Institute of Education

Millennium Cohort Study

User Guide (Surveys 1 – 5)

(9th Edition)

August 2020

CENTRE FOR
LONGITUDINAL
STUDIES

UKRI

Economic
and Social
Research Council

The Centre for Longitudinal Studies (CLS) is an ESRC Resource Centre based at the UCL Institution of Education. It provides support and facilities for those using the three internationally-renowned birth cohort studies: the National Child Development Study (1958), the 1970 British Cohort Study and the Millennium Cohort Study (2000). CLS conducts research using the birth cohort study data, with a special interest in family life and parenting, family economics, youth life course transitions and basic skills. The views expressed in this work are those of the author(s) (amend as necessary) and do not necessarily reflect the views of the Economic and Social Research Council. All errors and omissions remain those of the author(s).

This document is available in alternative formats.

Please contact the Centre for Longitudinal Studies.

tel: +44 (0)20 7612 6875

email: clsfeedback@ucl.ac.uk

Data queries: help@ukdataservice.ac.uk

Contents

Acknowledgements

Abdullahi Baba Ari	Assistant Data Manager
Kate Smith	MCS Survey Manager
Dr Alice Sullivan	Research Fellow
Mina Thompson	Administrator
Kelly Ward	Research Officer

We are also deeply grateful to the late Prof. Neville Butler who made an invaluable contribution to the Millennium Cohort Study until his final year (2007).

Overview of document

This document provides an overview of the MCS data from the first to the fifth sweep. The document is laid out as follows:

1. Part One provides an introduction to the study.
2. Part Two explains the sample, the achieved samples at different sweeps and related issues of weighting.
3. Part Three provides information about the survey development.
4. Part Four discusses the survey content.
5. Part Five explains the fieldwork.
6. Part Six provides an overview of characteristics of the MCS data
7. Part Seven provides information about the data of sweeps 1-5
8. Part Eight discusses ethical considerations
9. Part Nine provides information on the release of the data of sweeps 1-4 in a long format (Summer 2020)

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 .

PART 1: BACKGROUND

A renewed interest in child wellbeing in the late 1990s in the UK provided the context for the development of a new and distinctive child cohort study, after a gap of 30 years (since the 1970 British Cohort Study, the 1958 National Child Development Study and the 1946 National Survey of Health and Development). The Millennium Cohort Study (MCS) was developed as a multidisciplinary survey which could capture the influence of early family context on child development and outcomes throughout childhood, into adolescence and subsequently through adulthood. MCS's field of enquiry covers such diverse topics as parenting; childcare; school choice; child behaviour and cognitive development; child and parental health; parents' employment and education; income and poverty; housing, neighbourhood and residential mobility; and social capital and ethnicity.

To date there have been 7 surveys: the first (MCS1) when the children were around 9 months old, the second (MCS2) when the children were 3 years of age, the third (MCS3) when they were 5, the fourth (MCS4) when they were 7 and mostly recently the fifth (MCS5) returned when the children were 11 years old.

Sweep	Year of data collection	Cohort members' age
MCS 1	2001/2	9 months old
MCS 2	2003/4	3 years old
MCS 3	2005/6	5 years old
MCS 4	2008/9	7 years old
MCS 5	2012/3	11 years old
MCS 6	2015/6	14 years old
MCS 7	2018/9	17 years old

4

PART 2: SAMPLING

For a more comprehensive discussion of the sampling procedure used, please refer to the MCS Technical Report on Sampling (4th Edition) (Plewis 2007).

2.1. Sample Design of MCS1

Unlike its predecessor studies which followed the same survey design: (a systematic random sample of all children born in a particular week) the MCS had a new sample design. Firstly, the sample (fully discussed in Plewis, 2007) is drawn from a population of children born between 1 September 2000 and 31 August 2001 (for England and Wales), and between 24 November 2000 and 11 January 2002 (for Scotland and Northern Ireland) who were living in the UK at nine months of age and whose families were eligible to receive Child Benefit at that age. Sampling births across a 16 month period rather than a particular week not only makes it easier for fieldwork agencies by spreading interviews across a longer, less intense period but also means the MCS is well-placed to identify any season-of-birth effects, which have been shown to be important in other studies in a range of outcomes including academic achievement.

Secondly, the MCS is the first British birth cohort to include all four countries of the UK, reflecting increasing moves towards devolution, and allowing for the first time, researchers to not only look at relationships within each country but also make comparisons between the countries.

Thirdly, the MCS oversampled children from deprived backgrounds, so that the effects of disadvantage on children's outcomes could be better addressed. Fourthly, and finally, the MCS set out to reflect the increasing diversity of the UK, and clear evidence of differential health, educational and social outcomes across ethnic groups, by oversampling from areas of relatively high ethnic minority concentration.

To take account of this design the population was stratified. In England, the population was stratified into three strata: The first an *'ethnic minority'* stratum where the proportion of ethnic minorities in that ward in the 1991 Census was at least 30 per cent. The second, a *'disadvantaged'* stratum is 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 using the Child Poverty Index for England and Wales.[1] And finally, an *'advantaged'* stratum which captured children living in wards other than those above.

For Wales, Scotland and Northern Ireland, due to the low percentages of ethnic minority groups (around 1% of the population) (Plewis, 2007) there were only two

---

[1]For more information on the CPI and Indices of Deprivation in general see: https://www.gov.uk/government/publications/english-indices-of-deprivation-2010

strata: a *'disadvantaged'* stratum: children living in wards (known as Electoral Divisions in Wales) that fell into the poorest 25 per cent of wards using the Child Poverty Index. And an *'advantaged'* stratum: 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 used for stratification purposes are area-level measures. That means the design will be good at identifying those who are disadvantaged or from an ethnic minority background groups for those people who live in areas with others from a similar background but is less likely to find people who are equally 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 37per cent of disadvantaged families with a child under 16 were living in advantaged wards, 54 per cent were in disadvantaged wards and 10 per cent in ethnic minority ward (Plewis, 2007).

The sample is clustered by characteristics of electoral wards. Clustering is efficient as it is cheaper to draw a cluster sample of specific areas rather than sample the whole UK, it also keeps fieldwork costs down because it enables interviewer workloads to be concentrated and therefore travel costs are reduced. Moreover, from an analysis perspective clustering also brings the local neighbourhood context into the analysis as having multiple respondents in the same areas allows researchers to look at 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 of the same size.

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

Once the sample wards were selected, a list of all children turning nine months old during the 16 month survey window and living in those wards was generated from the Child Benefit register provided by the Department of Social Security (DSS), subsequently the Department for Work and Pensions (DWP). At that time, Child Benefit was a universal provision, payable (usually to the mother) from the child's birth.[2] 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

---

[2] Child Benefit claims cover virtually all of the child population except those ineligible due to recent or temporary immigrant status.

fraud within the family.[3] Also if families had already taken part in the DWP, Families and Children Survey (FACS) they were excluded from the sample.[4]

It was recognized that the Child Benefit records would not reveal all families who had moved into the sample wards as the child approached 9 months of age so for this reason Health Visitors were approached to find these families. It was thought that as local community Health Visitors would be aware of families transferring into areas. They were asked to see if families meeting the eligibility criteria who had recently moved into survey wards were willing to be recruited. Health Visitors reported 220 cohort families moving into the selected areas with children over 6 months of age, however only 56 had not been found by DWP.[5][6]

2.2. The achieved sample at MCS1

The MCS1 survey reached 18,552 families, which, after allowance for 246 sets of twins and 10 sets of triplets, amounted to 18,818 cohort children. Six families have 2 singletons in the sample. The table below shows how these respondents are distributed over the 4 countries of the UK. Further details by stratum appear in the Technical Report on Sampling (4th Edition) (Plewis 2007).

---

[3] This represents less than 3 per cent of cases (Hansen, 2012).
[4] This affected only 40 cases.
[5] There were several problems which may explain the rather disappointing result of this exercise. First, helping with the survey was not part of the Health Visitors' already demanding normal duties. Second, Health Visitors' caseloads do not neatly coincide with electoral wards. Third, there is no central list of Health Visitors for easy contact.
[6] DWP also discovered 1,389 new families in England who were living in the sample wards at sweep 1, but their addresses reached DWP too late to be included in the first survey so they were added to the sample at sweep 2.

Table 1: MCS1 Sample Size – Clusters, Children Families, by Country

	Number of sample 'wards'	Target sample as boosted	Achieved Responses **	
			Children	Families interviewed
Total UK	398	20,646	18,818	18,552
England	200	13,146	11,695	11,533
Wales	73	3,000	2,798	2,760
Scotland	62	2,500	2,370	2,336
N. Ireland	63	2,000	1,955	1,923

* Counting amalgamations into 'superwards' as a single unit.

** All productive contacts.

2.2.1. Response Rates

The overall response can be thought of as the combination of the leakages to numbers between the eligible population in the selected wards and the sample issued to field and the success the interviewers then have at securing interviews, from the issued sample. The leakages between the eligible population and the 'in scope' population are set out in the Technical Report on Sampling (4th Edition) (Plewis 2007) as:

- Families opting out of the survey
- Families excluded by DWP
- Families excluded from the sampling frame because their postcode could not be allocated to a ward
- Undetected in-migrants.

It is arguable that the eligible population should also include families who do not claim Child Benefit; but we make the simplifying assumption that numbers of such families who are permanently resident in the UK are negligible. The Technical Report on Sampling (4th Edition) (Plewis 2007) makes two alternative assumptions about how many undetected in-movers there are. The estimates quoted in Table 2 below assume that there is an undetected in-comer for every detected out-mover, on average, in each stratum.

The alternative estimate sets undetected in-moves to zero, which raises all overall response rates (except Northern Ireland) above the target or assumed response rate set in the design of the survey and shown in the first column. This table shows that, when undetected in-migration is counted as a leakage, the overall response rate is

68 per cent for the (unweighted) UK sample, modestly below the 71 per cent expected. It is below target in every stratum except the advantaged wards of Wales, but only markedly so in Northern Ireland, with overall response rates in the combined strata of 63 per cent where 71 per cent had been somewhat optimistically set, given the lack of a tradition of such surveys in that country. Northern Ireland is also the only country where inability to assign Child Benefit claimants to a ward was a significant problem. Survey work in the ethnic areas of England was also something of an unknown quantity. A cautious target of 65 per cent was missed by 3 percentage points.

Table 2: Response Rates by Stratum and Country for MCS1

		Expected Overall Response Rate	Achieved Overall Response Rate	In-scope Response Rate Fieldwork
England	Advantaged	75%	73%	86%
	Disadvantaged	70%	68%	82%
	Ethnic	65%	62%	76%
	Total	70%	68%	82%
Wales	Advantaged	75%	78%	89%
	Disadvantaged	70%	69%	83%
	Total	71%	72%	84%
Scotland	Advantaged	75%	73%	86%
	Disadvantaged	70%	68%	83%
	Total	71%	70%	85%
N. Ireland	Advantaged	75%	65%	81%
	Disadvantaged	70%	61%	78%
	Total	71%	63%	79%
UK	All	71%	68%	82%

Source: MCS Technical Report on Sampling (4th Edition) (Plewis 2007).

Out of the cases issued to field some have been deemed ineligible because they are known or thought to have moved out of the survey area before the child reached 9

months of age. Of the remaining eligible or 'in-scope' sample, the response in fieldwork averaged 82 per cent giving at least one interview. It varied by stratum as expected, but more so. The ethnic wards as anticipated had least 76 per cent, and the advantaged areas of Wales the highest 89 per cent, with both strata in Northern Ireland being below the stratum average for Great Britain.

Complete evaluation of sources of the known characteristics in case of survey loss before and after the Child Benefit stage are included in a later edition of the Technical Report on Sampling (4th Edition) (Plewis 2007). Those lost before issue to field do not appear systematically biased. A greater propensity of families in the disadvantaged areas to be excluded by DWP is balanced by a greater propensity of the inhabitants of advantaged areas to opt out.

2.3. Sampling at MCS2

The survey attempted to follow all the 18,553 families who took part in MCS1 where the child was still alive and living in the UK. It also attempted to make contact with another 1,389 'New Families' in England who appeared to have been living in sample wards at the time of MCS1, but whose addresses reached DWP records too late to be included in the first survey.

2.4. The MCS2 Achieved Sample

There are two components to the MCS2 issued sample, families that were productive in MCS1 and the so-called new families. There were 18,552 productive families in the first survey of the Millennium Cohort Study. The new families were families that although eligible, did not participate in MCS1. These were identified through DWP, of whom 1,389 families were eligible to be issued for MCS2 fieldwork.

From the paragraph above, the issued sample should have been 19,941 i.e. 18,552 +1,389, but 71 families from the MCS1 productive families were not issued to the field for various reasons. Their outcomes were known and recorded before the start of the fieldwork. Therefore, the MCS2 issued sample was 19,870; 18481 were the productive families in MCS1 and the 1,389 new families.

MCS2 response is reported in three groups: 1) all families initially thought to be eligible for MCS2 survey; 2) families that were productive in MCS1; and: 3) the New Families.

All response frequencies in this report are unweighted. The outcome codes were derived as:

Productive	All families with some data from one of 6 data collection instruments other than what was carried forward. The 6 data collection instruments were: Main Interview, Partner Interview, Proxy Partner Interview, BAS, Bracken, Height and Weight.
Ineligible	Emigrant families, deaths.
Uncertain Eligibility	Families that were away temporarily and those whose eligibility was uncertain, including untraced movers.
Unproductive	Refusals, non-contacts, other non-responses, including language problems, ill/incapacitated, deleted/lost data (lost CAPI).

2.4.1. All MCS2 Families Response

There were 19,941 families originally considered eligible for MCS2 survey, 15,590 of these were productive in the survey which is 78 per cent of all MCS2 families. There were 15,808 cohort members in the 15,590 productive families.

Table 3: MCS2 Overall response

Outcome code		N	(%)
Productive		15,590	78.0
Ineligible*		255	1.3
Uncertain Eligibility (including untraced movers)		868	4.4
Unproductive	**Refusal**	2,002	10.0
	Non-Contact	1070	5.4
	Other	156	0.8
Total		19,941	100.0

* Ineligible: deaths (n=16), permanent emigrants (n=169), failed eligibility (n=70).

Table 4 below shows that a slightly higher proportion of those that were productive in MCS1 (80 per cent) took part in MCS2 compared to the overall proportion in Table 3, which also includes New Families.

Table 4: MCS2 Overall response for families that were productive in MCS1

Outcome code		n	(%)
Productive		14,898	80.0
Ineligible*		167	0.9
Uncertain Eligibility (including untraced movers)		687	3.7
Unproductive	Refusals	17,39	9.4
	Non-Contact	930	5.0
	Other	131	0.71
Total		18,552	100.0

** **Ineligible:** deaths (n=14), permanent emigrants (n=153).*

2.4.2. The New Families Response

Only about 50 per cent of the New Families were productive.

Table 5: Overall Response for the New Families

Outcome code		N	(%)
Productive		692	50.0
Ineligible **		88	6.3
Uncertain Eligibility (including untraced movers)		181	13.0
Unproductive	Refusals	263	19.0
	Non-Contact	140	10.0
	Other	25	1.8
Total		1,389	100.0

*** **Ineligible:** deaths (n=2), permanent emigrants (n=16), failed eligibility (n=70).*

2.5. Sampling at MCS3

The sample issued for MCS3 comprised all those who had responded to the survey at least once, i.e. to MCS1 or the 631 additional cases who had responded to MCS2 in the New Families, less those known to have become ineligible through the death or emigration of the cohort child, and also less those deemed to have made a permanent refusal (also excluding the one case in the original MCS1 total of 18,533 subsequently discovered to have been invalid). Thus nearly all non-respondents to the second survey who had been interviewed in the first survey were given the opportunity to rejoin the survey at age 5. The non-respondents to the New Families sample were not reissued.

2.6. The MCS3 Achieved Sample

There were 19,244 families potentially eligible for inclusion in the issued sample. These were 18,552 families who were productive at MCS1 and 692 'New Families' who were productive at MCS2. However, 718 families were not issued to the field due to ineligibility (death or emigration), permanent refusal and sensitive family circumstances. Their outcomes were known and recorded before the start of the fieldwork. The families not issued due to sensitive family circumstances are recorded as 'unproductive other'. Two families who were not productive at either MCS1 or MCS2 were issued in error.

Therefore, the MCS3 issued sample was 18,528 (19,244 – (718 + 2)).

This section provides MCS3 response for the 19,244 families, i.e. including the 718 families not issued and excluding the two families issued in error.

All response frequencies here are unweighted. The outcome codes in this report were again derived as for MCS2. There were 19,244 families potentially eligible for the MCS3 survey, 15,246 of these were productive in the survey, which is 79.2 per cent of all MCS3 families. There were 15,459 cohort children in the 15,246 productive families.

Table 6: MCS3 Overall response

Outcome code	N	(%)
Productive	15,246	79.2
Ineligible **	300	1.6
Uncertain Eligibility (including untraced movers)	547	2.8

Unproductive	Refusals	2,798	14.5
	Non-Contact	63	0.3
	Other	290	1.5
Total		19,244	100.0

** *Ineligible:* deaths (n=18), permanent emigrants (n=282).

2.7. The MCS4 Sample

There were 19,244 families potentially eligible for inclusion in the issued sample. However, 2213 families were not issued to the field due to ineligibility due to death or emigration (n=362), permanent refusal (n=1,705), permanent untraced (n=136) and sensitive family circumstances (n=10). Their outcomes were known and recorded before the start of the fieldwork. The families not issued due to sensitive family circumstances are recorded as 'unproductive other'.

Therefore, the MCS4 issued sample was 17,031 (19,244 - 2213).

This section provides MCS4 response for the 19,244 families, i.e. including the 2213 families not issued. All response frequencies in this report are unweighted.

Outcome codes are:

Productive	All families with some data from one of five data collections instruments other than what was carried forward. The 5 data collection instruments were: Main Interview, Partner Interview, Proxy Partner Interview, Cohort Child Cognitive Assessments and Cohort Child Physical Measurements.
Ineligible	Emigrant families, deaths.
Uncertain Eligibility	Families that were away temporarily and those whose eligibility was uncertain, including untraced movers.
Unproductive	Refusals, non-contacts, other non-responses, including language problems, ill/incapacitated, deleted/lost data (lost CAPI).

In total, 13,857 families were productive in the survey, which is 72.20 per cent of all MCS families. There were 14,043 cohort children in the 13,857 productive families.

Table 7: MCS4 Overall response

Outcome code		n	(%)
Productive		13857	72.0
Ineligible		488	2.5
Uncertain Eligibility (including untraced movers)		848	4.4
Unproductive	Refusals	3,516	18.3
	Non-Contact	149	0.8
	Other	386	2.0
Total		19,244	100.0

2.8. The MCS5 Sample

There were 19,244 families potentially eligible for inclusion in the issued sample at MCS5. However, 2,851 families were not issued to the field due to ineligibility due to death or emigration (n=545), permanent refusal (n=2,215), permanent untraceability (n=86) and sensitive family circumstances (n=5). Their outcomes were known and recorded before the start of the fieldwork. The families not issued due to sensitive family circumstances are recorded as 'unproductive other'.

Therefore, the MCS5 issued sample was 16,393 (19,244 less 2,851).The response for all the 19,244 families, i.e. including the 2,851 families not issued are shown in Table 8. Response frequencies are unweighted.

In total, 13,287 families were productive in the survey, which is 69 per cent of all MCS families and 81 per cent of all families issued. There were 13,469 cohort members in the 13,287 productive families.

