12.0	1,811	9.4	2,195	11.4	3,029	15.7
Ineligible			167	0.9	300	1.6	126	0.7	78	0.4	45	0.2
Untraced movers			686	3.6	546	2.8	706	3.7	388	2.0	428	2.2
Non-contact			930	4.8	546	2.8	123	0.6	438	2.3	75	0.4
Not issued	692	3.6					2,212	11.5	2,851	14.8	3,828	19.9
Other unproductive			131	0.7	290	1.5	408	2.1	6	0.0	112	0.6
Total	19,243	100	19,243	100	19,243	100	19,243	100	19,243	100	19,243	100

Figure 1 presents the proportion of productive cases in MCS in all sweeps, showing that the sample decreased by 40 per cent by the time of the Age 14 Survey.

Figure 1: Proportion of productive cases in all MCS sweeps



Note: The total number of MCS respondents ever interviewed is 19,243.

We next show how the proportion of productive cases at MCS6 varies along key dimensions. First, Table 2 shows how the MCS6 proportion of productive cases varies by country of sampling. The productive proportion is higher than the UK average in England, while it is lower than the average in Scotland and Northern Ireland.

Table 2: Productive and unproductive cases by country of sampling in MCS6

	England		Wales		Scotland		Northern Ireland	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Productive	7,678	62.8	1,669	60.5	1,263	54.1	1,116	58.0
Refusal	1,853	15.2	455	16.5	401	17.2	320	16.6
Ineligible	35	0.3	2	0.1	6	0.3	2	0.1
Untraced	199	1.6	92	3.3	101	4.3	36	1.9
Non-contact	60	0.5	6	0.2	6	0.3	3	0.2
Not issued	2,335	19.1	503	18.2	548	23.5	442	23.0
Other unproductive	64	0.5	33	1.2	11	0.5	4	0.2
Total	12,224	100	2,760	100	2,336	100	1,923	100

Sample size=19,243. Percentages may not sum to 100 due to rounding.

Table 3 shows that the proportion of productive cases varies across sampling strata in each country. Respondents sampled from the socially advantaged stratum are more likely to be productive in all four countries, compared to those sampled from the disadvantaged stratum. Respondents sampled from the ethnic minority stratum are less likely to be productive than those in the advantaged stratum in England.

Table 3: Proportion of productive cases by stratum in MCS6

	England			Wales		Scotland		Northern Ireland	
	Adv.	Dis.	Ethn.	Adv.	Dis.	Adv.	Dis.	Adv.	Dis.
Productive	67.1	59.8	60.3	65.1	58.5	60.1	48.3	63.4	54.8
Unproductive	32.9	40.2	39.7	34.9	41.6	39.9	51.7	36.7	45.2

Note: Adv. stands for advantaged stratum. Dis. stands for disadvantaged stratum and Ethn. stands for ethnic minority stratum. Sample size=19,243.

In Table 4 we look at different response patterns. Table 4 shows that 47.2 per cent of all respondents participated in all six sweeps of MCS. In contrast, 22.1 per cent have interrupted response patterns (i.e., non-monotone response). In other words, they participated in a number of sweeps, and then dropped out before participating again in subsequent sweeps. 30.7 per cent of all respondents have monotone response patterns. That is, they participated in a number of sweeps before dropping out for all subsequent sweeps.

Table 4: Monotone vs. non-monotone responses in all MCS sweeps

Patterns	Freq.	Percent
Monotone	5,908	30.7
Non-monotone	4,247	22.1
All waves	9,088	47.2
Total	19,243	100

Table 5 shows the percentages of respondents participating in n sweeps (n=1…6). We see that 64.1 per cent of respondents participated in at least five out of six sweeps of MCS, which indicates that more than half of the sample have almost complete records.

