



Growing Up In Scotland

Sweep 4: 2008-2009

User Guide

Authors: Paul Bradshaw, Louise Marryat, Joan Corbett, Mireille Ferrandon and Sarah Tipping

1	Overview of the survey	3
1.1	Study Design	3
1.2	Sample Design	4
1.3	Development and Piloting.....	5
2	Data collection methods	5
2.1	Mode of data collection.....	5
2.2	Length of Interview	6
2.3	Timing of fieldwork.....	6
3	Child height and weight measurements.....	6
4	Response rates	7
5	Coding and editing	8
6	Weighting the data.....	8
6.1	Overview.....	8
6.2	Background	8
6.3	The sweep 4 sample	8
6.4	Longitudinal weights.....	9
6.4.1	Final Sweep 4 longitudinal weights	10
6.5	Cross-sectional weights.....	10
6.6	Sample efficiency	11
6.7	Applying the weights.....	12
6.8	Weighting variables	12
7	Using the data	13
7.1	Variables on the files	13
7.2	Variable naming convention	13
7.3	Variable labels	14
7.4	Derived variables.....	14
7.5	Household data	14
7.6	Childcare data	15
7.6.1	Childcare and Pre-school arrangements.....	18
7.7	Indicators and summary variables.....	19
7.7.1	Socio-economic characteristics: National Statistics Socio-economic Classification (NS-SEC).....	19
7.7.2	Socio-economic characteristics: Equivalised household annual income	20
7.7.3	Socio-economic characteristics: Index of Material Deprivation.....	20
7.7.4	Area-level variables: Scottish Government Urban/Rural Classification	21

7.7.5	Area-level variables: Scottish Index of Multiple Deprivation	22
7.7.6	Area-level variables: Carstairs Index.....	22
7.7.7	Height and weight measurements: Body Mass Index (BMI) scores.....	23
7.7.8	Child Development: Strengths and Difficulties Questionnaire.....	24
7.7.9	Parental Health: Depression, Anxiety and Stress Scale	25
7.7.10	Risk and protectiveness: the Parent Supervision Attributes Profile Questionnaire.....	25
7.8	Dropped Variables.....	26
7.9	Missing values conventions.....	26
8	Documentation.....	27
9	References.....	27
10	Related publications.....	27
11	Contact details	29
	Appendix A: Full non-response models.....	30
	Appendix B: Issues to be aware of when working with the data.....	38

1 Overview of the survey

The data files contain data from Growing Up in Scotland (GUS) Sweep 4, the fourth year of a longitudinal research study aimed at tracking the lives of two cohorts of Scottish children from the early years, through childhood and beyond. Funded by the Scottish Government Education Directorate, its principal aim is to provide information to support policy making, but it is also intended to be a broader resource for secondary analysis.

The aims of the study are:

- To provide reliable cross-sectional data on the characteristics, circumstances and experiences of children in Scotland aged between 0 and 5.
- To document differences in the current characteristics, circumstances and experiences of children from different backgrounds
- To generate information about longer-term outcomes across a range of key domains and to document differences in those outcomes for children of different backgrounds.
- To identify key predictors of adverse longer-term outcomes with particular reference to the role of early years service provision
- To measure levels of awareness and use of key services
- To examine the nature and extent of informal sources of help, advice and support for parents
- To generate parental assessments of the services accessed and used; and to improve understandings of choice and constraint in service use.

At sweep 4, data collection for the study included two main elements:

1. A face-to-face CAPI interview with the cohort child's main carer
2. Height and weight measurements in order to calculate the Body Mass Index (BMI)
- further details are included in sections 3 and 7

1.1 Study Design

The survey is based on two cohorts of children: the first aged approximately 10 months at the time of first interview and the second aged approximately 34 months. A named sample of approximately 10,700 children was selected from the Child Benefit records to give an achieved sample of 8,000 overall.

The configuration of cohorts and sweeps for the first four sweeps of data collection is summarised below. BC1 refers to the younger of the two cohorts ('birth cohort') and CC1 to the slightly older cohort ('child cohort').

Table 1.1 Sample design: sweeps 1 to 4

Sweep Launch year	Age at interview					
	0-1	1-2	2-3	3-4	4-5	5-6
1 2005	BC1		CC1			
2 2006		BC1		CC1		
3 2007			BC1		CC1	
4 2008				BC1		CC1

A key aim of using two cohorts is to allow the study to provide three types of data:

1. *Cross-sectional time specific data* – e.g. what proportion of 2-3 year-olds are living in single parent families in 2005?
2. *Cross-sectional time series data* – e.g. is there any change in the proportion of 2-3 year-olds living in single parent families between 2005 and 2007?
3. *Longitudinal cohort data* – e.g. what proportion of children who were living in single parent households aged 0-1 are living in different family circumstances at age 4-5?