Table 8: MCS5 Overall response

Outcome code	n	(%)
Productive	13,287	69.1
Ineligible	623	3.2

Uncertain Eligibility (including untraced movers)		474	2.5
Unproductive	**Refusals**	4,411	22.9
	Non-Contact	438	2.3
	Other	11	0.01
Total		19,244	100.0

Note refusals include broken appointments; non-contact includes ill away and other reasons; unproductive –other covers sensitive family circumstances.

2.8.1. Weighting

As discussed above the sample of births selected for the first survey of the MCS was clustered, geographically, and disproportionately stratified to over-represent areas with high proportions of ethnic minorities in England, residents of areas of high child poverty and residents of the three smaller countries of the UK respectively. The distribution of the cases in the dataset across strata for each country is given in Table 9 below.

Table 9: MCS Cases by stratum and country

	England		Wales		Scotland		N Ireland		UK	
Strata	N	%	n	%	N	%	n	%	N	%
Advantaged	4828	39.49	832	30.14	1145	49.04	723	37.69	7528	39.12
Disadvantaged	4806	39.31	1928	69.86	1191	50.96	1200	62.31	9125	47.42
Ethnic	2591	21.19	n/a	n/a	n/a	n/a	n/a	n/a	2591	13.46
Total	12225	100.00	2760	100.00	2336	100.00	1923	100.00	19244	100.00

The sample design weights or probability weights can be used to correct for MCS cases having unequal probabilities of selection that result from the stratified cluster sample design. The sample weights to be used depend on whether the analysis is confined to data relating to a single country, see Table 10 for country-specific

weights, or whether the analysis covers all countries of the UK, see Table 11 for UK weights.

Table 10: MCS sample design weights by stratum and country (weight1)

Strata	England	Wales	Scotland	N Ireland
Advantaged	1.32	1.77	1.23	1.41
Disadvantaged	0.71	0.65	0.75	0.76
Ethnic	0.24			

Table 11: MCS sample design weights by stratum for the UK (weight2)

Strata	England	Wales	Scotland	N Ireland
Advantaged	2.00	0.62	0.93	0.47
Disadvantaged	1.09	0.23	0.57	0.25
Ethnic	0.37			

Further details are included in *The Millennium Cohort Study: Technical Report on Sampling*, 4th Edition. Plewis, I. (Ed.) July 2007.

Two variables have been included in the dataset to facilitate such weighting by providing the sample weights attached to each case. These are:

- **weight1**: This variable should be used when your analysis is within one country only.
- **weight2**: This variable should be used when your analysis covers the whole of the UK.

2.8.3. Weighting (including non-response adjustment) for MCS1 and MCS2

One way of adjusting for possible biases generated by systematic unit non-response is to use non-response weights. Unit non-response in MCS1 and non-response from MCS1 to MCS2 was studied by Plewis (2007). The correlates of non-response for MCS1 and MCS2 were studied and used to produce non-response weights that can be used to adjust for non-response. For MCS2, there are three different types of weights to consider: 1) the sample design weights; 2) the non-response weights at wave 1 which when multiplied by the sample weights produce the overall weights at wave 1 (see Table 11.1 in *The Millennium Cohort Study: Technical Report on*

Sampling, 4th Edition. Plewis, I. (Ed.) July 2007); and 3) the non-response weights at sweep 2 which when multiplied by the overall weights at sweep 1 produce the overall weights at sweep 2 (see Table 3 in Plewis (2007) for the mean and standard deviation of these weights by stratum for whole UK analyses as well as further technical details on their calculation). Note that the sample at sweep 2 was supplemented by New Families who were eligible at MCS1, but excluded because their addresses held by the Child Benefit Office were not up to date. For these new families, their non-response weight at sweep 2 is defined to be 1. There were 97 sweep 2 productive families that were not used to generate non-response weights due to missing data on the variables used in the response model. These 97 productive families were given a non-response weight of 1.

All family level weights and response level variables are in a file called:

mcs longitudinal family level information. (The user needs to link this file to other files.)

The relevant variable names and value labels are below, where s1, s2, s3 and s4 denote sweeps 1, 2, 3 and 4 respectively and 'inc nr adjustment' denotes including non-response adjustment:

pttyp2   stratum within country fieldwork point number inc. superwards
weight1 mcs weight to use on single country analyses
weight2 mcs weight to use on whole uk analyses
aovwt1   s1: overall weight (inc nr adjustment) single country analysis
aovwt2   s1: overall weight (inc nr adjustment) whole uk analysis
bovwt1   s2: overall weight (inc nr adjustment) single country analysis
bovwt2   s2: overall weight (inc nr adjustment) whole uk analysis

2.8.4. Weighting (including non-response adjustment) for MCS3

Weighting methods to compensate for attrition are available for monotone patterns of non-response. For a monotone pattern, a sequential weighting procedure is typically used. The longitudinal weight at sweep 1 is defined as the sample (design) weight. For each sweep thereafter, the longitudinal weight is the product of the longitudinal weight at the previous sweep multiplied by a non-response weight for the current sweep. Typically, at each sweep the non-response weight is the estimated inverse of the probability of responding based on a logistic regression model. These logistic models use data from previous sweeps to predict response at the current sweep. However, for non-monotone patterns of non-response, some cases have missing data for previous sweeps and therefore the standard approach cannot be easily applied. For MCS, 1,444 unproductive families at MCS2 were recovered at MCS3, thus yielding a non-monotonic pattern of non-response.

In order to calculate non-response weights for MCS3, multiple imputation was used to impute the required missing data at sweep 2 for the logistic regression model for the probability of responding. With the missing data 'filled in', the pattern of non-response was monotone and then the standard sequential weighting procedure

could be used to estimate non-response weights. Note that imputation of missing values was only done for variables found in earlier non-response analyses to be related to non-response, not for all variables in the MCS2 with missing values. Multiple imputation was used to impute missing values at sweep 2 due to unit non-response for unproductive cases and item non-response for productive cases. For example, for the 1,444 unproductive families at MCS2 which were recovered at MCS3, missing housing tenure at MCS2 was imputed using their housing tenure at MCS1 and MCS3 along with other predictor variables in the imputation model. We expect that the imputation of missing values of housing tenure at MCS2 to be 'good' as the imputation model 'loosely speaking' involves 'interpolation' of the values at MCS1 and MCS3. Further detail on the non-response predictor variables and imputation models used will be provided in the Second Edition of the Technical Report on Response.

At sweep 3 all families in the MCS 'active' sample, the 1,922 families had a non-response adjusted weight at sweep 2 and therefore we didn't have to deal with missing weights at sweep 2. As a result of using multiple imputation, all 18,526 issued cases were used in the logistic modelling of response at sweep 3. Missing values were imputed 10 times and a logistic model of responding at sweep 3 was estimated 10 times, once for each imputed dataset. This yielded 10 estimated non-response weights at sweep 3 and the weights issued for sweep 3 were the average of the 10 weights. The overall weights, including non-response adjustment, for single country analysis and whole UK analysis are:

covwt1 s3: overall weight (inc nr adjustment) single country analysis
covwt2 s3: overall weight (inc nr adjustment) whole uk analysis.

Although non-respondents were typically systematically different from respondents at sweeps 1 and 2, Plewis (2007) found that these differences in the probability to respond were small compared to the unequal selection probabilities built into the sample design. The logistic modelling of sweep 3 non-response also found that these differences in the probability to respond were small compared to the unequal selection probabilities built into the sample design. It is, therefore, unlikely that any weighting adjustment for wave 3 non-response would have a substantial effect on most analyses.

2.8.5. Weighting (including non-response adjustment) for MCS4

At sweep 4 we used the same modelling approach and procedures as at sweep 3. In order to calculate non-response weights for MCS4, multiple imputation was used to impute the required missing data at sweep 3 for the logistic regression model for the probability of responding. With the missing data 'filled in', the pattern of non-response was monotone and then the standard sequential weighting procedure could be used to estimate non-response weights. Note that imputation of missing values was only done for variables found in earlier non-response analyses to be related to non-response, not for all variables in the MCS3 with missing values.

19

As a result of using multiple imputation to deal with missing data, all issued cases at sweep 4 were used in the logistic modelling of response at sweep 4. Missing values were imputed 10 times and a logistic model of responding at sweep 4 was estimated 10 times, once for each imputed dataset. This yielded 10 estimated non-response weights at sweep 4 and the weights issued for sweep 4 were the average of the 10 weights.

2.8.6. Weighting (including non-response adjustment) for MCS5

Response at sweep five (MCS 5)

In Table 12, response and non-response rates are presented by category. The table shows that the proportion of productive cases dropped over time from 96.4% in MCS1 to 69% in MCS5. The proportions in all other categories rose as the proportion of non-respondents grew.

Ineligible: includes child deaths, sensitive cases and temporary and permanent emigrants.

Untraced: untraced movers, possibly emigrants.

Table 12: Response rates in all MCS sweeps.

Categories	MCS1		MCS2		MCS3		MCS4		MCS5	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Not Issued	692	3.6	0	0.0	0	0.0	2,213	11.5	2,851	14.8
Productive	18,552	96.4	15,590	81.0	15,246	79.2	13,857	72.0	13,287	69.0
Ineligible	0	0.0	167	0.9	300	1.6	126	0.7	78	0.4
Untraced	0	0.0	687	3.6	547	2.8	706	3.7	388	2.0
Refusal	0	0.0	1,739	9.0	2,315	12.0	1,811	9.4	2,196	11.4
Non-Contact	0	0.0	930	4.8	546	2.8	123	0.6	438	2.3
Other unproductive	0	0.0	131	0.7	290	1.5	408	2.1	6	0.0
Total	19,244	100.0	19,244	100.0	19,244	100.0	19,244	100.0	19,244	100.0

Table 13 shows that response rates were very similar across all four countries with the highest response rate being in England.

Table 13: Response rates by country in MCS5.

Categories	England		Wales		Scotland		NI	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Not Issued	1,761	14.4	365	13.2	414	17.7	311	16.2
Productive	8,618	70.5	1,881	68.2	1,480	63.4	1,308	68.0
Ineligible	53	0.4	9	0.3	10	0.4	6	0.3
Untraced	200	1.6	75	2.7	84	3.6	29	1.5
Refusal	1,309	10.7	335	12.1	306	13.1	246	12.8
Non-Contact	279	2.3	95	3.4	42	1.8	22	1.1
Other unproductive	5	0.0	0	0.0	0	0.0	1	0.1
Total	12,395	100.0	2,798	100.0	2,370	100.0	1,955	100.0

Table 14 shows that the response rates vary across ward types within country. Advantaged households systematically have higher response rates than disadvantaged ones while the ethnic stratum in England has a relatively high response rate.

Adv: Advantaged ward. Dis: Disadvantaged ward. Ethn: Ethnic minority ward.

Table 14: Response rates by stratum in MCS5.

Categories	England			Wales		Scotland		NI	
	Adv.	Dis.	Ethn.	Adv.	Dis.	Adv.	Dis.	Adv.	Dis.
Unproductive	25.5	31.0	34.2	28.3	33.4	32.4	40.7	30.8	32.7
Productive	74.5	69.0	65.8	71.7	66.6	67.6	59.3	68.2	67.3

Table 15 shows that 54.3% of all respondents participated in all waves of MCS. In contrast, 23.9% have interrupted response patterns. In other words, they participated in a number of waves then dropped out before participating again in subsequent waves. 21.9% of all respondents have monotone response patterns. In other words, they participated in a number of waves before definitely dropping out.

Table 15: monotone vs. non-monotone response in MCS5.

Type of non-response	Freq.	%
Monotone	5,023	26.1
Non-monotone	3,773	19.6
All waves	10,448	54.3
Total	19,244	100.0

2.8.9 Predicting response at wave 5 for weight adjustment

The same procedure used for predicting non-response at wave 4 was again used at sweep 5. Missing data for predictor variables due to non-monotone non-response or item missingness were imputed using simple and multiple imputations. Sweep 5 non-response predictors were mostly the same as at sweep 4. Multiple imputations were carried out using the MI command in Stata 12.

As a result of the use of simple and multiple imputations, the sample used in the logit response model consisted of 16393 observations (i.e. the issued sample in MCS5). Weights were constructed for all respondents in MCS5. The dependent variable in the logit model is binary (1 for response and 0 otherwise) and the predictors are: the cohort member's gender, mother's age at first live birth, ethnicity, housing tenure, accommodation type, national vocational qualification, breastfeeding, main respondent's work status, whether the household is a new family which joined the survey in wave 2, and income item non-response. These variables came from all four previous waves.

Imputations were carried out in the following way:

- Simple imputations: ethnicity, accommodation type and NVQ were imputed using the most recent available data from previous waves with simple replacement imputations. The questions on accommodation type and NVQ were only asked if accommodation or NVQ have changed since the last wave of data collection.
- Multiple imputations: main respondent's work status and housing tenure were missing for 2744 observations. Breastfeeding was missing for the new families (617 observations). These three variables were imputed using 10 multiple imputations. Different imputation procedures were used depending on the nature of the variable: a logit procedure for work status and breastfeeding and a multinomial logit for housing tenure. The explanatory variables for the imputation of work status and housing tenure in wave 4 were the exact same variables from the previous three sweeps. For the imputation of breastfeeding I used different variable related to social class as

explanatory variables, these are: ethnicity, NVQ, number of parents in household, and type of accommodation.

It should be noted that 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. Income item non response was constructed as a binary variable which takes the value of 1 if the respondent did not answer the income question. Mother's age at first live birth was missing for only 49 observations; these were replaced by the average age of the non-missing cases.

Table 16 shows the odds ratios of the response logit model estimated using the 10 imputed datasets. The linear predicted values were generated from this model then an inverse-logit transformation was carried out to transform the predicted values into predicted probabilities. The non-response weights at sweep 5 were constructed as the inverse of the predicted probabilities. Two overall weights were constructed by multiplying the aforementioned non-response weights with the same weights from sweep 4. These overall weights adjust for both sampling and attrition. The weights are:

EOVWT1: Sweep 5 overall weight for single country analysis.

EOVWT2: Sweep 5 overall weight for whole of UK analysis.

Table 16: The logit response model.

Explanatory variables	Odds Ratio	Std. Err.	t-statistic	P>t
Boy	0.89	0.039	-2.60	0.009
Mother's age at first live birth, reference: [20-30[				
Before 20	0.82	0.044	-3.71	0.000
[30-40[	1.51	0.100	6.22	0.000
After 40	0.92	0.290	-0.27	0.784
Ethnicity, reference: White				
Mixed	1.04	0.130	0.34	0.737
Indian	1.14	0.163	0.91	0.365
Pakistani, Bangladeshi	2.05	0.195	7.50	0.000
Black	0.78	0.085	-2.28	0.022

Other	1.02	0.142	0.14	0.892
Housing tenure, reference: mortgage				
Own	0.89	0.108	-0.94	0.350
Rent LA or HA	0.76	0.047	-4.48	0.000
Rent privately	0.73	0.063	-3.63	0.000
Other	0.62	0.071	-4.18	0.000
Type of accommodation, reference: house, bungalow				
Anything else (flat, studio, other)	1.33	0.086	4.41	0.000
National Vocational Qualification, reference: NVQ 1				
NVQ 2	0.95	0.079	-0.65	0.514
NVQ 3	1.01	0.095	0.14	0.886
NVQ 4	1.21	0.110	2.05	0.040
NVQ 5	1.57	0.223	3.17	0.002
NVQ 6	0.85	0.073	-1.88	0.060
Breastfeeding attempted	1.36	0.068	6.17	0.000
Respondent in work	1.09	0.058	1.61	0.109
New family	0.93	0.101	-0.67	0.505
Income item non-response	0.21	0.009	-35.51	0.000
Constant	5.21	0.611	14.09	0.000
N	16393			

Number of imputations: 10; Minimum DoF: 82; LA and HA are local authority and housing association.

In tables 17 and 18, the means, minimums and maximums of the two weights are presented by ward type and for the UK as a whole.

Table 17: EOVWT1, Wave5 overall weight for single country analysis.

Ward type	Obs	Mean	Std. Dev.	Min	Max
England - Advantaged	3598	1.60	0.60	1.05	10.67
England - Disadvantaged	3316	1.14	0.53	0.58	8.06
England - Ethnic	1704	0.51	0.27	0.21	2.86
Wales - Advantaged	597	1.96	0.73	1.22	6.37
Wales - Disadvantaged	1284	0.90	0.40	0.46	5.23
Scotland - Advantaged	774	1.34	0.72	0.48	7.05
Scotland - Disadvantaged	706	1.09	0.67	0.30	5.65
Northern Ireland - Advantaged	500	1.56	0.86	0.49	7.94
Northern Ireland - Disadvantaged	808	1.14	0.69	0.28	5.24
Total	13287	1.22	0.69	0.21	10.67

Table 18: EOVWT2, S5 overall weight for whole of the UK analysis.

Ward type	Obs	Mean	Std. Dev.	Min	Max
England - Advantaged	3598	2.05	0.81	1.33	13.55
England - Disadvantaged	3316	1.50	0.71	0.75	10.37
England - Ethnic	1704	0.67	0.35	0.27	3.69
Wales - Advantaged	597	0.66	0.24	0.41	2.07
Wales - Disadvantaged	1284	0.31	0.13	0.16	1.79
Scotland - Advantaged	774	1.04	0.54	0.38	5.55
Scotland - Disadvantaged	706	0.84	0.50	0.24	4.47
Northern Ireland - Advantaged	500	0.61	0.32	0.20	3.18
Northern Ireland - Disadvantaged	808	0.43	0.24	0.11	1.87
Total	13287	1.23	0.87	0.11	13.55

2.9. Recommendations on the use of weights

Analysts wanting to estimate population quantities such as population proportions, quantiles, means and totals, should use weighted estimates. Weights with variables ending with a 1 are for UK country specific analyses and those ending with a 2 are for whole UK analyses. For sweep 1 datasets, use the weight variables beginning with 'a'; for sweep 2 datasets, use the weight variables beginning with 'b'; and for sweep 3 datasets, use the weight variables beginning with 'c', 'd' for sweep 4 and 'e' for sweep 5. For example, at MCS3 use covwt1 for single country estimates or covwt2 for whole UK estimates. These weights take into account the unequal selection probabilities of wards and adjust for non-response. Analysts wanting to estimate coefficients of regression models should include the stratum design variable (pttype2) as dummy variables in the model and unweighted methods can then be used to estimate model parameters. In order to properly estimate standard errors, the clustering of the MCS sample should be taken into account in any analysis by using the STATA svy commands or robust cluster option with PSU identifier if clustering is a nuisance or using a multilevel (hierarchical) model if the clustering is of substantive interest.

PART 3: SURVEY DEVELOPMENT

For a more comprehensive discussion of survey development, please refer to the:

- MCS1 Technical Report on Fieldwork (NatCen 2004)
- MCS2 Technical Report on Fieldwork (NOP 2006)
- MCS3 Technical Report on Fieldwork (NatCen 2007)
- MCS4 Technical Report on Fieldwork (NatCen 2010)
- MCS5 Technical Report on Fieldwork (Ipsos MORI 2013)

3.1. Development and Piloting of MCS1

The questionnaire was developed by the CLS team with input from 55 potential users of the dataset from academe and government departments who attended a consultation meeting on 11 October 2000. An instrument was initially piloted in January 2001 and redeveloped into a shorter version for the second Dress Rehearsal Pilot in April 2001.

3.1.1. First Pilot

The first pilot in January 2001 was conducted as a paper interview and computer-aided self-completion interview (CASI) in order to assess the timing of the instrument before the major work to convert the interview schedule into computer-aided personal interview (CAPI) format. The sample size was boosted from 30 to 60 thanks to the ONS consortium funding. Further details are in the NatCen Technical Report on Fieldwork (NatCen 2004).