Table 5: Number of times productive up to MCS6

Times productive	Freq.	Percent
One	1,932	10.0
Two	1,391	7.2
Three	1,559	8.1
Four	2,027	10.5
Five	3,246	16.9
Six	9,088	47.2
Total	19,243	100

Note: Percentages may not sum to 100 due to rounding.

14.2 Predicting response at MCS6

The procedure used for predicting responses at Sweep 5 was also used at Sweep 6 [8]. We estimate a logit model in which the dependent variable is binary (=1 for response and 0 otherwise). The predictors are:

1. The cohort member's gender
2. Mother's age at first live birth
3. The cohort member's ethnic group
4. Housing tenure in MCS5
5. Accommodation type in MCS5
6. The main respondent's highest educational qualification in all sweeps
7. Whether the cohort member was breastfed
8. Number of parents living in the household in MCS5
9. The main respondent's highest social and economic status in all sweeps
10. Ratio of number of times not answering the income question divided by the number of sweeps productive
11. Ratio of number of times reporting having a job divided by the number of times productive
12. Whether the household is a 'new' family. There were 701 children who joined the survey in Sweep 2 because their addresses were not known in Sweep 1; therefore they did not take part in the first sweep. These children and their families were labelled as 'new families'.

Missing data for predictor variables due to non-monotone non-response or item missingness were imputed using simple and multiple imputations, as described in Appendix A.

Table A1 in the Appendix shows the odds ratios of the response logit model estimated using the 25 imputed datasets. The linear predicted values were generated from this model; then the predicted values were converted into predicted probabilities using an inverse logit transformation. The non-response weights for Sweep 6 were constructed as the inverse of the predicted probabilities [Wooldridge 2007]. Two overall weights were constructed by multiplying the sampling weights in Sweep 1 by the attrition weights in each sweep of MCS (i.e., 2 to 6). The weights were scaled to make their total equal to the productive sample size. We note that the effectiveness of the response weights to correct for bias depends on the inclusion of all

[8] http://eprints.ioe.ac.uk/18759/1/Technical_Report_on_Response_in_Sweep5_for_web_TM.pdf

important predictors of unit non-response in the logit response model (Table A1) (Seaman and White 2013).

Weights variable names:

FOVWT1: Sweep 6 overall weight for single country analysis

FOVWT2: Sweep 6 overall weight for whole of UK analysis.

In Tables A2 and A3 of the Appendix, the means, minima and maxima of the two weights are presented by stratum.

For a description of how to use the weights in Stata and SPSS, refer to the respective guide: Stata, SPSS. Available on the MCS Survey Design pages of the CLS website [9].

[9] http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=880&sitesectiontitle=Survey+Design

15. Available datasets

There are two sets of 12 data files (one set of SPSS data files and one of STATA data files):

1) *The cross-sectional data from the household questionnaire, main, partner and proxy interviews:*
   - MCS6 parent interview data
   - MCS6 parent interview unfolding brackets data (in preparation)
   - MCS6 proxy parent interview data
   - MCS6 proxy parent interview unfolding brackets data (in preparation)

   This file contains one row per respondent.

2) *Cross-sectional household grid data:*
   - MCS6 Household Grid

   These files contain one row for each person (ever) in the household grid in productive families.

3) *Child questionnaire files*
   - MCS6 child interview data
   - MCS6 parent CM interview data

4) *Child assessment and measurement files:*
   - MCS6 child assessment data
   - MCS6 child measurement data
   - MCS6 accelerometer data (in preparation)
   - MCS6 time-use diary (in preparation)

5) Parental assessment (word activity)
   - MCS6 parent assessment

6) *Geographically linked data (IMD and rural/urban indicators):*
   - MCS6 IMDRU England
   - MCS6 IMDRU Wales
   - MCS6 IMDRU Scotland
   - MCS6 IMDRU Northern Ireland

7) *Derived variables:*
   - MCS6 family level derived variables.

   These files contain one row for each productive family at that sweep.
   - MCS6 parent derived variables
   - MCS6 CM derived variables

These files contain one row for each productive main, partner or CM respondent at that sweep.