1.2 Sample Design

The area-level sampling frame was created by aggregating Data Zones. Data Zones are small geographical output areas created for the Scottish Government. Data Zones are used by Scottish Neighbourhood Statistics to release small area statistics. The Data Zone geography covers the whole of Scotland. The geography is hierarchical, with Data Zones nested within Local Authority boundaries. Each data zone contains between 500 and 1,000 household residents. More information can be found on the Scottish Neighbourhood Statistics website: <http://www.sns.gov.uk>.

The Data Zones were aggregated to give an average of 57 births per area per year (based on the average number of births in each Data Zone for the preceding 3 years). It was estimated that this number per area would provide us with the required sample size. Once the merging task was complete, the list of aggregated areas was sorted by Local Authority¹ and then by the Scottish Index of Multiple Deprivation Score (SIMD). 130 areas were then selected at random. The Department of Work and Pensions then sampled children from these 130 sample points.

Within each sample point, the Child Benefit records were used to identify all babies and three-fifths of toddlers who met the date of birth criteria (see Table 1.2). The sampling of children was carried out on a month-by-month basis in order to ensure that the sample was as complete and accurate as possible at time of interview.

In cases where there was more than one eligible child in the selected household, one child was selected at random. If the children were twins they had an equal chance of being selected. If the eligible children were in different age cohorts the younger child had a higher chance of being selected given that those children had a higher chance of being included in the sample overall.

After selecting the eligible children, the DWP made a number of exclusions before transferring the sample details. These exclusions included cases they considered 'sensitive' and children that had been sampled for research by the DWP in the last 3 years.

¹ Local Authority has been used as a stratification variable during sampling, this means the distribution of the GUS sample by Local Authority will be representative of the distribution of Local Authorities in Scotland. However, the sample sizes are such that we would not recommend analysis by Local Authority. The small sample sizes would give misleading results.

Table 1.2 Eligible child dates of birth for inclusion in the Growing Up in Scotland study by cohort

Sample Number	Dates of Birth required	
	Birth Cohort	Child Cohort
1	01-June-2004 - 30-Jun-2004	01-June-2002 - 30-Jun-2002
2	01-Jul-2004 - 31-Jul-2004	01-Jul-2002 - 31-Jul-2002
3	01-Aug-2004 - 31-Aug-2004	01-Aug-2002 - 31-Aug-2002
4	01-Sep-2004 - 30-Sep-2004	01-Sep-2002 - 30-Sep-2002
5	01-Oct-2004 - 31-Oct-2004	01-Oct-2002 - 31-Oct-2002
6	01-Nov-2004 - 30-Nov-2004	01-Nov-2002 - 30-Nov-2002
7	01-Dec-2004 - 31-Dec-2004	01-Dec-2002 - 31-Dec-2002
8	01-Jan-2005 - 31-Jan-2005	01-Jan-2003 - 31-Jan-2003
9	01-Feb-2005 - 28-Feb-2005	01-Feb-2003 - 28-Feb-2003
10	01-Mar-2005 - 31 Mar-2005	01-Mar-2003 - 31 Mar-2003
11	01-Apr-2005 - 30-Apr-2005	01-Apr-2003 - 30-Apr-2003
12	01-May-2005 - 31-May-2005	01-May-2003 - 31-May-2003

1.3 Development and Piloting

Policy priorities and key topics of interest for the sweep 4 questionnaire were initially discussed and agreed by the study's Scottish Government Project Managers and Policy Advisory Group. The questionnaire was then developed by the GUS team at ScotCen with input from colleagues at the Centre for Research on Families and Relationships (CRFR) in reference to these priorities and topics. A subset of new questions was included in a small cognitive pilot in October 2007, with a full instrument initially piloted in paper form in November 2007. This instrument was revised and converted into CAPI for the second 'Dress Rehearsal' Pilot in January 2008.

2 Data collection methods

2.1 Mode of data collection

Interviews were carried out in participants' homes, by trained social survey interviewers using laptop computers (otherwise known as **CAPI** – Computer Assisted Personal Interviewing). The interview was quantitative and consisted almost entirely of closed questions. There was a brief, self-complete section in the interview in which the respondent, using the laptop, input their responses directly into the questionnaire programme.

At sweep 1, primarily because of the inclusion of questions on the mother's pregnancy and birth of the sample child, interviewers were instructed as far as possible to undertake the interview with the child's mother. Where the child's mother was not available, interviews were undertaken with the child's main carer.