3.1.2. Dress Rehearsal Pilot

The second pilot took place during April 2001 and was fully computer-based (CAPI and CASI). As a 'dress rehearsal' for the main stage, all the contact and administrative processes were tested as well as the near final form of the survey instruments. Thirteen wards were selected for this pilot, including one in each of Wales and Scotland. The wards in England and Wales were chosen from those that were to be used in the main stage. As the Scottish wards had not yet been selected, a large deprived ward was purposively picked.

The DWP sampling route was tested with letters sent from the DWP at Newcastle to parents of babies born between 12 June and 22 July 2000 on the Child Benefit register in the chosen wards. The use of an advance letter sent by interviewers was also piloted.

1

In addition, Health Visitors (HVs) were approached in the 12 English and Welsh wards in order to pilot their contribution. Two HV supervisors declined to help, as we had not received Multi-centre Research Ethics Committee (MREC) approval at that time.

3.2. Development and Piloting of MCS2

The questionnaire was developed by the CLS team with input from a team of external MCS2 collaborators. The questionnaire development was discussed at a consultative meeting on 22 April 2002. An instrument was initially piloted in May 2003, and redeveloped for the second Dress Rehearsal Pilot in June 2003.

3.2.1. First Pilot

The first pilot in May 2003 was carried out as CAPI and CASI interviews of around 30 families in order to establish the time taken to carry out the early drafts of the interview, self-completion and child assessments. It was also designed to identify other problems such as flow, question wording recall and filtering.

3.2.2. Dress Rehearsal Pilot

The dress rehearsal for the study took place in June 2003. All of the procedures planned for main-stage sampling and fieldwork were tested, including the taking of saliva samples from the children; home and neighbourhood observations; and the self-completion questionnaire for older siblings. The sample used for the MCS2 dress rehearsal consisted of respondents from the MCS1 dress rehearsal. Forty-eight families were interviewed in 13 wards in England, Wales and Scotland.

3.3. Development and Piloting of MCS3

The questionnaire was developed by the CLS team with input from a team of external MCS3 collaborators. The questionnaire development was discussed at a consultative meeting in July 2004. An instrument was initially piloted in May 2005, and redeveloped for the second Dress Rehearsal Pilot in September/October 2005.

3.3.1. First Pilot

The first pilot in May 2005 was carried out as CAPI and CASI interviews of 49 families in order to establish the time taken to carry out the early drafts of the interview, self-completion and child assessments and measurements. It was also designed to identify other problems such as flow, question wording recall and filtering. The sample was a quota sample recruited by interviewers.

3.3.2. Dress Rehearsal Pilot

The dress rehearsal for the study took place in September/October 2005. All of the procedures planned for main-stage sampling and fieldwork were tested.

The sample used for the MCS3 dress rehearsal consisted, in England, Scotland and Wales, of respondents from the MCS1 dress rehearsal and additional families sampled for MCS3. Northern Ireland was included in the dress rehearsal for the first time at MCS3; and all families in Northern Ireland were newly sampled for MCS3. The dress rehearsal sample was drawn from Child Benefit records in 14 wards of the UK and109 families were interviewed.

The dress rehearsal also included a postal teacher survey in Wales, Scotland and Northern Ireland. This was in order to collect data equivalent to the Foundation Stage Profile in England (which was obtained through data linkage for consenting families in England).

3.4. Development and Piloting of MCS4

The data collection instruments were developed by the CLS team with input from a team of external MCS4 advisors. The development work started with a consultative conference in July 2008 at which the convenors of the MCS4 advisory groups presented their recommendations. A consultation on the first draft questionnaire for parents and cohort members took place in January/February 2007 and on the first draft teacher questionnaire in February/March 2007. The first pilot took place in March-June 2007 and the Dress Rehearsal Pilots for families and teachers in July-August 2007 and October-December 2007, respectively.

3.4.1. First Pilot

The first pilot in March/April 2007 was carried out as CAPI and CASI interviews of 38 families in order to establish the time taken to carry out the early drafts of the parent interviews and self-completion; child self-completion; and child assessments and measurements. It was also designed to identify other problems such as flow, question wording recall and filtering. Of the 38 interviewed families, 26 had previously been interviewed at MCS3 pilot 1, and 12 were newly recruited by interviewers. It was a quota sample and covered Great Britain only.

The teacher survey pilot took place in May-June 2007. Of the families who took part in the main pilot, 32 gave consent for their child's teacher to be approached. Of these, 23 returned a questionnaire after 2 reminders, giving a response rate of around 72 per cent.

3.4.2. Dress Rehearsal Pilot

The dress rehearsal for the study took place in July/August 2007. All of the procedures planned for main-stage sampling and fieldwork were tested.

The longitudinal dress rehearsal sample, drawn from Child Benefit records in 14 wards of the UK, consisted, in Great Britain, of respondents sampled for the MCS1 dress rehearsal and additional families sampled for MCS3. In Northern Ireland it consisted of respondents sampled at MCS3, and 102 families were interviewed. This was in excess of the target sample of 100 families.

The dress rehearsal also included a postal teacher survey which was carried out in October-December 2007. In all, 84 teachers were approached (consenting families in the main dress rehearsal) and 38 questionnaires were returned after 2 reminders, giving a response rate of 45 per cent.

3.5. Development and Piloting of MCS5

The data collection instruments were developed by the CLS team with input from a team of external MCS5 advisors. The development work started with a consultative conference in July 2010 at which the convenors of the MCS5 advisory groups presented their recommendations. A consultation on the first draft questionnaires took place in November 2010-January 2011. The first pilot took place in March-April 2011 for families and in May-June 2011 and the Dress Rehearsal Pilot for families and teachers in August-September 2011 and October-November 2011, respectively.

3.5.1. First Pilot

The first pilot in March/April 2011 was carried out as CAPI and CASI interviews of 45 families in order to establish the time taken to carry out the early drafts of the parent interviews and self-completion; child self-completion; and child assessments and measurements and to test the feasibility of saliva sample collection. It was also designed to identify other problems such as flow, question wording recall and filtering. It took place in five areas on Great Britain only. All of the families newly recruited by interviewers using quota sampling.

The teacher survey pilot took place in August-September 2011 covering England and Wales only. Of the 37 families in England and Wales who took part in the main pilot, 31 gave consent for their child's teacher to be approached. Of these, 19 returned a questionnaire after reminders, giving a response rate of around 61 per cent.

3.5.2. Dress Rehearsal Pilot

The dress rehearsal for the study took place in August/September 2011. All of the procedures planned for main-stage sampling and fieldwork were tested.

The longitudinal dress rehearsal sample, drawn from Child Benefit records in 14 wards of the UK, consisted, in Great Britain, of respondents sampled for the MCS1 dress rehearsal and additional families sampled for MCS3. In Northern Ireland it consisted of respondents sampled at MCS3. Additional families in England were sampled through the Department for Education's National Pupil Database and in England through the Welsh Government's record of pupils in Wales. In total, 126 families were interviewed. This was in excess of the target sample of 100 families.

The dress rehearsal also included a postal teacher survey which was carried out in October-November 2011. In all, 103 teachers were approached (consenting families in the main dress rehearsal in England and Wales) and 56 questionnaires were returned after reminders, giving a response rate of 54 per cent.

PART 4: SURVEY CONTENT

The chart below shows the content of the MCS surveys at a glance.

MCS 1	MCS2	MCS 3	MCS4	MCS5
2001/2	2003/4	2005/6	2008/9	20012/13
9 months	AGE 3	AGE 5	AGE 7	AGE 11
Main	Main	Main	Main	Main
Partner	Partner	Partner	Partner	Partner
	Cohort Member	Cohort Member	Cohort Member	Cohort Member
	Older Siblings	Older Siblings		
			Teachers	*Teachers E & W on*
18,552	**15,590**	**15,246**	**13,857**	**13,287**
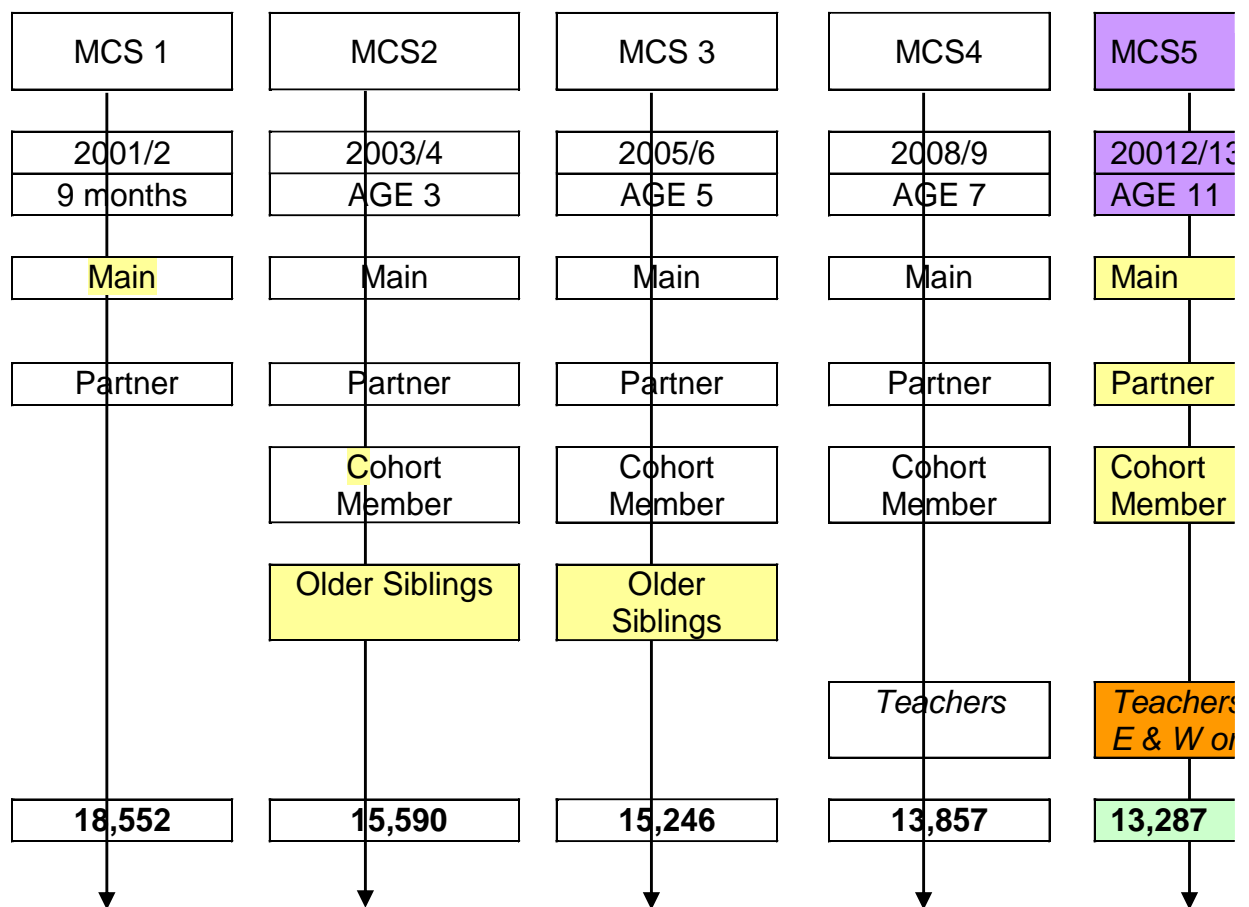
Tables 19-23 below show in detail elements included at each sweep of the MCS. For more details of the content for all surveys, please refer to the respective questionnaires.

Table 19: MCS1 – Summary of MCS1 Survey Elements.

Respondent	Mode	Summary of content
Mother/Father	Interview	Household Module
Mother/Main*		Module A: Non-resident parents
		Module C: Pregnancy, labour and delivery
		Module D: Baby's health and development
		Module E: Childcare
		Module F: Grandparents and friends
		Module G: Parental health
	Self-completion	Module H:
		- Baby's temperament & behaviour
		- Relationship with partner
		- Previous relationships
		- Domestic tasks
		- Previous pregnancies
		- Mental health
		- Attitudes to relationships, parenting, work, Etc
	Interview	Module J: Employment, income, education
		Module K: Housing and local area
		Module L: Interests
Father/Partner*	Interview	Module B: Father's involvement with baby
		Module C: Pregnancy, labour and delivery (where applicable)
		Module F: Grandparents and friends
		Module G: Parental health
	Self-completion	Module H: Self-completion
		- Baby's temperament & behaviour
		- Relationship with partner
		- Previous partners
		- Previous children

		- Mental health
		- Attitudes to marriage, parenting, work, etc
	Interview	Module J: Employment and education
		Module L: Interests
* In the majority of cases, the Main interview was undertaken by the mother/mother figure while the Partner interview was undertaken by the father/father figure. See Table 20.		

Table 20: MCS2 – Summary of MCS2 Survey Elements

Respondent	Mode	Summary of content
Mother/Father	Interview	Household Module
Mother/Main*		Module A: Non-resident parents
		Module C: Pregnancy, labour and delivery
		Module D: Baby's health and development
		Module E: Childcare
		Module F: Grandparents and friends
		Module G: Parental health
	Self-completion	Module H:
		- Child's temperament & behaviour
		- Relationship with partner
		- Previous relationships
		- Domestic tasks
		- Previous pregnancies
		- Mental health
		- Attitudes to relationships, parenting, work, etc
	Interview	Module J: Employment, income, education
		Module K: Housing and local area
		Module L: Interests and time with baby
		Module N : Older siblings
Father/Partner*	Interview	
		Module B: Father's involvement with baby
		Module C: Pregnancy, labour and delivery (where applicable)
		Module F: Grandparents and friends
		Module G: Parent's health
	Self-completion	Module H: Self-completion
		- Baby's temperament & behaviour
		- Relationship with partner
		- Previous partners
		- Previous children
		- Mental health
		- Attitudes to marriage, parenting, work, etc
	Interview	Module J: Employment and education
		Module L: Interests
Interviewer	Observations	Home environment
		Neighbourhood
Child	Assessments	BAS Naming Vocabulary
		Bracken Basic Concept Scale
		Height and weight
		Oral fluids
Older sibling	Self-completion**	

* In the majority of cases, the Main interview was undertaken by the mother/mother figure and the Partner interview was undertaken by the father/father figure. See Table 21.
** England only.

Table 21: MCS3 – Summary of MCS3 Survey Elements

Respondent	Mode	Summary of content
Mother/Father	Interview	Module HD: Household demographics
Mother/Main*		Module FC: Family context
		Module ES: Early education, schooling and childcare
		Module AB: Child and family activities and child behaviour
		Module PA: Parenting activities
		Module CH: Child health
		Module PH: Parental health
		Module EI: Employment, education and income
		Module HA: Housing and local area
		Module OM: Other matters
	Self-completion	Module SC: Self-completion
		- Child's temperament and behaviour
		- Child's relationship with siblings
		- Parenting and parent-child relationship
		- Mental health and drug-taking
		- Relationship with partner
		- Previous relationships, children living elsewhere, non-resident parents
		- Attitudes and ethnic identity
		- Racial harassment and discrimination
		- Work-life balance and life satisfaction
		- Older Siblings' temperament and behaviour
	Interview	Module OS: Older siblings
		Module Z: Consents and contact information
Father/Partner*	Interview	Module FC: Family context
		Module ES: Early education, schooling and childcare (some)
		Module PA: Parenting activities
		Module PH: Parental health
		Module EI: Employment, education and income
		Module OM: Other Matters
	Self-completion	Module SC: Self-completion
		- Parenting and parent-child relationship
		- Mental health and drug-taking
		- Relationship with partner
		- Previous relationships, children living elsewhere
		- Attitudes and ethnic identity
		- Racial harassment and discrimination
		- Work-life balance and life satisfaction
	Interview	Module Z: Consents and contact information
Interviewer	Observations	Cognitive assessment

Respondent	Mode	Summary of content
Child	Assessments	Story of Sally and Anne
		British Ability Scales: Picture Similarities
		British Ability Scales: Naming Vocabulary
		British Ability Scales: Pattern Construction
	Measurements	Height, weight and waist circumference
Older sibling	Self-completion**	
Teacher	Self-completion***	Questions equivalent to Foundation Stage Profile in England
In the majority of cases, the Main interview was undertaken by the mother/mother figure and the Partner interview was undertaken by the father/father figure. See Table 22. *** England only.* **** Wales, Scotland and Northern Ireland only.*		

Table 22: MCS4 – Summary of MCS4 Survey Elements.

Respondent	Mode	Summary of content
Mother/Father	Interview	Module HD: Household demographics
Mother/Main*		Module FC: Family context
		Module ES: Early education, schooling and childcare
		Module AB: Child and family activities and child behaviour
		Module PA: Parenting activities
		Module CH: Child health
		Module PH: Parental health
		Module EI: Employment, education and income
		Module HA: Housing and local area
		Module OM: Other matters
	Self-completion	Module SC: Self-completion
		- Child's temperament and behaviour
		- Child's relationship with siblings
		- Parenting and parent-child relationship
		- Mental health
		- Relationship with partner
		- Previous relationships, children living elsewhere, non-resident parents
		- Attitudes, ethnic identity, racial harassment and discrimination
		- Personality (OCEAN)
		- Life satisfaction
	Interview	Module Z: Consents and contact information
Father/Partner*	Interview	Module FC: Family context
		Module ES: Early education, schooling and childcare (some)
		Module PA: Parenting activities

Respondent	Mode	Summary of content
		Module PH: Parental health
		Module EI: Employment, education and income
		Module OM: Other Matters
	Self-completion	Module SC: Self-completion
		- Parenting and parent-child relationship
		- Mental health
		- Relationship with partner
		- Previous relationships, children living elsewhere
		- Attitude, ethnic identity and racial harassment and discrimination
		- Personality (OCEAN)
		- Life satisfaction
	Interview	Module Z: Consents and contact information
Interviewer	Observations	Cognitive assessment
Child	Assessments	Story of Sally and Anne
		British Ability Scales: Word Reading
		British Ability Scales: Pattern Construction
		Progress in Maths (Millennium Cohort Study edition)
	Measurements	Height, weight, body-fat and waist circumference and physical activity monitoring
	Self-completion	Hobbies, friends and family, feelings, school
Teacher	Self-completion	Child's abilities and behaviour
		Suspensions and exclusions
		Language of schooling and language needs
		Special Educational Needs/Additional support needs
		Parental interest in education
		Setting and streaming
		Teacher demographics
		Study child's class
* In the majority of cases, the Main interview was undertaken by the mother/mother figure and the Partner interview was undertaken by the father/father figure. See Table 23 below.		

Table 23: MCS5 – Summary of FIFTH Survey Elements.