Data can be downloaded from the UK Data Service at
https://discover.ukdataservice.ac.uk/series/?sn=2000031

16. Accompanying documents

16.1 Questionnaires

MCS6 CAPI questionnaire
MCS6 Young person questionnaire
MCS6 Physical measurement activity monitor time use
MCS6 Age 14 survey SDQ

16.2 Technical Reports

MCS6 Technical Report
MCS6 Technical Report Appendix A Pre-notification and briefing docs
MCS6 Technical Report Appendix B Household materials
MCS6 Technical Report Appendix C Welsh and ethnic minority languages
MCS6 Technical Report Appendix D

16.3 Guides

MCS6 Time-Use diary documentation
MCS6 Technical Report on response
User Guide to analysing MCS data using Stata
User Guide to analysing MCS data using SPSS
MCS6 Derived Variables User Guide

17. Appendix A: Response and weights

Imputations

Multiple imputations were carried out using the MI command in Stata 13. As a result of the use of simple and multiple imputations, the sample used in the logit response model consisted of 15,415 observations (i.e., the issued sample in MCS6). Weights were constructed for all respondents in MCS6. Imputations were carried out as described in Appendix I.

Replacement of missing values

Since ethnicity is a fixed attribute over time and the main respondent's highest educational qualification is unlikely to change from one sweep to the other, we replaced the missing values on these two variables in MCS6 using the most recent available information from previous sweeps. Mother's age at first live birth was missing for only 60 cases. These were replaced by the average of non-missing cases.

Multiple imputations

Four variables were imputed using multiple imputation with chained equations in Stata 13. The imputation was carried out for item missingness and missingness caused by non-monotone response patterns. Accommodation type had the largest proportion missing with 22 per cent. We generated 25 multiple imputations. The number of parents in the household and housing tenure were missing for 2,177 respondents in the analytical sample (i.e., issued sample in MCS6). Whether the cohort member was breastfed was missing for 612 respondents – mostly new families who joined the survey in Sweep 2 – and accommodation type was missing for 3,435 respondents.

Breastfeeding and type of accommodation were imputed using the following variables as predictors of item specific missingness: highest educational qualification, whether family is a new family, cohort member's gender, cohort member's ethnic group, the main respondent's highest social and economic status, and sampling stratum.

The number of parents living in the household was imputed using highest educational qualification, whether family is a new family, cohort member's gender, cohort member's ethnic group, the main respondent's highest social and economic status, sampling stratum, in addition to the number of parents in the household in MCS1 and MCS4.

Housing tenure was imputed using highest educational qualification, whether family is a new family, cohort member's gender, cohort member's ethnic group, the main respondent's highest social and economic status, and sampling stratum, in addition to housing tenure in all previous sweeps.

We note that multiple imputation returns valid estimates assuming the data are Missing at Random (MAR) (Enders (2010); Seaman *et al.*, (2013); Sterne *et al.*, (2009). This implies that any differences between the missing values and the observed values can be explained by the variables that were included in the imputation models. Put differently, conditional on the variables in the imputation model, missingness is not due to unobserved or observed variables not included in the model.

No imputation required

The ratio of number of times the income question was not answered divided by the number of productive sweeps did not have any missing data since it was constructed using non-missing observations from all sweeps. The same applies for the ratio of number of times the informant reported having a job divided by the number of times productive. The main respondent's highest social and economic status was constructed as the maximum of social and economic status reported in each sweep.

Finally, some variables, such as cohort member's gender and whether the household is a new family, did not have any missing values and therefore did not require any imputation.

Table A1: Logit response model

Explanatory variables	Odds ratio	SE	T	P>t	[95% Conf.	Interval]
New family	1.03	0.11	0.28	0.78	0.84	1.26
Cohort member is a boy	0.87	0.03	-3.49	0.00	0.81	0.94
Cohort member's ethnic group (Reference: White)						
Mixed	0.85	0.10	-1.42	0.15	0.68	1.06
Indian	1.35	0.19	2.17	0.03	1.03	1.78
Pakistani, Bangladeshi	2.01	0.19	7.31	0.00	1.67	2.43
Black	0.79	0.09	-2.19	0.03	0.64	0.97
Other, NA, not known, refusal	1.22	0.17	1.43	0.15	0.93	1.61
Main respondent's highest social and economic status (Reference: Managerial and Professional)						
Intermediate	0.83	0.05	-2.87	0.00	0.74	0.94
Small employers and self-employed	0.77	0.07	-2.65	0.01	0.64	0.94
Lower supervisory and technical	0.66	0.06	-4.47	0.00	0.54	0.79
Semi-routine and routine	0.52	0.03	-9.81	0.00	0.46	0.59
NA	0.45	0.05	-7.96	0.00	0.37	0.55
Highest educational qualification (Reference: NVQ level 1)						
NVQ level 2	1.00	0.08	-0.05	0.96	0.85	1.16
NVQ level 3	1.11	0.10	1.21	0.23	0.94	1.32
NVQ level 4	1.35	0.12	3.46	0.00	1.14	1.60
NVQ level 5	1.84	0.22	5.16	0.00	1.46	2.33
Overseas qualifications only	1.22	0.16	1.47	0.14	0.94	1.58
None of these	0.80	0.07	-2.60	0.01	0.67	0.95
Number of parents in household (reference: One parent)						
Two parents/carers	1.20	0.06	3.70	0.00	1.09	1.32
Child was breastfed at least once	1.23	0.06	4.49	0.00	1.12	1.35
Accommodation type (Reference: Other)						
House or bungalow	1.14	0.11	1.39	0.17	0.95	1.37
Housing tenure (Reference: Own outright)						
Own – mortgage/loan	1.09	0.10	0.94	0.35	0.91	1.32
Rent from local authority	0.73	0.08	-2.96	0.00	0.59	0.90
Rent from housing association	0.72	0.08	-2.87	0.00	0.57	0.90
Rent privately	0.74	0.08	-2.79	0.01	0.60	0.91
Other (rent free, living with parents)	0.67	0.09	-3.02	0.00	0.51	0.87

Ratio income item non-response	0.26	0.03	-11.43	0.00	0.20	0.32
Ratio times having a job	0.33	0.03	-14.08	0.00	0.29	0.39
Mother's age at first birth	1.05	0.00	10.68	0.00	1.04	1.06
Constant	1.81	0.36	2.96	0.00	1.22	2.69
N			15,415			

Note: The analytical sample in table 6 includes all issued cases in MCS6 with 15,415 observations.

Table A2: FOVWT1, Sweep 6 overall weight for single country analysis

Sampling stratum	N	Mean	Std. Dev.	Min	Max
England – advantaged	3,240	1.24	0.66	0.68	8.96
England – disadvantaged	2,876	1.00	0.68	0.38	10.90
England – ethnic	1,562	0.42	0.30	0.14	4.28
Wales – advantaged	542	1.54	0.87	0.79	10.80
Wales – disadvantaged	1,127	0.79	0.49	0.32	6.93
Scotland – advantaged	688	1.06	0.79	0.32	7.95
Scotland – disadvantaged	575	0.95	0.81	0.20	9.23
Northern Ireland – advantaged	458	1.25	0.86	0.34	6.07
Northern Ireland – disadvantaged	658	1.03	0.88	0.20	8.45
All strata	11,726	1.01	0.73	0.14	10.90