Respondent	Mode	Summary of content
Main/Partner	Interview	Module HD: Household demographics
Main*		Module FC: Family context
		Module ES: Education and schooling
		Module AB: Child and family activities and child behaviour
		Module PA: Parenting activities
		Module CH: Child health
		Module PH: Parental health
		Module EI: Employment, income and education
		Module HA: Housing and local area
		Module OM: Other matters
	Self-completion	Module SC: Self-completion
		- Strength & difficulties questionnaire
		- Discipline
		- Relationship with cohort member
		- CM's pubertal development
		- Attitudes, racial harassment and discrimination; anti social behaviour; consumerism
		- Mental health
		- AUDIT (alcohol consumption)
		- Relationship with partner
		- Life satisfaction
	Interview	Module Z: Consents and contact information
Partner*	Interview	Module FC: Family context
		Module ES: Education and schooling (partial)
		Module PA: Parenting activities
		Module PH: Parental health
		Module EI: Employment, income and education
		Module OM: Other Matters
	Self-completion	Module SC: Self-completion
		- Relationship with cohort member
		- Attitude, racial harassment and discrimination; anti social behaviour; consumerism
		- Mental health
		-AUDIT (alcohol consumption)
		- Relationship with partner
		- Life satisfaction
	Interview	Module Z: Consents and contact information

Respondent	Mode	Summary of content
Interviewer	Observations	Cognitive assessment
Cohort Member	Assessments	British Ability Scales: Verbal Similarities
		CANTAB Spatial Working Memory (Memory task)
		CANTAB Cambridge Gambling Task (Decision making task)
	Measurements	Height, weight and body fat
	Self-completion	Activities outside school, Internet & social networking, Life satisfaction, happiness &self esteem, Friends & unsupervised time, Pocket money, family financial position & materialism, Anti social behaviours, School Secondary school; Attitudes; Other children (incl. bullying); Risky behaviours (incl. smoking & alcohol), Mental health, Future ambitions
Teacher	Self-completion	Child's abilities and behaviour
		Suspension & truancy
		CM profile (including EAL, SEN, help & support, peers, bullying
		Move to secondary school
		Future education
		Parents
		Class groupings & setting
		Child's class
		Teacher profile
* In the majority of cases, the Main interview was undertaken by the mother/mother figure and the Partner interview was undertaken by the father/father figure. See Table (n) below.		

PART 5: FIELDWORK

For a more comprehensive discussion of survey development, please refer to the:

- MCS1 Technical Report on Fieldwork (NatCen 2004)
- MCS2 Technical Report on Fieldwork (NOP 2006)
- MCS3 Technical Report on Fieldwork (NatCen 2007)
- MCS4 Technical Report on Fieldwork (NatCen 2010)
- MCS5 Technical Report on Fieldwork (Ipsos MORI 2013)

5.1. Fieldwork for MCS1

Following a competitive tender process NatCen was appointed to carry out the fieldwork for MCS1. The fieldwork in Northern Ireland was sub-contracted by NatCen to the Central Survey Unit of NISRA (the Northern Ireland Statistics and Research Agency). For the most part it took place in 2002, having started in England and Wales in June 2001, and in Scotland and Northern Ireland in September 2001. It finished in January 2003.

5.1.1. Briefings

Briefings for the 232 interviewers who were to work in England and Wales were held in 17 regional one-day meetings between 31 May and 15 June 2001. A further 42 interviewers working in Scotland were briefed at 4 sessions between 29 August and 6 September. These training sessions were conducted jointly by researchers from NatCen and CLS. In Northern Ireland, some 50 interviewers were briefed at 4 sessions between 17 and 28 August.

5.1.2. Fieldwork Timetable

The fieldwork for MCS1 (and MCS2) was carried out in 17 consecutive waves. Each issued wave of fieldwork contained babies born in a 4-weekly birth cycle (apart from the last), with the first wave covering the births between 1 and 28 September 2000 in England and Wales. This rhythm of recruiting the sample was dictated by the cycle of DWP procedures, scanning the Child Benefit database every 4 weeks. Interviewers arranged interviews as soon as possible after the addresses were issued, aiming to reach the families while the baby was as close as possible to 9.5 months of age. Interviews with partners could be delayed until the child's first birthday (as were some main interviews where the address had been issued late).

The process for drawing each wave of the DWP sample is as follows:

Prior to fieldwork, the DWP sent opt-out letters to all parents of children with an eligible birth date who were registered (for Child Benefit purposes) as living within

one of the sampled wards, apart from any cases flagged as sensitive. Batches of letters, including an information leaflet, were sent every 4 weeks to families whose babies were approximately 7 months old. The letter invited parents to take part in the study and gave them the opportunity to opt out of the study by telephoning or writing to the DWP. Any parents who opted out of the study were then removed from the sample.

The final stage was for the DWP to remove cases which they discovered had subsequently moved out of the sampled wards and to update the addresses for cases which had moved within or between sampled wards. At this stage any late opt-outs or newly sensitive cases were also removed.

The data were sent by the DWP to CLS in two stages, a week apart, in order to ensure that any late opt-outs or changes of addresses could be notified as near to the start of fieldwork as possible. After the final data were received serial numbers were assigned to each valid case and the data were sent to NatCen, for issue to the field.

The fieldwork timetable for the project detailing the dates of birth and fieldwork is shown in the table below.

Table 24: Fieldwork timetable for MCS1.

Fieldwork Wave	Baby's Date of Birth	Fieldwork Period
Wave 1	1 – 28 Sep 2000	11 June – 8 Jul 2001
Wave 2	29 Sep – 26 Oct 2000	9 Jul – 5 Aug 2001
Wave 3	27 October – 23 Nov 2000	6 Aug – 2 Sep 2001
Wave 4	24 Nov – 21 Dec 2000	3 Sep – 30 Sep 2001
Wave 5	22 Dec 2000 – 18 Jan 2001	1 Oct – 28 Oct 2001
Wave 6	19 Jan – 15 Feb 2001	29 Oct – 25 Nov 2001
Wave 7	16 Feb – 15 Mar 2001	26 Nov – 23 Dec 2001
Wave 8	16 Mar – 12 Apr 2001	24 Dec 2001 – 20 Jan 2002
Wave 9	13 Apr – 10 May 2001	21 Jan – 17 Feb 2002
Wave 10	11 May – 7 June 2001	18 Feb – 17 Mar 2002
Wave 11	8 June – 5 Jul 2001	18 Mar – 14 Apr 2002

Wave 12	6 Jul – 2 Aug 2001	15 Apr – 12 May 2002
Wave 13	3 Aug – 30 Aug 2001	13 May – 9 June 2002
Wave 14	31 Aug – 27 Sep 2001	10 June – 7 Jul 2002
Wave 15	28 Sep – 25 Oct 2001	8 Jul – 4 Aug 2002
Wave 16	26 Oct – 23 Nov 2001	5 Aug – 22 Sep 2002
Wave 17	24 Nov 2001-11 Jan 2002	23 Sep –10 Jan 2003

Note: NatCen numbered these waves 2-18 as they counted the Dress Rehearsal Pilot as Wave 1.

Waves 1-13 of fieldwork took place in England and Wales from June 2001 to July 2002. The last wave in England and Wales, wave 13, which included babies born on 31 August, was delayed by 4 weeks for operational reasons, so this wave contained interviews mostly conducted at 10 rather than 9 months for these 2 countries. The last wave in Scotland and Northern Ireland, wave 17, was the extended sample spanning 7 weeks of births. The latest interview (with a partner) took place in Northern Ireland on the last-but-one eligible day, 10 January 2003. Fieldwork in Scotland (and with all main informants) finished before the end of 2002.

The aim was that the fieldwork for each wave should be as self-contained as possible, with the minimum amount of overlap. Interviewers were briefed to interview families when the baby was 9 months and 15 days old, ideally, in order to standardise the data being collected as far as possible. Allowing for delayed interviewing due to tracing problems, the window of opportunity to interview was brief, up to 11 months of the babies' age for the main interview and up to 12 months for the partner.

Seventy-five per cent of main interviews took place while the baby was aged 9 months – 3,579 (19 per cent) at 10 months with 541 (3 per cent) at 8 months – representing babies born towards the end of the 4-week span interviewed early in the fieldwork period. However, 479 interviews took place late, 475 at 11 months and only 4 in months 12-13. Seventeen were not interviewed because the time window had expired by the time they were found. They are included in the 'other ineligible', Table 7.2 in the Technical Report on Sampling (4th Edition) (Plewis 2007).

5.1.3. Languages

In order to comply with the recommendations made by the Multi-Centre Research Ethics Committee (MREC), a simplified leaflet was produced for interviewers to give to respondent families on the doorstep. This leaflet, the advance letter and the thank-you letter were translated into the most common non-English languages spoken in

the 19 selected 'ethnic' wards. The languages appropriate for translation were: Bengali, Gujarati, Kurdish, Punjabi, Somali, Turkish and Urdu. The first leaflet had already been translated into Welsh. Some interviews were carried out in verbal translation (in these and other languages) by relatives or friends. In certain circumstances where no one was available to translate into English, NatCen provided translator interviewers. Other languages encountered in non-trivial numbers included Arabic, Hindi and Tamil. Two hundred and twenty-six (1 per cent) main interviews were carried out in a language other than English and a further 547 (3 per cent) were done in a mix of English and another language. For partners the corresponding figures were 306 (2 per cent) and 94 (1 per cent).

5.1.4. In-field Tracing

On the whole, the addresses supplied by DWP proved to be current. Unfortunately, in a proportion of those issued to the field, the families had moved, either after the baby was aged 7 months or else before the baby reached that age, but had not informed the DWP of their move. Where a family was not living at the issued address and the interviewer could not establish a new local address, cases were returned to CLS for tracing. Where a new address was found within a selected ward, cases were re-issued to the field. Where a family had moved to a non-selected area, but were resident at their old address when the baby was aged 9 months, they could be interviewed at the new address.

5.1.5. Data Collection Errors

In a number of cases, interviewers made errors in data collection which were identified by the fieldwork agency during the data preparation stage. Where possible, the data were cleaned to correct these errors. In a small number of cases (identified below) this has not been possible and users should exercise caution when using data for these cases. These cases mostly involved incorrect application of the proxy module and are identified on the variable 'errtype'.

Table 25: MCS1 Data Collection Errors

	Error Type	N	Action taken
1	Proxy module done in error, i.e. the proxy section of the Main interview was completed about a partner who was not eligible to be interviewed by proxy.	117	**Data deleted from proxy module, household outcome code re-classified to 'partial household' and partner outcome code re-classified to unproductive.**
2	Partner interview done by proxy in error, i.e. the main respondent has completed the partner interview on behalf of partner. Partner should have done the interview him/herself.	42	**Data deleted from partner interview, household outcome code re-classified as 'partial household' and partner outcome code re-classified to unproductive.**
3	Partner answered proxy in person, should have done normal partner	6	**Data transferred from proxy section to equivalent variables in**

	interview, i.e. the partner completed the proxy module in person (about him/herself).		**partner interview, household outcome code re-classified as 'main and partner in person' and partner outcome code re-classified to 'partial interview in person'.**
4	Main interview done by father, partner interview by mother, i.e. the data indicate that the mother did the main interview and the father did the partner interview but the main interview was actually conducted with the father (in error) and the partner interview was actually conducted with the mother (in error).	2	**NONE**
5	Father did both main and partner interviews, i.e. the data indicate that the mother completed the main interview and the father completed the partner interview but actually the father conducted both interviews (should have only done the partner interview).	1	**NONE**
6	Main interview done by partner, no other interview, i.e. the data indicate that the mother completed the main questionnaire and the father did not respond to the partner questionnaire but actually the father completed the main interview (in error) and there was no partner interview.	1	**NONE**
7	**Grandmother (person 1 in household) was incorrectly coded as natural mother. The actual natural mother (who was person 3) completed the main interview.**	1	**Relevant variables corrected.**

5.2. Fieldwork for MCS2

Following a competitive tender process, the fieldwork for MCS2 was carried out by NOP Research. The work in Northern Ireland was sub-contracted to Millward Brown Ulster. This survey was conducted mainly during 2004. The main-stage started in England and Wales in September 2003, and in Scotland and Northern Ireland in December 2003. Fieldwork finished in early 2005.

5.2.1. Briefings

Interviewers who were to work in England and Wales were briefed before the start of fieldwork in 13 regional 3-day meetings. Interviewers working in Scotland were briefed at 3 additional sessions. These training sessions were conducted jointly by researchers from NOP and CLS. In Northern Ireland, some interviewers were briefed in just one session by Millward Brown and CLS researchers. There were 5 further briefings during the course of fieldwork as new interviewers were added.

Some 150 interviewers were initially briefed to work on the survey; but by the time fieldwork was complete around 200 interviewers had worked on the survey. Further details may be found in the NOP Technical Report on Fieldwork (NOP 2006).

5.2.2. Fieldwork Timetable

Fieldwork started in September 2003 in England and Wales finished in April 2005. In Scotland and Northern Ireland, fieldwork started in December 2003 and finished in January 2005.

Table 26: Fieldwork timetable for MCS2

Fieldwork Wave	Baby's Date of Birth	Fieldwork Period
Wave 1	1 – 28 Sep 2000	September 2003
Wave 2	29 Sep – 26 Oct 2000	October 2003
Wave 3	27 Oct – 23 Nov 2000	November 2003
Wave 4	24 Nov – 21 Dec 2000	December 2003
Wave 5	22 Dec 2000 – 18 Jan 2001	January 2004
Wave 6	19 Jan – 15 Feb 2001	February 2004
Wave 7	16 Feb – 15 March 2001	March 2004
Wave 8	16 Mar – 12 April 2001	April 2004
Wave 9	13 April – 10 May 2001	May 2004
Wave 10	11 May – 7 Jun 2001	June 2004
Wave 11	8 Jun – 5 Jul 2001	July 2004
Wave 12	6 Jul – 2 Aug 2001	August 2004
Wave 13	3 Aug – 30 Aug 2001	September 2004
Wave 14	31 Aug – 27 Sep 2001	October 2004
Wave 15	28 Sep – 25 Oct 2001	November 2004
Wave 16	26 Oct – 23 Nov 2001	December 2004
Wave 17	24 Nov 2001-11 Jan 2002	January 2005

5.2.3. Languages

A breakdown of interviews by 'language interviewed in' is provided in the Technical Report on Fieldwork (NOP 2006).

5.2.4. In-field Tracing

Families who had moved from the issued address were traced in the field by NOP interviewers. Families who could not be successfully traced by interviewers were returned to CLS for additional tracing by the Tracing team. Details of in-field tracing activities can be found in the Technical Report on Fieldwork (NOP 2006).

5.3. Fieldwork for MCS3

Following a competitive tender process the NatCen was appointed to carry out the fieldwork for MCS3. The fieldwork in Northern Ireland was sub-contracted by NatCen to the Central Survey Unit of NISRA (the Northern Ireland Statistics and Research Agency). The main stage of this fieldwork took place within the calendar year of 2006, starting in England and Wales in January 2006, and in Scotland and Northern Ireland in April 2008. The survey also included a follow-on survey of teachers outside England extending into 2007.

5.3.1. Briefings

Interviewers were briefed in 3-day training sessions. These sessions were conducted jointly by researchers from NatCen and CLS. For further details see NatCen (2007).

5.3.2. Fieldwork Timetable

The fieldwork timetable for MCS3 was driven by the requirement to interview the family during the child's first year of compulsory schooling (Reception Class in England and Wales and Primary One in Scotland and Northern Ireland). As a result, fieldwork was compressed into school years. In England and Wales, the cohort's birth dates span a single school year. However, in Scotland and Northern Ireland the birth dates 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 preference in addition to date of birth. For this reason, school year was known with less certainty in advance in Scotland. During the first wave of fieldwork in Scotland, interviewers were asked to find out, before conducting the interview, whether the child had started school. If the child had not yet started school, the interview was deferred until the second wave of fieldwork.

Table 27: Fieldwork timetable for MCS3 – Main Survey

Wave	Country	Dates of birth	Fieldwork
E1	England	1 Sep 2000 – 28 Feb 2001	Jan – May 2006
E2	England	1 Mar 2001 – 11 Jan 2002	Apr – Jul 2006
W1	Wales	1 Sep 2000 – 28 Feb 2001	Jan – May 2006
W2	Wales	1 Mar 2001 – 11 Jan 2002	Apr – Jul 2006
S1	Scotland	1 Sep 2000- 28 Feb 2001 (starting school in Aug 2005)	Apr – Jul 2006
S2	Scotland	1 Sep 2000- 28 Feb 2001 (starting school in Aug 2006) and 1 Mar 2001 – 11 Jan 2002	Aug – Dec 2006
N1	Northern Ireland	24 Nov 2000 – 1 July 2001	Apr – Jul 2006
N2	Northern Ireland	2 July 2001 – 11 Jan 2002	Sep – Dec 2006

Table 28: Fieldwork timetable for MCS3 – Teacher Survey in Wales, Scotland and Northern Ireland

Teacher Wave	Country	Main Fieldwork Wave	Teacher Fieldwork
T1	Wales, Scotland and Northern Ireland	W1 & W2, S1 and N1	Sep 2006- Jan 2007
T1 – mop-up	Wales, Scotland and Northern Ireland	W1 & W2, S1 and N1	Jan – May 2007
T2	Wales, Scotland and Northern Ireland	W2, S1 & S2 and N1 & N2	Mar – Jun 2007

5.3.3. Languages

A breakdown of interviews by 'language interviewed in' is provided in the Technical Report on Fieldwork (NatCen 2007).

5.3.4. In-field Tracing

Families who had moved from the issued address were traced in the field by NatCen interviewers. Families who could not be successfully traced by interviewers were returned to CLS for additional tracing by the Tracing Unit. Details of in-field tracing activities can be found in the Technical Report on Fieldwork (NatCen 2007).

5.4. Fieldwork for MCS4

Following a competitive tender process the NatCen was appointed to carry out the fieldwork for MCS4. This was a planned extension to their existing contract for MCS3. The fieldwork in Northern Ireland was sub-contracted by NatCen to the Central Survey Unit of NISRA (the Northern Ireland Statistics and Research Agency). The first wave of the main stage fieldwork commenced in England and Wales in January 2008 and in Scotland and Northern Ireland in April 2008. The survey also included a follow-on survey of extending into 2009.

5.4.1. Briefings

Interviewers new to the study were briefed in 3-day training sessions. Interviewers who had worked on MCS3 were briefed in 2-day training sessions. Some of these sessions were large 'conference style' briefings'. These sessions were conducted jointly by researchers from NatCen and CLS (see NatCen 2010).

5.4.2. Fieldwork Timetable

The fieldwork timetable for MCS4 was driven by the requirement to interview the family during the child's third year of compulsory schooling (Year 2 in England and Wales, and Primary Three in Scotland and Northern Ireland). As at MCS3, fieldwork was compressed into school years. In England and Wales, the cohort's birth dates span a single school year. However, in Scotland and Northern Ireland the birth dates 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 preference in addition to date of birth.

Table 29: Fieldwork timetable for MCS4 – Main Survey

Wave	Country	Dates of birth	Fieldwork
E1	England	1 Sep 2000 – 28 Feb 2001	Jan – May 2008
E2	England	1 Mar 2001 – 11 Jan 2002	April – Aug 2008
W1	Wales	1 Sep 2000 – 28 Feb 2001	Jan – May 2008
W2	Wales	1 Mar 2001 – 11 Jan 2002	April – Aug 2008
S1	Scotland	1 Sep 2000- 28 Feb 2001 (started school in Aug 2005)	April – Aug 2008
S2	Scotland	1 Sep 2000- 28 Feb 2001 (started school in Aug 2006) and 1 Mar 2001 – 11 Jan 2002	Aug – Dec 2008
N1	Northern Ireland	24 Nov 2000 – 1 Jul 2001	April – Aug 2008
N2	Northern Ireland	2 Jul 2001 – 11 Jan 2002	Sep – Dec 2008

Table 30: Fieldwork timetable for MCS4 – Teacher Survey

Teacher Wave	Country	Main Fieldwork Wave	Teacher Fieldwork
Wave 1	England and Wales	Interviews in E1, E2, W1, W2 up to end-Apr 2008	Jun-Nov 2008
Wave 2a	Scotland and Northern Ireland	Interviews in S1 and N1 up to end-Apr 2008	Jul-Dec 2008
Wave 2b	England and Wales	Interviews in E1, E2, W1, W2 up to end-May 2008	Jul-Dec 2008
Wave 3	England, Wales, Scotland, Northern Ireland	Interviews in E1, E2, W1, W2, S1, N1 up to end-Aug 2008	Oct 2008-Feb 2009
Wave 4	Scotland and Northern Ireland	Interviews in S2 and N2 up to end-Dec 2008	Feb-Jul 2009

5.4.3. Languages

A breakdown of interviews by 'language interviewed in' is provided in the Technical Report on Fieldwork (NatCen 2010).

5.4.4. In-field Tracing

Families who had moved from the issued address were traced in the field by NatCen interviewers. Families who could not be successfully traced by interviewers were returned to CLS for additional tracing by the Cohort Maintenance Team. Details of in-field tracing activities can be found in the Technical Report on Fieldwork (NatCen 2010).

5.5. Fieldwork for MCS5

Following a competitive tender process the Ipsos MORI was appointed to carry out the fieldwork for MCS5. The first wave of the main stage fieldwork commenced in all countries in January 2012.

5.5.1. Briefings

All interviewers had a 3-day training session. In total, 23 briefings were conducted. 19 were conducted for Wave 1 (between January 2012 and February 2012). An additional 2 briefings were conducted for Wave 2 (in August 2012) and 2 mop up

briefings were conducted (one in March 2012 and one in May 2012). In total, 325 interviewers were briefed. The size of the briefings varied between regions and attendance ranged from between 13 to 21 interviewers. These sessions were conducted jointly by researchers from Ipsos MORI and CLS (see Ipsos MORI 2013).

5.5.2. Fieldwork Timetable

The fieldwork timetable for MCS5 was driven by the requirement to interview the family during the child's last year of primary schooling (Year 7 in England and Wales, and Primary Seven in Scotland and Northern Ireland). As at MCS3 and MCS4, fieldwork was compressed into school years. In England and Wales, the cohort's birth dates span a single school year. However, in Scotland and Northern Ireland the birth dates 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 preference in addition to date of birth.

Table 31: Fieldwork timetable for MCS5 – Main Survey

Wave	Country	School year in Year 6/Primary 7	Fieldwork dates
1a	All	2011-12	January 2012 – February 2013
1b	England, Wales, Northern Ireland	2011-12	March 2012 – February 2012
1c	Wales	2011-12	April 2012 – January 2013
2	Scotland	2012-2013	August 2012 – February 2013
2	Northern Ireland	2012-2013	September 2012 – February 2013

Table 32: Fieldwork timetable for MCS5 – Teacher Survey (England and Wales only)

Teacher Wave	Teacher Fieldwork
Wave 1	March-April 2012
Wave 2	April-June 2012
Wave 3	May-July 2012

Wave 4	June-July 2012
Wave 5	July-September 2012
Wave 6	July-October 2012
Wave 7	September-October 2012
Wave 8	October-November 2012
Wave 9	November 2012-January 2013
Wave 10	January-February 2013
Wave 11	January-February 2013

5.5.3. Languages

A breakdown of interviews by 'language interviewed in' is provided in the Technical Report on Fieldwork (Ipsos MORI 2013).

5.5.4. In-field Tracing

Families who had moved from the issued address were traced in the field by Ipsos MORI interviewers. Families who could not be successfully traced by interviewers were returned to CLS for additional tracing by the Cohort Maintenance Team. Details of in-field tracing activities can be found in the Technical Report on Fieldwork (Ipsos-MORI 2010). Additional tracing using administrative data was carried out by CLS.

PART 6: THE DATA OF MCS

6.1. Structure of the MCS data

The most important source about the data structure of the MCS is the questionnaires. There is a variety of questions that generate data of different structures. For example, a question that is asked from each parent of a family generates one or two pieces of information per family depending on the parent/carer availability. Similarly, if a question is asked from each Cohort Member, it can generate one to three pieces of information. This is because some families have more than one Cohort Member (twins and triplets).

The figure below visualises the link between the questionnaire and the datasets. Each questionnaire section and question contain notation on who is asked the question. For example, on whether the section is addressed to 'MAIN PARENT' or 'MAIN AND PARTNER' respondents, or whether the question is asked from both parents about each of the cohort members.

Figure 1: Interview setting -> Questionnaire -> Datasets
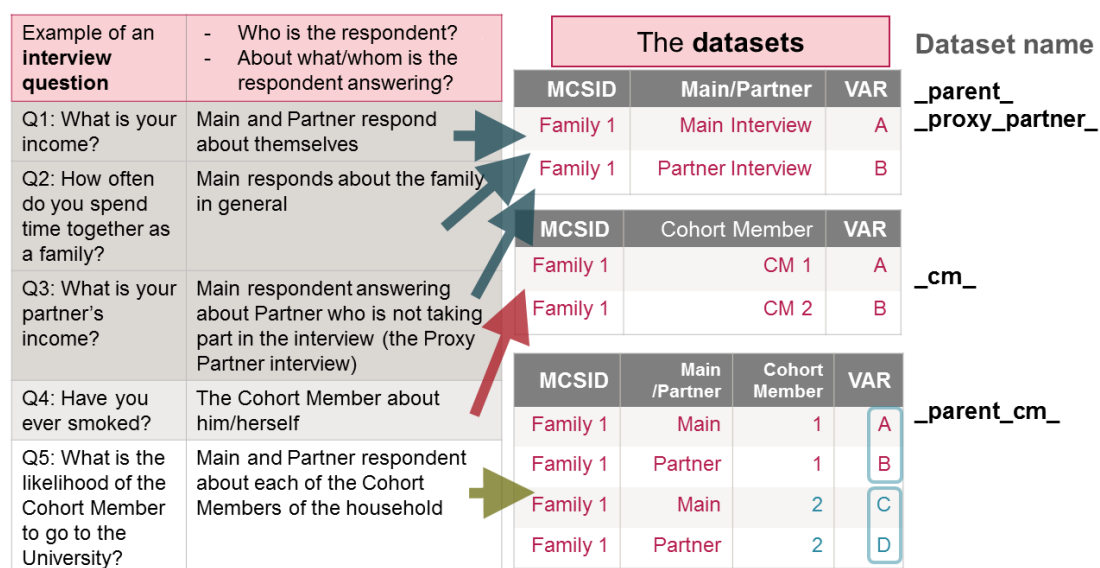
6.2. Dataset structures & dataset naming

The key structures of the datasets correspond to questionnaire information. Because of the variety of the questions asked in MCS the possible data types are primarily four (excluding the household grid).

The figure below shows the basic information for each dataset type.

family level dataset contains one row per family

_cm_ level dataset contains one row per cohort member per family

parent level dataset contains one row per parent/carer per family

_parent_cm_ level dataset contains one row per parent/carer about each cohort member per family

Dataset structures

Variables are allocated to datasets according to the type of information they hold. The questionnaire specification (to whom the question is asked and whether it loops over each CM) determines the type of information held. For example, if a question is answered by the Main or Partner respondent about themselves (Parent level dataset) or answered by the Main/Partner about each of the Cohort Members (Parent - CM level dataset). The same frame applies to derived variable datasets.

family level dataset

MCSID	VAR
Family identifier	
Family 1	A
Family 2	B
Family 3	C

_cm_ level dataset

MCSID	Cohort Member	VAR
Family identifier	Index of the Cohort Member (CNUM)	
Family 1	1	A
Family 2	1	B
Family 2	2	C
Family 3	1	D

parent level dataset

MCSID	Main/Partner	VAR
Family identifier	Variables on Main/Partner respondent (ELIG, RESP, PNUM)	
Family 1	Main Interview	A
Family 1	Partner Interview	B
Family 2	Main Interview	C
Family 2	Partner Interview	D
Family 3	Main Interview	E

_parent_cm_ level dataset

MCSID	Main/Partner	Cohort Member	VAR
Family identifier	Variables on Main/Partner respondent (ELIG, RESP, PNUM)	Index of the Cohort Member (CNUM)	
Family 1	Main Interview	1	A
Family 1	Partner Interview	1	B
Family 2	Main Interview	1	C
Family 2	Partner Interview	1	D
Family 2	Main Interview	2	E
Family 2	Partner Interview	2	F
Family 3	Main Interview	1	G

Figure 2: Overview of dataset structures and the identifiers they contain

6.2.1. The household grid

Another dataset structure that appears is the household grid. The datasets mcs*_hhgrid contain one row per person that appears in the household grid.

Family MCSID	Person Number PNUM	Cohort Member Number CNUM	Info about person & CM (age, gender, present in hh)	Relationship to Cohort Member CREL	Relationship to Person 1	Relationship to Person 2	Relationship to Person 3	Relationship to Person 4
Family 1	1		…	Parent				
Family 1	2		…	Parent				
Family 1	3		…	Sibling				
Family 1	4		…	…				
Family 1		1	…					
Family 2	1		…	Parent				
Family 2	2		…	Grandparent				
Family 2		1	…					
Family 2		2	…					
Family 3	1		…	Partner				
Family 3		1	…					

Figure 3: Overview of the household grid data structure

The household grid dataset provides information about each person who appeared in the household at the interview. It provides key demographic information about the person that lived with the Cohort Member at the time of the interview. The variable PRES shows whether the person was present at the interview. The variable CREL shows the relationship of each of the persons that are present to the Cohort Member. Finally, this datasets also provides in a 'relationships grid'. This contains the relationship of each person to the others present. This is particularly helpful for research on family structure and environment.

6.2.2   Additional dataset types

Further dataset types appear in a few sweeps. For example, the datasets that contain the survey conducted on the older siblings of the Cohort Members (key identifier: PNUM person number from household grid). Another dataset type is the teacher survey, which again has a _cm_ structure (one row per cohort member).

6.3. MCS key identifiers

The format of the different dataset types, require different identifiers that help the user manage the data for each family, parent/carer respondent (within the family), Cohort Member (within the family) and person (within the family). This section describes the identifiers and how they can be used.

The key identifiers of MCS are:

- **MCSID** is a family / household anonymised identifier and it is the same for all of its members, Cohort Members and parents.
- **CNUM** is Cohort Member number within a family. Namely, the CNUM in ascending order standing from 1 indicates the first Cohort Member within a family. The majority of the Cohort Members have CNUM=1, however, in

families with twins and triplets the second and third Cohort Member have CNUM=2 and CNUM=3 respectively.

- **PNUM** is Person number for the individuals living in the household apart from the Cohort Member(s). This includes grandparents, siblings of the Cohort Members, parents, etc. The PNUM is given in an ascending order starting from 1 randomly to each person that lives in the household that the Cohort Member lives or has lived (in a previous sweep). This means that a grandparent may be PNUM=1 and if s/he moves out in a later sweep s/he keeps the same PNUM. If another person moves in to the household or a sibling is born will receive the next available PNUM.

- **ELIG** provides the information on whether the individual has been eligible for the role of the Main or Partner respondent. For example, the mother of the Cohort Member may have PNUM=2 and the father PNUM=3, however, only the mother may have been eligible to be interviewed for the survey (ELIG=1) and not the father (ELIG=4).

- **RESP** marks whether the person selected at ELIG has participated in the interview or not (e.g. refusal). This variable is the outcome of the ELIG.

6.3.1 Parent/person identifiers: PNUM vs ELIG/RESP

PNUM and CNUM are permanent cross-sweep identifiers. Namely, they do not change between sweeps and the person or CM holds this number for the rest of the survey. Moreover, the PNUM gets assigned at random. It is likely that the parents of the CM have a low PNUM as they have been present at Sweep 1 or 2 but this is not always the case.

ELIG and RESP are sweep-specific identifiers. Namely, they mark whether the person with a specific PNUM has been selected to participate in the interview and provide information about the Cohort Member(s) in this specific sweep. The individuals eligible to respond to a survey can change from the one sweep to the other. This is why PNUM is always added in a dataset that there is a parent respondent like the _parent_ level dataset and the _parent_cm_ level dataset.

Each dataset depending on its type and structure contains a different set of identifiers.

MCSID	*PNUM*	*ELIG*	*RESP*	*CNUM*
Family identifier	Person number	Eligibility for interview	Response to Interview	Cohort Member Number
Family 1	1	Main Interview	Main Interview	Be aware of more than one Cohort Member per family (i.e. twins & triplets) like Family 1 — 1
Family 1	6	Partner Interview	Partner Interview	1
Family 1	1	Main Interview	Main Interview	2
Family 1	6	Partner Interview	Partner Interview	2
Family 2	1	Main Interview	Main Interview	1
Family 2	3	Partner Interview	Partner Interview	1

PNUM	ELIG	RESP	CNUM
Person number PNUM is the personal number of an individual that has entered the household that a cohort member lives in. It remains the same across sweeps.	**Role of the respondent** The main carers of the Cohort Member are eligible to respond to the survey either as the Main respondent or the Partner of the Main respondent.	**Availability of interview data** from the individuals who have been identified as eligible for the roles of the Main or Partner respondent (e.g. lack of data in case of refusal)	**Cohort Member number** Information about one Cohort Member will expand on two rows if there is information provided by the Main respondent and Partner (_parent_cm_ level dataset)

Figure 4: Example of how key identifiers look like in a _parent_cm_ dataset

For example, a dataset on the CM level _cm_, such as the self-completion questionnaire of the Cohort Member, contains only MCSID and CNUM and not any parent identifiers (PNUM, ELIG, RESP) since these are irrelevant. Similarly, a dataset on the parent level _parent_ contains MSCID, PNUM, ELIG and RESP but does not contain CNUM. A dataset on the _family_ level contains only the MCSID that is the family identifier. Figure 3 provides an example of the _parent_cm_ level dataset that requires all the identifiers. This is because within a family (MCSID) a parent (PNUM, ELIG, RESP) provides information about each of the Cohort Members of the family (CNUM).

6.4. Merging / Joining datasets of MCS

Merging the data of MCS is possible by using the key identifiers:

- **MCSID** (family ID),
- **CNUM** (Cohort Member number within an MCS family),
- **PNUM** (Person number across sweeps, excluding the Cohort Members),
- **ELIG** (Person eligibility as Main or Partner respondent in a certain sweep).

6.4.1 Merging between sweeps data of the same structure

Merging two datasets with the same structure it is possible by using the relevant identifiers.

Structure	Identifiers needed to merge
family	MCSID
parent	MCSID **and** a parent/person identifier (e.g. PNUM or ELIG) For selection of PNUM or ELIG as a key identifiers see the relevant section
_cm_	MCSID **and** the cohort member identifier CNUM
_parent_cm_	MCSID **and** the cohort member identifier (CNUM) **and** a parent/person identifier (e.g. PNUM or ELIG)
hhgrid	MCSID **and** a(n) row/individual identifier. This is possible by calculating a row identifier that combines the PNUM and the CNUM. Strategies on this are provided in the Data Handling guide.

6.4.2 Merging between different data structures of the same sweep

It is possible to merge data of different data structures between the same sweep.

The specific research purpose guides what is needed from the data. Based on the specific purposes, the data can be manipulated and merged to the format required.

For example, if the research project requires one row per Cohort Member, other datasets can be merged with the _cm_ dataset using MCSID and CNUM. For datasets that do not have CNUM, like _parent_ or _hhgrid_ or _family_ datasets, then it is possible to merge using only MCSID and consider that the merge will be many to many, since both dataset provide individual rows.

Let us imagine that the variable of focus (outcome or dependent variable in regression analysis) is in a _cm_ dataset and information is needed from the _parent_ information. For example, the outcome is on child's well being and the information that is used to explain the child's well being comes from a _parent_ file. In this case, there are two options:

- Many to many merge between _cm_ and _parent_ using MCSID: In this option, there will be a row per cohort member per parent found in both files for each MCSID.

- Many to 1 merge between _cm_ and coded variables in a _family_ level from the _parent_ dataset using MCSID: In this option, the researcher recodes the information that is needed from the _parent_ level dataset to one row per family. For example, aggregate income for both parents of each family or higher qualification of the parents available for interview. Then the new file with one row per family (that resembles the _family_ file) can be merged with the _cm_ file using the MCSID.

If the specific research purposes requires a _parent_cm_ structure, then researchers needed to connect the other information (e.g. _cm_ file, _parent_ file, etc) to the _parent_cm_ using the relevant identifiers. The file _parent_cm_ contains rich information with question from the parent(s) about the cohort member(s) in each family. As it contains all the person identifiers (CNUM, PNUM, ELIG), it can merge to _cm_ and to _parent_ without restructure.

Syntax and worked examples are provided in the MCS Data Handling guide.

6.4.3. Analysing long datasets with clustered design with regression

If the research design requires multiple individuals per row per family, it is possible to cluster the standard errors if the project requires it:

[STATA] reg Y X1 X2 X3 [pw=weight_variable], robust cluster(cluster_variable)

The cluster variable can be the family level MCSID, or any other cluster the researcher would like to take into account.

6.5 Finding variable(s) or datasets of MCS

MCS covers a wide range of aspects of cohort members' personal and family life.

It is important to read through the questionnaire to understand how a question was asked and what information exists in the data.

The following platforms provide overviews on the information available.

- **Questionnaires** (available in pdf format from the UKDS without logging in).
- **MCS Longitudinal Data dictionary** (available in xlsx format from the UKDS without logging in and from the CLS website). The longitudinal data dictionary contains information on the datasets, variable names and variable topics, cognitive and psychological scales that are available through the MCS data resources.
- **Closer discovery search engine** discovery.closer.ac.uk/ . It is possible to search for a question or a topic across different sweeps of MCS or another longitudinal study. For example, the user can find a topic such as 'smoking' in

different sweeps (points in life of the cohort member) or in different studies (across different cohorts, i.e. cross-cohort research).

- **Nesstar platform** nesstar.esds.ac.uk/webview/index.jsp . It is possible to search for 'MCS' and can browse through the metadata and basic statistics of all the datasets and variables available through the UKDS under End User License.

6.6 Variable naming and labelling conventions

Standard variable naming conventions are applied in the variables of MCS that make it easy for the user to find a variable. The first character of each variable corresponds alphabetically to the order of the Sweep. For example, in the datasets of the 1st sweep the variable names start with the character A (B for the 2nd sweep, C for the 3rd sweep, D for the 4th sweep, E for the 5th sweep and F for the 6th sweep).

The second character of the variable indicates the interview section: who is the respondent.

H is for household interview and household grid

X is for the proxy interview variables

P is for questions that the Main and Partner respondents are responsible for answering

C is for questions that the Cohort Member is answering on his/her own

Figure 5: Variable Naming Conventions

35

Key identifiers (PNUM, CNUM, ELIG, RESP) do not have a second character to signify the dataset they belong to. For example, for MCS6 the key identifiers are FPNUM00 FCNUM00 FELIG00 FRESP00. This means that merging within a sweep can be done without renaming the key identifiers. This convention has been applied in MCS6 and the data of MCS1-5 have been updated (2020 onwards) based on this convention.

Underscore and suffix are used in variable names to indicate a specific version of the variable, for example:

- VARNAME_R30 : R stands for recoded. In this variable categories that have counts of less than 30 have been recoded to a generic value. The original variable (without _R30) is provided in the Secure Access datasets.
- VARNAME_TR2 : TR stands for truncated. In this variable values (like SOC, SIC codes) have been truncated to the first 2 or 3 digits from the left. The original non truncated variable is provided under Secure Access.

All the variables available through End User License and through Secure Access can be found on the MCS Longitudinal Data Dictionary. This can help users find the datasets of the original variable and of the truncated or recoded version.