Table A3: FOVWT2, Sweep 6 overall weight for whole of the UK analysis

Sampling stratum	N	Mean	Std. Dev.	Min	Max
England – advantaged	3,240	1.60	0.88	0.86	12.47
England – disadvantaged	2,876	1.31	0.91	0.49	14.03
England – ethnic	1,562	0.56	0.40	0.18	5.53
Wales – advantaged	542	0.52	0.29	0.27	3.48
Wales – disadvantaged	1,127	0.27	0.16	0.11	2.37
Scotland – advantaged	688	0.82	0.59	0.25	6.26
Scotland – disadvantaged	575	0.74	0.63	0.16	7.30
Northern Ireland – advantaged	458	0.49	0.32	0.14	2.43
Northern Ireland – disadvantaged	658	0.40	0.33	0.08	3.33
All strata	11,726	1.01	0.86	0.08	14.03

18. References

Calderwood, L., Smith, K., Gilbert, E., Rainsberry, M., Knibbs, S., and Burston, K. (Winter 2015). Securing participation and getting accurate answers from teenage children in surveys: Lessons from the UK Millennium Cohort Study. *Social Research Practice*, *1*.

Cole T.J., Bellizzi M.C., Flegal K.M., Dietz W.H. (2000). Establishing a standard definition for child overweight and obesity worldwide: International survey. *BMJ, 320*:1240.

Cole T.J., Freeman J.V., Preece M.A. (1995). Body mass index reference curves for the UK, 1990. *Archives of disease in childhood, 73*(1): 25-9.

Conroy, R.M. (2005). Stings in the tails: Detecting and dealing with censored data. *Stata Journal, 5*, 395-404.

DWP (2016). *Households below average income*. 1994/95-2014/15. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/532416/households-below-average-income-1994-1995-2014-2015.pdf

Enders, C. K. (2010). *Applied missing data analysis*, Guilford Press.

Gilbert, E., Calderwood, L., Rainsberry, M., Knibbs, S., and Burston, K., (2015). 'Tweets, branding and swag: Engaging teenagers in research'. European Survey Research Association Conference, Reykjavik.

Goodman, R (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581-6; Psychological and Development Inventories.

Hansen, K. (2008). Millennium Cohort Study First, Second and Third Surveys: A Guide to the Datasets, Third Edition. Retrieved from www.cls.ioe.ac.uk/shared/get-file.ashx?id=598&itemtype=document

HBAI Team, Information Directorate, Department for Work and Pensions (2005). Households below average income statistics: Adoption of new equivalence scales. (http://www.dwp.gov.uk/asd/hbai/nsfr_newequiv.pdf;). Accessed 22 May 2008.

Johnson, J., Atkinson, M., Rosenberg, R. (2015). 'Millennium Cohort Study: Psychological, developmental and health inventories'. (Jan 2015), CLS User Guide to the Data.

Kessler, R.C., Barker, P.R., Hirpi, E., Howes, M. J, Normand, S-L. T., Manderscheid, R.W., Walters, E.E., Zaslavsky, A.M (2003) Screening for serious mental illness in the general population, Archives of General Psychiatry 60(2) 184-9.

Pan, H. and Cole, T.J. (2012). LMSgrowth, a Microsoft Excel add-in to access growth references based on the LMS method. Version 2.77. http://www.healthforallchildren.co.uk/

Plewis, I. (Ed). (2007). Millennium Cohort Study First Survey: Technical report on sampling (4th edition). London: Centre for Longitudinal Studies, Institute of Education.

Seaman, S., Galati, J., Jackson, D. and Carlin, J. (2013). What is meant by 'missing at random'? *Statistical Science*, *28*, 257-68.

Seaman, S. R. and White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*, *22*, 278-95.

Smith, K., Calderwood, L., Knibbs, S., and Burston, K. (2015). 'Let's talk about sex: asking 14-year-olds about their lives in a home setting.' European Survey Research Association Conference, Reykjavik.

StataCorp. (2007). Stata Statistical Software: Release 10. College Station, TX: StataCorp Lp.

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M. and Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338.

Stewart, M.B. (1983). On least-squares estimation when the dependent variable is grouped. *Review of Economic Studies*, *50*, 737-53.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, *141*, 1281-1301.