6.7. Abbreviations in variables labels

Abbreviation	Description
R	Response. This is used for variables where the respondent was asked a question in a loop (example, work history). The number next to R shows the order in the loop, e.g. R1 is job 1, R2 is job 2, etc
DV	Derived variable
HHQ	Household questionnaire
PX	Proxy partner questionnaire
OS	Older Siblings
S*	Sweep *, where * is the numeric order of the sweep: 1, 2, 3, 4, 5, 6, 7
MC	Multi Coded
(A) or MERGED	Combined information from original response and back-coded open text (see related section for more information)

(O) or CODED	Information from the back-coded open text only. The open text is usually an option 'Other… please write/specify' (see related section for more information)
IW or IWR	Interview or Interviewer
FF	Feed Forward information. Information collected in previous sweeps is used to inform the routing and how the question will be asked in the subsequent sweep (see dependent interviewing). For example, if a person has been diagnosed with a condition, the person will not be asked the same question again.

The main **abbreviations** used in **variable labels** are: R1, R2, etc, or MC, or (A) and (O). In R1, R2 and so on, R stands for Response and it applies to questions where the respondent has the option to provide more than one responses in a loop. When the question is in a loop, an R with a number has been added in the label of each variable.

For example, in the questions about absent or non-resident parents of the CM, the Main respondent can provide information on whether the absent / non-resident parent is in contact with the CM, for each absent / non-resident parent, R1 for the first absent parent, R2 for the second absent parent, etc.

6.7.1 Multi-coded & Back-coded variables

The abbreviation **MC** in variable labels stands for **Multi-Coded**.

These are the questions that include the instruction 'Select/Code all that apply'. In these questions the participant may have given more than one response.

These have been stored in the data and the variable label is either MC1 MC2 MC3 or MC [Sub-label 1] MC [Sub-label 2] etc.

The first type of labelling MC1 MC2 MC3, and so on, suggests that the first selection of the participant is stored in the first variable, the second selection in the second variable and so on. This means that the variable that has label MC1 includes what the respondents selected as part of their first response.

The second type of labelling MC [Sublabel 1] indicates that this is a binary variable (e.g. yes / no) on whether that specific option of the sublabel has been selected or not.

The abbreviations **(O)** or **(A)** are used when the question had an open text element. This is when the questionnaire instructs: 'Other, please specify' or 'Write in'.

The text that the interviewer writes in gets back-coded (categorised / classified) by the survey agency. These back-coded responses are given a (O) in their variable label. The detailed coding matrix used for the back-coding for each sweep is available through the technical report of each sweep. Also, the questionnaires provide the exact wording of the response options that was used during the interview.

They are the outcome of categorising and classifying the open text that has been provided.

In some questions the original response in the Multi-Coded question has been collapsed with the back-coded response given in the 'Other...' prompt into one variable. This variable holds all the information provided in the question either by selecting one of the questionnaire options or by providing an open text response that has been back-coded. It has been marked as (A) or 'MERGED' in the variable label.

Figure 6: Example of merging Multi-Coded & open text (back-coded) variable

6.7.2 How to use back-coded variables

Although it depends on the specifics of a research project, it is recommended to use the combined information ((A) or MERGED) wherever it is available. However, in a few variables the information available may be in two different sources: response to the original variable (no label marking) and the back-coded information of the open text question ((O) or CODED). In this case the data user may decide to combine the information of the two variables in a combined information form.

PART 7: THE DATA OF SURVEYS 1-5

Researchers can find the respective variables for the cognitive scales

7.1. Cohort Member Cognitive Assessments

A number of cognitive assessments have been administered to the MCS cohort members since they were aged 3. The following assessments were administered to the MCS cohort members at different sweeps:

Assessments by Sweep Collected

Assessment	MCS Sweep			
	MCS 2	MCS 3	MCS 4	MCS5
BAS Naming Vocabulary	X	X		
Bracken School Readiness	X			
BAS Picture Similarity		X		
BAS Pattern Construction		X	X	
BAS Word Reading			X	
BAS Verbal Similarities				X
NFER Number Skills			X	
CANTAB Spatial Working Memory Task				X
CANTAB – Cambridge Gambling Task				X
Sally Anne assessment		X	X	
Our Adventures (Wales only)			X	

7.1.1 The British Ability Scales

The British Ability Scales (BAS) is a battery of individually administered tests of cognitive abilities and educational achievements suitable for use with children and adolescents aged from 2 years 6 months to 7 years 11 months.

1) BAS Naming Vocabulary (MCS 2 and MCS 3)

Following consultation with advisers and piloting, the BAS Naming Vocabulary scale was administered by interviewers to cohort members during the MCS2 data collection.

The Naming Vocabulary is a verbal scale for children aged 2 years 6 months to 7 years 11 months. It assesses the spoken vocabulary of young children. The test items consist of a booklet of coloured pictures of objects which the child is shown one at a time and asked to name. The scale measures expressive language ability, and successful performance depends on the child's previous development of a vocabulary of nouns. Picture recognition is also crucial; however, the pictures are large and brightly coloured and are unlikely to cause problems except for children with major visual impairments or with no experience of picture books. The items require the child to recall words from long-term memory rather than to recognise or understand the meaning of words or sentences.

2) Scores

Naming Vocabulary scores may reflect:

- Expressive language skills
- Vocabulary knowledge of nouns
- Ability to attach verbal labels to pictures
- General knowledge
- General language development
- Retrieval of names from long-term memory
- Level of language stimulation.

Low scores may reflect reluctance to speak.

3) BAS Picture Similarity (MCS3)

Children are shown a row of 4 pictures on a page and asked to place a card with a fifth picture under the picture most similar to it. This assessment measures children's problem solving abilities.

4) BAS Pattern Construction (MCS3 and MCS4)

The child constructs a design by putting together flat squares or solid cubes with black and yellow patterns on each side. The child's score is based on accuracy and speed. This assessment tests spatial awareness but can also be used to observe dexterity and coordination, as well as traits like perseverance and determination.

5) BAS Word Reading (MCS 4)

Word Reading is an assessment from the British Ability Scales: Second Edition (BAS 2) which assesses children's English reading ability.

The child reads aloud a series of words presented on a card. The assessment consists of 90 words in total. The words are organised into 9 blocks of 10 words in ascending order of difficulty. The child is asked to read each word in a block out loud to the interviewer. The number of blocks of words the child is asked to attempt to read is dependent on the child's performance during the assessment. This assessment is designed to be used with children aged from 5 years to 17 years and 11 months. All of the children in MCS4 started at the first item, as this was the starting point for children of their age.

A child's progression through the assessment is dependent on the number of words they read correctly. If a child makes 8 errors in a block of 10 words, then the assessment stops.

In Wales a different test was carried out (see the relevant section below).

6) BAS Verbal Similarities (MCS5)

Verbal Similarities is an assessment from the British Ability Scales: Second Edition (BAS 2) which assesses children's verbal reasoning and verbal knowledge.

The interviewer reads out three words to the child who must then say how the three things are similar or go together.

This assessment is designed to be used with children aged from 5 years to 17 years and 11 months. All of the children in MCS5 start at the 16th item, as this is the starting point for children of their age. There are decision points after items 28 and 33 where the child's performance so far decides whether the test stops or continues to the next set of questions. The test stops at the decision point unless the child has less than three failures on all items so far. In this case they are routed to the next set of questions. If the child has obtained less than three passes however, they are routed back to the previous starting point (e.g. item 8).

After five consecutive failures the test is automatically stopped provided that at least three items have been passed prior to this, otherwise they are routed back to the previous starting point.

If the child fails either of the first two items administered they are provided with teaching to help them to understand the concept of the test. If the child subsequently gives a correct answer to the same question it is acknowledged but they do not receive a point for that question.

7) Scores for the BAS assessments

There are three types of score provided for each scale of the BAS: raw score, ability score and T-scores or standardised scores. Each type has its uses and limitations.

Raw Scores

Raw scores are simply the number of items the cohort member child answered correctly. They do not take into account the stop and start points of the items administered; for this reason, the raw scores have little meaning and should not be used.

Ability Scores

The ability scores are a transformation of the raw scores that take into account the specific item set administered. They are not adjusted for anything else, so are the scores to consult for unadjusted cognitive scores.

There are some issues to keep in mind when using ability scores. The first is that it not a truly continuous scale. The table below shows the correspondence between some example raw scores and ability scores. As can be seen from this table, there are ability scores that cannot be obtained.

Correspondence of raw scores to ability scores for an example BAS scale:

Raw Score	Ability Score
4	26
5	30
6	33
7	37
8	40
9	43

For convenience, the ability scores for each scale start with a value of 10, which reflects a raw score of 0 on the easiest possible set of items in a scale. The upper limit of ability scores varies from scale to scale. Because the ability scale uses an

arbitrary numbering system, comparing ability scores from different scales is not meaningful, just as comparing raw scores from different scales is not meaningful.

The other issue is that the ability scores are not adjusted for age. Children of a large range of ages take the same BAS tests, and the general trend is that older children score higher. When using ability scores, one should control for child age. The issue of age and the BAS scales is discussed in further detail in the section below on BAS Scales and Age.

T-Scores and Standardised Scores

Also available for all scales are T-scores or a standardised score. These scores are adjusted for the cohort member child's age group and for the mean scores of the BAS norming group. They are computed using the BAS manual's conversion tables. For each 3-month age group, there is a table showing the conversion of ability scores to T-scores or standardised scores. The T-scores have a mean of 50 and standard deviation of 10 within the norming sample of a given age group. A cohort child who has an ability score that is the same as the mean for the norming group in his or her age group will have a T-score of 50. A child with a T-score of 60 had an ability score that was one standard deviation above the norming sample mean for his or her age group.

All of the scales used with the MCS sample in sweeps 2 through 4 have T-scores, with the one exception of Word Reading at MCS 4. That scale has a standardised score rather than a T-score. The only difference between the standardised score and the T-scores is that the former does not have a mean of 50 and standard deviation of 10. It is otherwise computed the same as the T-score, adjusting for age group and norming sample mean and standard deviation.

There are pros and cons to using T-scores or standardised scores. While these scores take into account child age, they are based on 3-month age groupings of the norming sample. They don't take into account the score variation with each group of 3 months. They also are based on the relationship between age and score in the norming sample rather than within the MCS sample. Using the age of the MCS sample one is using as a control will be a more accurate adjustment for age than using the T-scores (see the section on age equivalence below for more information). However, if one is looking at univariate relationships and cannot control for MCS child age, it could be beneficial to use the T-scores or standardised scores, especially in cases in which the variables of interest may be related to child age.

As the T-scores and standardised scores remove the mean and standard deviation of the norming sample from each score, they may hide differences in variance at different ages. If one is interested in how variance in BAS scores differs across age or sweep, one may want to avoid using the T-scores or standardised scores so that the actual variance in the sample is clear.

43

As was the case for the ability scores, the T-scores and standardised scores are not truly continuous.

8) Further information

For more information about the development, administration, scoring and interpretation of the BAS scores see:

Elliott, C.D., Smith, P, and McCulloch, K (1996). British Ability Scales Second Edition (BAS II): Administration and Scoring Manual. London: NFER-Nelson.

Elliott, C.D., Smith, P, and McCulloch, K (1997). *British Ability Scales Second Edition (BAS II): Technical Manual.* London: NFER-Nelson.

7.1.2 The Bracken School Readiness Score (MCS2)

The Bracken Basic Concept Scale – Revised (BBCS-R) is used to assess the basic concept development in children in the age range of 2 years 6 months to 7 years 11 months. BBCS–R measures the comprehension of 308 functionally relevant educational concepts in 11 subtests or concept categories. Following consultation with advisers and piloting, only subtests 1-6 were administered by interviewers to the members of the cohort during the MCS2 data collection.

The sub-tests administered together form the Bracken School Readiness Assessment (BSRA) which evaluates 88 concepts relating to:

- *Colours*: represents both primary colours and basic colour terms.
- *Letters*: measures knowledge of both upper- and lower-case letters.
- *Numbers/Counting*: measures recognition of single- and double-digit numbers, and samples the ability to assign a number value to a set of objects.
- *Sizes*: includes concepts that describe one, two, and three dimensions.
- *Comparisons*: measures ability to match and/or differentiate objects based on one or more of their salient characteristics.
- *Shapes*: includes one-, two-, and three-dimensional shapes. The one-dimensional category includes linear shapes; two-dimensional shapes are represented by concepts such as the circle, square, and triangle; and three-dimensional shapes include concepts such as the cube and pyramid.

The readiness concepts assessed in these sub-tests are argued to be directly related to early childhood education and to predict readiness for more formal education. The test is individually administered and suitable for children aged 2 years 6 months to 7 years 11 months.

The "Raw scores" variables in the datasets contain the total number of correct answers for each of the six BRSA sub-tests.

Percentage mastery: The raw score as a percentage of the maximum possible score for each sub-test.

School Readiness Composite Percentage mastery: The total number of correct answers in all 6 sub-tests as a percentage of the maximum possible score (bdsrcm00).

1) Age-adjusted scores

The following variables are derived from bdbsrc00 (School Readiness Composite) which is the total number of correct answers adjusted for age.

School Readiness Composite Standard Score: The total number of correct answers on all six sub-tests (bdsrcs00).

Normed scores: Derived from standard tables in the BSRA manual and defined with reference to the standardisation sample used in developing the assessments. The standardisation sample was composed of 1,100 children aged between 2 years 6 months and 8 years 0 months representative of the general US population and was stratified by age, gender, race/ethnicity and parental education:

- *Standard Scores* with mean of 15 and standard deviation of 15.
- *Percentile ranks* representing the percentage of children in the normative sample who are ranked at or below the child's score (bdsrcp00).
- *Descriptive Classification* of the normed scores into five categories: Very advanced; Advanced; Average; Delayed; and Very delayed (bdsrcn00).

2) Further information

For more information about the development, administration, scoring and interpretation of the BSRA see:

Bracken, B.A. (2002). *Bracken School Readiness Assessment: Administration Manual.* San Antonio, Texas: Psychological Corporation.

7.1.3 NFER Number Skills (MCS4)

This test was adapted from the NFER Progress in Maths test which is aimed for 7-year-olds and was originally developed and nationally UK standardised in 2004. The whole test has a maximum raw score of 28. The national mean raw score in 2004 was 19.3 with a standard deviation of 5.3. The scores were nationally age standardised to a mean of 100 and SD of 15.

The edition of this test used in the MCS is an adaptive version of the test created by Cres Fernandes of NFER. All children have to complete an initial test and based on their score they are routed to easier, medium or harder sections. The sections were

devised to save administration time, as it means each child completes around half the original number of questions.

An item response scaling method (Rasch) was used to scale the results of the easy, medium and hard subtest scores to the equivalent original raw scores. The variable **maths7scale** can be considered to be the estimated raw score based on the original test. The variable **maths7sas** is the standardised age adjusted score based on the national standardisation lookup tables in 2004.

7.1.4 Our Adventures (MCS4) – Wales Only

Our Adventures is part of the All Wales Reading Test, which was developed in Wales to assess the reading skills of children in Welsh schools. The test is available in Welsh and English.

In MCS4, parents of children in Wales were given the option of having their child's reading skills assessed in either Welsh or English. The Welsh version of Our Adventures was used for children whose parents opted for the Welsh medium to be used, and the Word Reading assessment was used for children whose parents opted for the English medium to be used.

It was decided to use the Welsh medium All Wales Reading Test, rather than a Welsh translation of the Word Reading assessment because the Word Reading assessment is designed only to assess English reading ability and if translated the results are not valid.

The Our Adventures assessment is a paper booklet that shows a story in pictures and words; underneath each picture is a sentence that has one missing word, and a list of words that can complete the sentence. The child has to circle the word that best completes the sentence. There are a total of 59 items, and the assessment has a time limit of 30 minutes. The assessment continues until the time limit has been reached, or the child completes the last item.

This assessment is designed to be used with children from age 6 years 10 months to 9 years 9 months.

7.1.5 Sally and Anne Task (MCS3 and MCS4)

The Sally and Anne Task was first used in the Millennium Cohort Study at MCS3 (age 5) and is based on a social cognition, or false belief, task developed by Wimmer and Perner (1983)[7] and subsequently modified by Baron-Cohen et al (1985).[8] The

---

[7] Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and the containing function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103-128.
[8] Baron-Cohen S, Leslie AM, Frith U (1985). Does the autistic child have a 'theory of mind'? *Cognition*, *21* (1): 37-46

latter modification has become the standard version used and involves puppets used to act out a story for the child. The child is introduced to a character, Sally, who leaves a desirable object such as a ball in her basket, before leaving the scene. In her absence, another character, Anne, removes the object and places it in a box. Children are asked to predict, on Sally's return to the room, where Sally will look for the object (or, sometimes, where she thinks the object is – the belief question). In addition, children are asked two control questions: a reality question (where is the object, really?) and a memory question (where did Sally put the object at the beginning?).

In MCS the task was adapted for use in a home survey setting: instead of puppets, pictures were used, and the interviewers followed a script that was written in the CAPI. At MCS3, the Sally and Anne task was mainly used to train the cohort child (to help them engage with the kinds of activities that would be involved in the cognitive assessments) and to develop rapport with the interview.

The results of Sally and Anne at MCS3 showed that only 20% of the cohort children got the correct answer to the key belief question, which was much lower than anticipated (the expected result is in excess of 80% for children with no developmental problems; Baron-Cohen (1985)). This led the survey team to question whether the results had been affected by the change in mode. As a result the task was repeated at MCS4 (age 7) to see whether the cohort children's belief systems had changed in the intervening two years. The result from age 7 showed that the majority (67%) of MCS cohort children still answered the belief question incorrectly. **These out of range results have led the survey team to conclude that the change in mode caused the unusual results in the MCS cohort and therefore the data from the Sally and Anne Task should be used with extreme caution.**

7.1.6 CANTAB Spatial Working Memory (MCS5)

The Memory task is a touch-screen assessment that tests the child's ability to retain spatial information and to manipulate remembered items in working memory. It also assesses use of strategy. The aim of this test is that, by process of elimination, the child should find one blue 'token' in each of a number of coloured boxes displayed on the screen and use them to fill up an empty column (black hole) on the right hand side of the screen. To see if a blue token is beneath a coloured box, the child has to touch it with their index finger. If a blue token is revealed to be beneath a coloured box, the child moves it to the black hole by touching the black hole with their index finger. Touching any box in which a blue token has already been found is an error, as is touching any box which has been found to be empty while searching for the same token. The child decides the order in which the boxes are searched. Performance at the harder levels of this task is enhanced by the use of a search strategy. The number of boxes is gradually increased from three to eight boxes. The colour and position of the boxes used are changed from trial to trial to discourage the use of the same search strategies from trial to trial.

The child's overall score is calculated from three different aspects of their performance: errors, strategy and latency. Their performance is scored on each of the assessed trials.

- Errors are the number of times the child revisits a box which has previously been found to be empty or in which a token has been previously found.
- Strategy is the order in which the child decides to search the boxes. On the harder levels the child will perform better if they make use a search strategy.
- Latency is calculated from three different measures of 'time taken'. They are the average time the child takes to first touch the screen when a new trial is presented, the average time the child takes between when they place the token in the black hole and the next time they touch a box and the average time it takes the child to find the final token from the time each trial was presented on screen.

See http://www.camcog.com/spatial-working-memory.asp for more information.

7.1.7 CANTAB Cambridge Gambling Task (MCS5)

This task assesses decision-making and risk-taking behaviour. Unlike other 'Gambling' tasks, CGT dissociates risk taking from impulsivity. On each trial, the participant is presented with a row of ten boxes across the top of the screen, some of which are red and some of which are blue. At the bottom of the screen are rectangles containing the words 'Red' and 'Blue'. The participant must guess whether a yellow token is hidden in a red box or a blue box and gamble points based on their confidence with this choice. There are six outcome measures which cover risk taking, quality of decision making, deliberation time, risk adjustment, delay aversion and overall proportion bet (for more information see: http://www.camcog.com/Cambridge-Gambling-Task/Decision Making and Response Control)

7.1.8. General influences on test scores

It is important to note that the child's performance may have been affected by influences extraneous to those that the assessment is intended to measure. The conditions listed below can lead either to a higher or lower score than would normally be obtained.

Non-standard administration of the scale	The specification of CAPI program and the training of interviewers was designed to ensure standard administration of the assessment.
Non-standard scoring	Scoring algorithms used ensure standard scoring in all cases.
Administration disrupted by noise or other interruptions **Difficulty in establishing rapport with the child** **Child has difficulty in concentrating on the tasks or is easily distracted** **Child is excessively anxious to the extent that concentration/flexibility of thought seems impaired** **Child is reluctant to respond and/or refuses to persevere on more difficult items** **Child has permanent/temporary sensory impairment (particularly vision/hearing) or motor impairment** **Child is on medication of a type that could affect performance** **Child is overtired or ill**	The training of interviewers was designed to ensure that risks were minimised. Details of any interruptions, distractions, behaviours, health circumstances, etc., were recorded in CAPI by interviewers. See table below.

7.2. Cohort Member Behavioural Development

The SDQ is a behavioural screening questionnaire for 3- to 16-year-olds. It measures **25 items on psychological attributes** (for information on other scales see Psychological and developmental inventories, Johnson, 2012).

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

i) Emotion Symptoms Scale

1. Complains of headaches/stomach aches/sickness
2. Often seems worried
3. Often unhappy
4. Nervous or clingy in new situations
5. Many fears, easily scared.

ii) Conduct problems

1. Often has temper tantrums
2. Generally obedient*
3. Fights with or bullies other children
4. Can be spiteful to others
5. Often argumentative with adults.

ii). Hyperactivity Scale

1. Restless, overactive, cannot stay still for long
2. Constantly fidgeting
3. Easily distracted
4. Can stop and think before acting*
5. Sees tasks through to the end*.

iv) Peer Problems

1. Tends to play alone
2. Has at least one good friend*
3. Generally liked by other children*
4. Picked on or bullied by other children
5. Gets on better with adults.

v) Pro-social Scale

1. Considerate of others' feelings
2. Shares readily with others
3. Helpful if someone is hurt, upset or ill
4. Kind to younger children
5. Often volunteers to help others.

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

Each of the 5 sub-scales can be used alone or together to create:

- 1-4 when taken together generates a **total difficulties score**.
- 1 and 4 create an internalising problem score.
- 2 and 3 create an externalising conduct score.
- 5 alone measures pro-social behaviour.

SDQ derived variables are available in the derived variables datasets (_cm_ dataset since there is one row per cohort member). More information is available in the derived variables user guide.

For more information about the scoring and interpretation of the Strengths and Difficulties Instrument see:

Goodman, R. (1997). 'The Strengths and Difficulties Questionnaire: A Research Note.' *Journal of Child Psychology and Psychiatry.* 38: 581-586.

 Goodman, R. (2001), 'Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ).' *Journal of the American Academy of Child and Adolescent Psychiatry.* 40: 1337-1345.

Goodman, R., Meltzer, H. and Bailey, V. (1998). 'The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version.' *European Child and Adolescent Psychiatry.* 7: 125-130.

7.3. Cohort Member Physical Measurement

7.3.1. Height

The original height variables – *HTCM* for sweeps 2, 3 and 4 – have not been edited.

Copies of the variables were made – *DHCMC for sweeps 2 and 3 and *CHTDV* for sweep 4 – and appropriate changes were made to them as follows:

- Where interviewer notes gave clear warnings that the height values entered were incorrect, the values were removed.
- Where the interviewer notes gave a value to replace an incorrect entry, these were changed.
- The variables *DHTAM* **of sweep 2 and 3** are flags to show if any changes were made. There were very few interviewer comments at MCS4 relating to measurements.

A variable was included to categorise the "outcome" of each height measurement based on the amended variables. The original variables *YHTRL* pf sweeps 2 and 3 record whether or not the measurement was "successful." However, many cases

where this was flagged as "unsuccessful" still have a height measurement recorded. Researchers may wish to use the variables *DHTOC* of sweeps 2 and 3 to exclude these cases (as well as cases containing historic data) from analyses.

At MCS4, the variable ***CHTIS* of sweep 4** indicates whether "measurement circumstances" (of variable *CHTRZ* A to D) and/or "other information" (*CHTEXT* A to B of sweep 4) was given in relation to the height measurement, and flags up the highest and lowest 100 or so values where no other circumstances are mentioned.

At MCS5 the height measurement is provided in variable *CHTCMA*. Variable *CUNHT* gives reason why height measurement was not taken (where applicable), and there are two sets of binary variables documenting circumstances that may have affected the height measurement (*CHTRX* from A to Q and *CHTEX* from A to P of sweep 5).

7.3.2. Weight

The original weight variables – *YWTCM* and *YWTGM* of sweeps 2 and 3 and *CWTCM* of sweep 4– were not edited.

Copies of the variables were made – *DWTKC* of sweeps 2 and 3, and *CWTGC* of sweep 4 – and appropriate changes were made to them as follows:

- Where interviewer notes gave clear warnings that the weight values entered were incorrect, the values were removed from *DWTKC* of sweeps 2 and 3, and *CWTDV* of sweep 4.
- Where the interviewer notes gave a value to replace an incorrect entry, these were changed in *DWTKC* and *DWTKC* of sweeps 2 and 3. There were very few interviewer comments at MCS4 relating to measurements.
- The variables *DWTAM* of sweeps 2 and 3 are flags to show if any changes were made.

A variable was included to categorise the "outcome" of each weight measurement, based on the amended variables. Researchers may wish to use the variables *DWTOC* of sweeps 2 and 3 to exclude cases containing historic data from analyses.

At MCS4, the variable *CWTIS* indicated whether "measurement circumstances" (*CWTRZ* from A to D of sweep 4) and/or "other information" (*CWTEX* from A to B of sweep 4) was given in relation to the weight measurement, and flags up the highest and lowest 100 or so values where no other circumstances are mentioned.

At MCS5 the weight measurement is provided in variable *CWTCMA*. Variable *CWTUN* of sweep 5 gives reason why weight measurement was not taken (where applicable), and there are two sets of binary variables documenting circumstances that may have affected the weight measurement (*CWTRL* from A to G and *CWTEX* from A to M of sweep 5).

7.3.3. Height, weight and BMI

The amended height and weight variables were used to calculate BMI.

The formula to compute BMI is weight (in kilos) divided by height squared (height measured in metres). This is computed for cases where we have a valid value given for both height and weight, and will be missing if either or both measurement is missing.

7.3.4. Outliers

All height and weight observations have been included in the data, even where they might be considered outliers. All observations have been used to calculate the BMI measure. We leave it to individual researchers to take decisions on whether they consider any of the measurements to be outliers and what they do with such observations. Users should be warned that the dataset contains a few values that other users have considered implausible.

7.4. Income data

The MCS has collected income in a number of different ways over the different sweeps. At sweeps 1-5 income data were collected in a single banded question in addition a set of detailed questions which collected information on a range of different measures detailed in the Table below.

Table 38: Income data collection across the cohort studies

Income	MCS 1		MCS 2		MCS 3		MCS 4		MCS5	
	Main	Partner	Main	Partner	Main	Partner	Main	Partner	Main	Partner
Gross Earnings	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Net Earnings	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Usual net Earnings					✓	✓	✓	✓	✓	✓
Earnings from second job	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Irregular earnings from occasional work	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Earnings from Self-employment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Housing benefit	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Income	MCS 1	MCS 2	MCS 3		MCS 4		MCS5	
Child benefit	-	-	✓	✓	✓	✓	✓	✓
Guardian's Allowance	-	-	✓		✓	✓	✓	✓
Carer's allowance	-	-	✓	✓	✓	✓	✓	✓
State pension	-	-	✓	✓	✓	✓	✓	✓
Widow's pension	-	-	✓	✓	✓	✓	✓	✓
War disablement allowance	-	-	✓	✓	✓	✓	✓	✓
Severe Disablement Allowance	-	-	✓	✓	✓	✓	✓	✓
Disability Allowance	-	-	✓	✓	✓	✓	✓	✓
Job seekers allowance	-	-	✓	✓	✓	✓	✓	✓
Pension credit	-	-	✓	✓	✓	✓	✓	✓
Income support	-	-	✓	✓	✓	✓	✓	✓
Incapacity benefit	-	-	✓	✓	✓	✓	✓	✓
Working tax credit	✓	✓	✓	✓	✓	✓	✓	✓
Child tax credit		✓	✓	✓	✓	✓	✓	✓
Child care tax credit	✓	✓						
Statutory sick pay	-	-	✓	✓	✓	✓	✓	✓
Grant from the social fund for maternity expenses	-	-	✓	✓	✓	✓	✓	✓
Other social fund grant	-	-	✓	✓	✓	✓	✓	✓
Maternity Allowance	-	-	✓		✓		✓	
Statutory Maternity Allowance		-	✓		✓		✓	

Income	MCS 1		MCS 2		MCS 3		MCS 4		MCS5	
Child Maintenance Payment	*		*		✓		✓		✓	
Other Regular Payments**	✓		✓		✓		✓		✓	

✓ = Collected in full with respondents reporting amount.

- = Partially collected, no amount given.

* = Collected as an option of other regular payments.

** = Includes: education grants/student shops or work; training/government training scheme; employers maternity/paternity pay; maintenance allowance or other regular; regular cash help from parents; regular cash help from other relatives; rent from boarders, lodgers or sub-tenants; other income from organisations; pension from a former employer; income from investments; allowance for a foster child. Monthly amount given as a total from these sources.

7.4.1 Banded data

Respondents were shown a card with weekly, monthly and annual bands of total take-home income from all these 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 received as separate components, so they may well have been omitted from estimates of total net income as reported. Bands of different sizes were used for lone and 'couple' families.

7.4.2. Missing income data (item non-response)

Analysis of the collected data shown in the Table below indicates that more than 1,500 of MCS families, at each sweep, do not provide banded income data either by saying they didn't know their family income or refusing.

Imputation of missing and continuous income from banded data

We imputed income for the cases where it was missing using interval regression (Stewart 1983). This method allowed us to impute a continuous value within a band where income band was available, 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 in our case, *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 given in the following table.

Variable	Categories
Main respondent's age at interview	Continuous
Housing tenure	Own
	Private renting
	Renting from Local Authority or Housing Association
	Other
DV combined labour market status of main and partner respondents	Both in work/on leave
	Main in work/on leave, partner not in work/on leave
	Partner in work/on leave, main not in work/on leave
	Both not in work/on leave
	Lone parent in work/on leave,
	Lone parent not in work/on leave
Point type	Advantaged
	Disadvantaged
	Ethnic
DV interview government office region	North East
	North West
	Yorkshire and the Humber
	East Midlands
	West Midlands

Variable	Categories
	East of England
	London
	South East
	South West
	Wales
	Scotland
	Northern Ireland
Receipt of state benefit?	No
	Yes
Main respondent's ethnic group – 6 category census classification (UK)	White
	Mixed
	Indian
	Pakistani and Bangladeshi
	Black or Black British
	Other ethnic group (inc. Chinese and other Asian)
DV combined education highest NVQ	NVQ level 1
	NVQ level 2
	NVQ level 3
	NVQ level 4
	NVQ level 5
	Overseas qual only
	None of these
Main type of accommodation	A house or bungalow
	A flat or maisonette

Variable	Categories
	A studio flat
Number of children including cohort child	1
	2
	3
	4+
DV summary of parents/carers in household	Two parents/carers
	One parent/carer

Equivalisation

We used modified OECD scales 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 such children 1.07; and so on. This is shown below.

Table 40: OECD household equivalence scales

Equivalence scales before housing cost	OECD scale*
First adult (Main respondent)	0.67
Spouse	0.33
Dependent child age between 14<=18 years old (16<=18 for McClements)	0.33
Child aged under 14 years (<16 for McClements)	0.20

* The user guides to initial findings of MCS 1 and MCS 2 used a simplified version of this scale where all children under 16 years were given a score of 0.23

References

- Conroy, R.M. (2005). 'Stings in the tails: Detecting and dealing with censored data.' *Stata Journal*. 5: 395-404.
- DWP (2007). Households below average income. 1194/5-2005/6.
- (http://www.dwp.gov.uk/asd/hbai/hbai2006/pdf_files/chapters/chapter_4_hbai07.pdf)
- Hansen, K. (2008). *Millennium Cohort Study First, Second and Third Surveys: A Guide to the Datasets*, Third Edition.

(http://www.cls.ioe.ac.uk/studies.asp?section=00010002000100110002http://www.cls.ioe.ac.uk/studies.asp?section=00010002000100110002; accessed on 16/05/2008)

- 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)
- StataCorp. (2007). Stata Statistical Software: Release 10. College Station, TX: StataCorp Lp.
- Stewart, M.B. (1983). 'On least-squares estimation when the dependent variable is grouped.' *Review of Economic Studies.* 50: 737-753.

7.5. The Cohort Member's Self-completion

As the cohort members grow up the focus of the survey increasingly is focussed on them. From MCS4 it was appropriate to gain the cohort member's own views on their developing lives. The cohort members were given their own self completion questionnaire at MCS4 (age 7). This involved a short, easy to read, 8 page paper self completion which the interviewer gave to them during the home visit. It took around 20 minutes to complete. At Age 7 the age appropriate topics covered included: Hobbies, friends and family, feelings and school.

At MCS5 (age 11) the self completion questionnaire was extended significantly to reflect the greater complexity if the cohort member's lives and their ability to answer a longer, more complex instrument. The age 11 questionnaire was 28 pages long and took around 30 minutes to complete. Once again it was a self completion paper questionnaire. The topics at age 11 included:

- Activities outside school
- Internet & social networking
- Life satisfaction
- Happiness
- Self esteem
- Friends
- Unsupervised time
- Pocket money
- Family financial position & materialism
- Anti social behaviours,
- School
- Secondary school
- Attitudes
- Other children (incl. bullying)
- Risky behaviours (incl. smoking & alcohol)
- Mental health

- Future ambitions

At MCS5 the questionnaire was offered in audio assisted mode to cohort members who had lower levels of literacy using an MP4 player. However, less than 2% (1.8%) of cohort members completed the self completion with audio support.

7.6. Feed Forward Data

Some information was fed forward from earlier sweeps. The feed forward data were associated with the Main respondent and the Partner respondent from the previous sweep. It was fed forward into the MCS2, MCS3, MCS4 or MCS5 interview if the interviewer indicated that the Main respondent was the same as at the previous sweep or that the Partner respondent was the same as at the previous sweep . The name of the Main respondent and Partner at the previous sweep was made available for interviewers. In some cases, the interviewer coded variable is discrepant with the derived variable indicating There are derived variables indicating this and this implies that information was fed forward into the Main/Partner interview because the CAPI thought the respondents were the same but in fact the respondents were different.

7.7. Data Cleaning

7.7.1. MCS1 Data Cleaning

Details of coding and editing activities can be found in the Codebook and Edit Instructions prepared by NatCen, included in this deposit and their Technical Report on Fieldwork (NatCen 2004).

7.7.2. MCS2 Data Cleaning

MCS2 data were received from NOP in SPSS format. The data went through an extensive process of restructuring to produce the current datasets.

Because the Household Grid information was not fed forward from MCS1, construction of the current household grids had to be carried out by a process of matching each individual recorded at MCS2 with that at MCS1. Twenty-eight per cent of individuals did not require matching because the family did not take part in MCS2, either they were new families entering the study for the first time, or younger siblings of the cohort member.

Of the remaining, 37 per cent matched on name, sex and date of birth. Cohort members matching on full name accounted for another 18 per cent (date of birth of cohort members was not re-collected unless there was a discrepancy with that fed forward). Of the remaining 17 per cent, 9 per cent were either new entrants or leavers from the household and 6 per cent matched on full name only. The remaining 3 per cent matched on less reliable measures. The full list of matches was checked

by eye to reveal any discrepancies (false positives and false negatives). A fuller analysis of how this compared to the final cleaned data used in the deposited data will appear in due course.

Data for child assessments, child measurements and home observations also needed to be matched as their number in the household was not passed between instruments. As there are only a relatively small number this was done by hand.

7.7.3. MCS3 Data Cleaning

Household grid information was fed forward from the interview at MCS1 and MCS2. Where responses conflicted, the value used was a majority where response from the previous two surveys were available or the latest where only one previous interview was available for comparison. Checks were also applied to investigate implausible or unlikely values, grandmothers under 30, natural siblings more than 30 years apart, etc.

7.7.4. MCS4 Data Cleaning

Essentially the same strategy was employed as at MCS3, but with the addition of more checks on inter-family relationships with an emphasis on relationship to the cohort child and the main and partner respondents.

7.7.5. MCS5 Data Cleaning

The data collection was split across three instruments for the main and partner respondent elements. This was initially reconciled by IPSOS-MORI and further integrity checks were then conducted by CLS. This was complicated by the use of a different person ID to that used at CLS, resulting in a mapping exercise between those used in the data collection and that seen in the output data. The unintended consequence of this was that the household grid and subsequent relationships needed to be re-organised, this was done at CLS. Checks were constructed to ensure that the people present in the household are longitudinally consistent, through checks on date of birth, sex, and relationship to Cohort Member. As is the case in self report of relationships, in many cases this led to correction of relationship's previously collected, and were the data collection asked for confirmation of a change being required, this was accepted as being the correct relationship. In some cases, e.g. where a relationship is corrected from partner to married, it has not been possible to reconcile whether this requires historical correction or just applies to the existing data collection.

7.8. Coding and Editing

7.8.1. MCS1

Details of coding and editing activities can be found in the Codebook and Edit Instructions prepared by NatCen, included in this deposit and their Technical Report on Fieldwork (NatCen 2004). Special thanks to Professor Neville Butler who was tireless in developing coding frames for the open-text answers to health questions, and in supervising the ICD10 coding at CLS of responses on mothers' and fathers' longstanding illness.

7.8.2. MCS2

Details of coding and editing activities can be found in the NOP Technical Report on Fieldwork (NOP 2006).

Recoding of occupation codes

In 2007, ONS were commissioned to re-code the occupation variables for MCS2.

Coding Approach

Automated Coding

In total, ONS received 52,868 records. The first stage of coding for ONS was to run the entire sample through its corporate automated coding tool "ACTR" (Automated Coding by Text Recognition). ACTR automatically coded 24,281 records, leaving 28,587 records.

Manual Coding

The 28,587 records not automatically coded were distributed equally between the coders in ONS, who were asked to make a variety of assumptions, as follows:

- Where the job title is non-descript, code using the job description.
- Where the job title and the job description differ, the record was coded to the job title.
- If the job title is not sufficiently detailed to assign a SOC 2000 code to the unit group (4-digit) level, code to the most detailed level possible.
- In the cases where there are 2 possible codes for the job title and a subjective approach was called for, the code assigned was always to the lowest level. For example, "Armed Forces" were coded to "Other rank" rather than "Officer".

Quality Checking

<u>Automated Coding</u>

At present, ACTR is tuned to code an ONS survey, for which the accuracy has been adjudicated as 99.80 per cent. As the MCS is new to ACTR all records coded by ACTR were checked and found to be 98 per cent accurate, with incorrect records manually changed. The reduced quality for the MCS was due to ACTR not being tuned for the survey, as it was the first time ACTR had seen it. Information from the MCS will be used to tune ACTR so the quality of ACTR will be enhanced should the MCS be automatically coded in the future.

<u>Manual Coding</u>

Once the manual coding was completed, a 10 per cent sample of the manually coded records was drawn by the ONS Methodology Division. The sample selected maintained the SOC code distribution, and was checked by someone other than the coder who initially coded the record. Coder accuracy was found to be 95 per cent, with queried records changed where appropriate.

Where there was ambiguity as to how to code a record it was decided that these would be coded after the majority of coding was completed in a "committee" format, all coders discussing and then coding the record together.

The final quality check involved grouping the job titles and showing all the different SOC codes associated with them. This allowed the coders to identify areas of inconsistency and make changes accordingly.

7.8.3. MCS3

Details of coding and editing activities can be found in the NatCen Technical Report on Fieldwork (NatCen 2007).

7.8.4. MCS4

Details of coding and editing activities can be found in the NatCen Technical Report on Fieldwork (NatCen 2010).

7.8.5. MCS5

Details of coding and editing by IPSOS-MORI can be found in the MCS5 Coding and Editing Report (IPSOS-MORI, 2013). Further editing e.g. value labels for multi-coded questions received from IPSOS-MORI included the response within the value label, this has been moved to the variable label and the description changed to Yes/No as appropriate were conducted at CLS.

7.9. Administrative Linked and Geographical data

MCS has collected consents to link to a range of other data: A detailed guide MCS Ethical review and Consent has been produced and is available from www.cls.ioe.ac.uk/mcssample

New linked data to administrative records and to geographical information of the cohort members are released through the UKDS.

For the linked data currently available please take a look at :

https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/

and at https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000031

7.10. Geographically Linked Data including IMD & Rural Urban Indicators

7.10.1. Index of Multiple Deprivation

The Index of Multiple Deprivation (IMD) measures relative levels of deprivation in small areas. As there is no unified definition for these measures across the UK, these are held as country specific variables. Whilst the IMD definitions are not directly equivalent, they could be broadly compared using the within country deciles.

The deciles were created using the rank for each sub-measure provided. As a practical example, in England there were 32,482 LSOAs, each decile containing 3,248 or 3,249 LSOAs. This data was then linked to address at interview at Lower Super Output Area Level. The IMD measures used were based on the following definitions:

- **England**: ODPM Indices of Deprivation 2004
- **Wales**: Welsh Assembly Index of Multiple Deprivation 2005
- **Scotland**: Scottish Assembly Index of Multiple Deprivation 2004
- **Northern Ireland**: NISRA Multiple Deprivation Measure 2005 May 2005.

The websites for ONS, Welsh Assembly, Scottish Assembly and NISRA have specific details:

England: www.communities.gov.**uk**/documents/communities/pdf/131209.pdf

Wales: http://wales.gov.uk/topics/statistics/theme/wimd/2005/?lang=en

Scotland: www.**scotland**.gov.uk/Publications/**2004**/10/20089/45181

Northern Ireland: www.**nisra**.gov.uk/archive/**deprivation**/nimdm**2005**fullreport.pdf

64

7.10.2. Rural Urban Indicators

MCS postcodes have been classified into different types of rural and urban areas. Again these are country specific. An overview is provided by ONS at http://www.ons.gov.uk.

The data for the Rural Urban measures were linked at Lower Super Output Area Level and used the following definitions:

- **England and Wales**: ONS Rural Urban Classification (2005)
- **Scotland**: Scottish Executive Urban Rural Classification (version 2005–2006)
  Northern Ireland: (NISRA) Urban Rural Classification 2005.

The Birkbeck definition of Rural Urban in England is that used by DEFRA. More information on this is available from ONS at the above URL.

PART 8: ETHICAL CONSIDERATIONS

8.1. MREC for MCS1

Medical research ethical clearance was obtained from the National Health Service Ethical Authority (in February 2001, MREC/01/6/19). This was as a general precaution for future health data collection and was specifically required because of the proposal to involve Health Visitors. Any research involving NHS staff needs to be given such clearance. We were directed to the South West Multi-Centre Research Ethics Committee in March 2001, who felt that opt-out sampling could be coercive and might fail to obtain properly informed consent. They did, however, accept that written opt-ins would tend to exclude vulnerable people, so procedures were devised in consultation with the Committee to give potential respondents more information before they committed themselves for interview. Advance letters introducing the interviewer were sent shortly before her/his first visit and they were asked to arrange interviews generally after their first visit, whose main purpose should be to give information. A simplified information sheet was produced, and translated into several languages.

8.2. MREC for MCS2

For MCS2, ethical approval was again sought for the pilot and main surveys – on this occasion from the London Multi-Centre Research Ethics Committee. Following their deliberations, the members of the Committee sought additional information on various aspects of the survey, commented on aspects of tracing procedures adopted for families discovered to have moved, and requested that a number of specific changes be made to information leaflets and consent forms. Ethical approval was obtained in September 2004, MREC/03/2/022).

8.3. MREC for MCS3

Both pilot surveys and the main survey of MCS3 were considered by the London Multi-Centre Research Ethics Committee of the NHS. Their letter granting a favourable ethical opinion for the *Economic and Social Research Council Millennium Cohort Study Third Survey 2005: Dress Rehearsal and Main Survey 2nd amendment* (12 December 2005) was granted on 15 December 2005, with the REC Reference No. 05/MRE02/46.

8.4. MREC for MCS4

Both pilot surveys and the main survey of MCS4 were considered by the Northern and Yorkshire Multi-Centre Research Ethics Committee of the NHS. Their letter granting a favourable ethical opinion for the Economic and Social Research Council

Millennium Cohort Study Fourth Survey: Dress Rehearsal and Main Survey 2nd amendment (3 January 2008) was granted on 5 February 2008, with the REC Reference No. 07/MRE03/32.

8.5. MREC for MCS5

Ethical approval for the Pilot 1 was obtained on 24th March 2011 from the Northern and Yorkshire REC: Ref: 11/H0903/3/

For the Dress Rehearsal and Main Stage approval was granted by the Yorkshire and Humber REC on 29th July 2011: Ref:11/YH/0203. On the 13th December 2011, confirmation of a favourable opinion was received in relation to a substantial amendment put to the Yorkshire and Humber REC covering the addition of the DWP data linkage consent collection to the study.

8.6. Codes of Practice

In order to support our assurances of confidentiality to informants, ethics committees, and government agencies to whose records links are being made, CLS extended the Cohort Studies Code of Practice to cover all those working with MCS data and developed a Data Security Policy, setting out the secure, isolated computing environment which handles any named data files within CLS.

8.7. Consents

At each sweep of the survey a series of consents were asked of the respondents. These are detailed below.

Table 43: Consents at each sweep of MCS

Survey	Consent	Who from	Elements	Document
MCS1	Verbal	Parent	Interview	
	Written	Mother	Maternity & birth linkage Linkage to child's National Health Service registration	MCS1 Technical Report
MCS2	Written	Parent	Child assessments & measurements	MCS2 Technical Report - appendices

Survey	Consent	Who from	Elements	Document
			Older sibling's questionnaire (Aged 10-15)	
			Linkage to Child's Records of school performance & attendance	
			Linkage to Older Siblings Records of school performance & attendance (Aged 5+)	
			Linkage to NHS Medical records (birth to age 7)	
			Saliva Sample	
MCS3	Written	Parent1	Main interview & self-completion	[MCS3 Technical Report on Fieldwork - appendices](#)
			Child assessments & measurements	Consent 1: data collection parent 1 and cohort child
			Linkage to NHS medical records and accidents (birth to age 7 – IF NOT GOT AT 3)	Consent 1: data collection parent 1 and cohort child
			Linkage to Foundation Stage Profile (England)	Consent 3: cohort child health records
			Teacher postal survey (S, W & NI)	Consent 4: E cohort child school admin data England
				Consent 4: NSW cohort child school data Ireland Wales Scotland1
			Older siblings questionnaire (England)	
				Consent 5: E older sibling questionnaire placement
			Linkage to older siblings school records (Age 7-16)	

Survey	Consent	Who from	Elements	Document
				Consent 6: E older sibling school records parents
	Written	Parent 2	Partner Interview & self-completion	MCS3 Technical Report on Fieldwork - appendices Consent 2: data collection parent 2
MCS4	Written	Main Parent	Main interview & self-completion Linkage to health & economic records Child assessments & measurements Child self-completion questionnaire Child physical activity monitor Class teacher postal survey Child's health records (birth to age 14) Child's educational records (to age 16) Sibling's health records (birth to age 14) Sibling's educational records (to age 16)	MCS4 Technical Report on Fieldwork: appendices MCS4_Consent 1: Main respondent MCS4_Consent 1: Main respondent MCS4 Consent 2 Cohort child data MCS4 Consent 3: Cohort child linkage MCS4 Consent 3: Cohort child linkage MCS4 Consent 3: Cohort child linkage MCS4 Consent 3: Cohort child linkage MCS4 Consent 3: Cohort child linkage MCS4_ Consent 4_Siblings MCS4_ Consent 4_Siblings

Survey	Consent	Who from	Elements	Document
	Written	Partner	Partner interview & self-completion	MCS4 Technical Report on Fieldwork - appendices
				MCS4 Consent 5: Partner respondent
			Linkage to health & economic records	MCS4 Consent 5: Partner respondent
MCS5	Written	Main Parent	Main interview & self-completion	MCS5 Technical Report on Fieldwork – appendix 2
				MCS5 Consent from Main Parent/Guardian
			To approach the cohort member to complete assessments & measurements & child questionnaire (England and Wales – includes permission to approach child's class teacher)	MCS5 Consent for Child Elements
			To approach the cohort member to complete assessments & measurements & child questionnaire (Scotland and Northern Ireland)	MCS5 Consent for Child Elements
			Linkage to DWP records	MCS5 Consent from Main Parent/Guardian
	Written	Partner	Partner interview & self-completion & linkage to DWP records	MCS5 Partner Consent
	Verbal	Cohort member	Child assessments & measurements	MCS5 consent from child

Survey	Consent	Who from	Elements	Document
			Child questionnaire	
			Approach class teacher for postal survey (England & Wales)	
			Child assessments & measurements	
			Child questionnaire (Scotland & Northern Ireland)	

PART 9: RELEASE OF DATASETS OF SWEEPS 1-4 TO A LONG FORMAT (SUMMER 2020)

Up to MCS6 the carers/parents have been the participants providing the majority of the information of the interview (about the family, themselves and the Cohort Member(s)). However, since MCS4 the number of questions answered by the Cohort Members (CM) has increased steadily. In MCS7 the main respondents are the Cohort Members.

To aid research focused on CMs, sweeps 1-4 are now in long format and contain one row per respondent. The old format was wide, and it provided one row per family. The old / wide datasets are not maintained or distributed after Summer 2020. Researchers need to download the new / long formatted datasets.

This section provides resources available to users of the previous / wide format datasets. The new / long format is the same as in sweeps 5 and 6, so many researchers will find this familiar.

A variety of resources are released to help users who have previously analysed the data in the old / wide format.

These resources provide technical support on the new / long format:

- PART 6 of this user guide with detailed information on the identifiers of MCS, and,
- Data Handling Guide with guidance and syntax on handling MCS data

Additional resources released help users locate variables:

- MCS Longitudinal Data Dictionary is a spreadsheet that provides information on all the variables available in MCS datasets across sweeps with topic information and,
- MCS1-4_Wide-Long Correspondence is a spreadsheet that provides users with the correspondence between the variables of the old / wide datasets and the new / long ones.

There are some further improvements to the data of sweeps 1-5, through enhancements to the metadata resources making the data easier to browse and identify variables of interest.

Sweep	MCS 1	MCS 2	MCS 3	MCS 4	MCS 5	MCS 6	MCS 7
Year of data collection	2001/2	2004/5	2006	2008	2012	2015/6	2018
CM's age	9 months	3	5	7	11	14	17
Education (England & Wales)			Year 1 Key stage 1	Year 3 Key stage 2	Year 7 Key stage 3	Year 10 Key stage 4	A-Level Key stage 5
Old format of datasets	Wide format (one row per family)	Wide format (one row per family)	Wide format (one row per family)	Wide format (one row per family)	Long format (person level - one row per respondent)	Long format (person level - one row per respondent)	Long format (person level - one row per respondent)
New format of datasets	Long format (person level - one row per respondent)	Long format (person level - one row per respondent)	Long format (person level - one row per respondent)	Long format (person level - one row per respondent)	Long format (person level - one row per respondent)	Long format (person level - one row per respondent)	Long format (person level - one row per respondent)

9.1. What are the differences between the old and new datasets?

The main difference lies in the structure of the data. The information is split into separate datasets with one row per respondent. For example, the parent interview is located in the mcs*_parent_interview dataset with one row per parent/carer respondent. Similarly, the Cohort Member interview or measurements are located in the mcs*_cm_interview datasets.

Figure 7: Correspondence between wide and long format variables

The previous / wide dataset had a large number of variables and required the user to consider whether a variable contained information about the main parent or the partner, or about the 1st, 2nd or 3rd Cohort Member of the family. The specific information on "who answered the question / who provided the information" was previously found in the variable label and the variable name. Figures 7 and 8 show how information of the old / wide formatted datasets corresponds to the new / long formatted datasets. A question asked from each parent would create 2 variables in the old / wide format: one for the main respondent and one the partner respondent. However, in the new / long format this question will be in one variable. In the long format, the identifiers MCSID (family) and person (PNUM/ELIG) help the user find the row for the main respondent and the row for the partner respondent.

Differences between Wide and Long format
variable naming and labelling conventions in detail

CAPI name	Variable in the old wide format Sweep M/P CAPI a/b/c	Variable in the new long format Sweep P/C/H/X VAR
	M/P = Main or Partner a/b/c = CM 1/2/3	P = Parent C = Child X = Proxy H = Household
PRFE	In mcs1_parent_interview.sav AMPRFE00 "S1 MAIN How felt when became pregnant" APPRFE00 "S1 PARTNER How felt when became pregnant"	In mcs1_parent_level.sav APPRFE00 "S1 How felt when became pregnant"
HAPN	In mcs1_parent_interview.sav AMHAPNA0 "S1 MAIN Happy sounds during nappy changing etc C1" AMHAPNB0 "S1 MAIN Happy sounds during nappy changing etc C2" AMHAPNC0 "S1 MAIN Happy sounds during nappy changing etc C3"	In mcs1_parent_cm_level.sav ACHAPN00 "S1 Happy sounds during nappy changing etc"
APBA	In mcs1_parent_interview.sav AMAPBAA0 "S1 MAIN Whether lived apart from baby C1" AMAPBAB0 "S1 MAIN Whether lived apart from baby C2" AMAPBAC0 "S1 MAIN Whether lived apart from baby C3" APAPBAA0 "S1 PART Lived apart from baby since birth C1" APAPBAB0 "S1 PART Lived apart from baby since birth C2" APAPBAC0 "S1 PART Lived apart from baby since birth C3"	In mcs1_parent_cm_level.sav APAPBA00 "S1 Whether lived apart from baby"

Figure 8: Variable presentation in wide and long format datasets

The identifiers are important in MCS because of the hierarchical structure of the information (long formatted). Person identifiers in each dataset structure specify the parent/carer (PNUM or ELIG) and the Cohort Member (CNUM). MCSID remains the same and is a family / household identifier. Specifically:

- the information provided by the parents about themselves is in the _parent_ datasets
- the information provided by the Cohort Members about themselves is in the _cm_ datasets
- the information provided by the parents about the Cohort members is in the _parent_cm_ datasets

74

This user guide includes a detailed description of the long format datasets of MCS and the importance of the each of the identifiers (PART 6). Furthermore, a Data Handling guide is available that visualizes the data of MCS and contains data handling examples and syntax in R, SPSS and STATA.

Further improvements to the data are:

- Resolution of two duplicate cases. This has affected the total number of families. The final number of productive families per sweep can be found in the mcs_longitudinal_family_file
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8172
- Through the MCS Longitudinal Data Dictionary of MCS it is possible to identify the location of the variables: End User License, Special License or Secure Access. This is available for download from the UKDS without requiring logging in. Section 4 provides an overview on the information that is available under each license.
- The variable labels have been harmonised across sweeps. This means that questions with the same wording and CAPI code (the four character code that appears in the questionnaire) in different sweeps have the same variable label.
- There have been minor updates for a small number of families (3) due to the resolution of a duplicate case and removal of another. Due to this the total number of families has altered slightly. The mcs_longitudinal_family_file
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8172
contains the up to date information on the total productive families per sweep.

9.2. How can I know how the variables I have used correspond to the new variables and datasets?

The information on the correspondence between the wide (before Summer 2020) and long (Summer 2020 onwards) datasets is available in the spreadsheet named MCS1-4_Wide-Long Correspondence_v* .

9.3. How can I restructure the long format data into wide format for my research?

It is possible to restructure each dataset of MCS into a wide format, with one row per family. When restructuring into a wide format, one needs to take into account the identifiers of each datasets. For example, for the _cm_ and _parent_ level there are 2 identifiers (MCSID and CNUM / PNUM / ELIG), whereas for the _parent_cm_ there are 3 (MCSID and CNUM and PNUM / ELIG).

For more information please take a look at reshape in STATA and CASESTOVARS in SPSS.

9.4. Information in End User License (EUL) and Secure Access (SA)

Information about specific questions that contain sensitive details has been distributed to separate datasets.

The table below shows the availability and distribution.

Complete information on variables available under EUL and SA is in the **MCS_Longitudinal_Data_Dictionary** (spreadsheet available for download from the UKDS without logging in required).

If research requires a coding of the soc codes in categories based on the full information data (for example, aggregate information from the ONS or Labour Force Survey), then the data users can apply for a data enhancement. More information is provided here: https://cls.ucl.ac.uk/data-access-training/data-enhancements/ If you would like to discuss on what is possible before applying for a data enhancement please contact clsfeedback@ucl.ac.uk adding 'DAC' in the e-mail topic.

Theme	End User License	Special Access
Ethnicity	Derived variables on ethnicity exist in the derived variables datasets with information on ethnicity up to 11 categories.	The information that respondents provided at the survey interview.
SIC & SOC codes	The datasets contain sic and soc coding across sweeps. Up to 3 digits of SOC codes and up to 2 of SIC.	The entire SOC and SIC code.
ICD codes on long standing limited illness	The binary variable on whether the respondent answered about the presence of a limiting longstanding illness exist in all datasets. Only specific ICD codes are provided in the EUL datasets.	The entire ICD codes.
BNF codes (medication)	MCS_Longitudinal_Data_Dictionary has information on similar variables.	The entire BNF codes.
Hearing & Visual conditions	The information on whether the cohort member has a hearing or visual impairment is available under EUL. The same applies for some of	Detailed information on condition..

Theme	End User License	Special Access
	the variables that follow in the same questionnaire.	
Country of birth of parents/grandparents of the Cohort Members	The country of birth is provided under EUL but the countries with low counts are given as 'Continent Other'	Detailed country of birth as provided by data linkage (mcs1) and respondents (subsequent to mcs1).
Number of rooms	The number of rooms is provided in grouped categories.	Exact number of rooms as provided by the respondents.

Centre for Longitudinal Studies
Institute of Education
20 Bedford Way
London WC1H 0AL
Tel: 020 7612 6860
Fax: 020 7612 6880
Email: clsfeedback@ucl.ac.uk
Web: cls.ucl.ac.uk