At the following sweeps, interviewers were instructed to undertake the interview with the same respondent as in the previous sweep. At Sweep 4, this means the same respondent as Sweep 3, or Sweep 2 / Sweep 1 if the household skipped some of the sweeps. Where this was not possible or appropriate, interviews were conducted with the child's main carer. In practice, most interviews

were undertaken with the previous sweep respondent (98.7% of interviews were with the previous respondent) and this was usually the child's mother (97.7% of interviews were with the child's mother).

2.2 Length of Interview

Overall, the average interview lasted around 64 minutes. The child cohort interview had a slightly longer average length at 66 minutes, than the birth cohort interview at 63 minutes. The median interview length was 60 minutes for the birth cohort and 61 minutes for the child cohort.

2.3 Timing of fieldwork

Fieldwork was undertaken over a fourteen month period commencing in April 2008. The sample was issued in twelve monthly waves at the beginning of each month and each month's sample was in field for a maximum period of two and a half months. For example, sample 2 was issued at the beginning of May 2008 and remained in field until mid-July 2008.

To ensure that respondents in both samples were interviewed when their children were approximately the same age, each case was assigned a 'target interview date'. For the birth cohort this was identified as the date on which the child turned 46.5 months old, and for the child cohort the date the child turned 70.5 months old. Interviewers were allotted a four-week period based on this date (two weeks either side) in which to secure the interview. In difficult cases, this period was extended up to and including the child's subsequent birthday which allowed a further four weeks. The vast majority of interviews were achieved within the four-week target period.

3 Child height and weight measurements

The relationship between general build and health is of great interest to the Scottish Government, especially in relation to children. This is particularly so, as both the height and the weight of the population appear to have been changing very rapidly over the last two decades. These changes reflect the changes in the children's diet and lifestyle. This survey will provide a reliable source of data on the changes that are taking place in all of these areas.

Although many people know their child's height and weight, these measurements are not usually up to date or are not known with the precision required for the survey. The reason for wanting to know accurate heights and weights is in order to relate them to other health measures.

The interviewers were asked to measure the height and weight of all children in both cohorts. However, in some cases it may not have been possible or appropriate to do so, for example if it was clear that the child was unwilling or that the measurement would be far from reliable.

It was recommended that height and weight measurements be taken on a floor which was level and not carpeted. If all the household was carpeted, a floor with the thinnest and hardest carpet was chosen (usually the kitchen or bathroom).

For the weight measurements, there was an option to weigh the child whilst being held by an adult. In this case, the adult was weighed on his/her own first and then the adult and the child were weighed together. Both weights were entered in the computer, which calculated the child's weight.

The interviewer was asked to code whether they experienced problems with the height and/or weight measurements and, if they did, to indicate whether they felt the end result was reliable or unreliable at WdXhei14 and WdXwei19. As a rough guide, if the measurement was likely to be more than 2 cms (3/4 inch) from the true figure for height or 1 kg (2 lbs) from the true figure for weight, it was coded as unreliable.

If the respondent was not willing to allow the sample child to have his/her height or weight measured, for example saying that they were too busy or already knew their measurements, a Refusal code was entered for the measurements variables WdXhei01 and WdXwei01, with the reason for refusal at WdXhei02 or WdXwei02.

If the height or weight was refused or not attempted, the respondent was asked to estimate their child's height or weight, in metric or imperial measurements.

Detailed protocols of how to take height and weight measurements are included as appendices to the main interviewer instructions deposited with the dataset and available from the data archive website.

The data has been used to estimate an approximate BMI (Body Mass Index) score for each child. Further details on the data and variables associated with the height and weight measurements can be found in section 7.7.7.

4 Response rates

Details of the number of cases issued and achieved and the response rates are detailed in Table 4.1.

Table 4.1 Number of issued and achieved cases and response rates

	Birth Cohort	Child Cohort	All Sample
Achieved interviews at sweep 1	5217	2858	8075
Achieved interviews at sweep 2	4512	2500	7012
Achieved interviews at sweep 3	4193	2332	6525
Cases to field at sweep 4:			
All issued to field*	4394	2460	6854
Eligible i.e. achievable or 'in-scope'**	4374	2453	6827
Cases achieved at sweep 4	3994	2200	6194
Response rate			
As % of all eligible cases at sweep 4	91%	90%	91%
As % of all sweep 1 cases	77%	77%	77%

* The number of cases issued to the field at sweep 4 is higher than the number of Interviews achieved at sweep 3 because some of the sweep 1 and sweep 2 respondents missed at sweep 3 came back at sweep 4.

** Cases which were considered out-of-scope or unachievable were mostly ineligible addresses – usually due to the family having moved away from Scotland.

5 Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted. The GUS Sweep 4 Coding Instructions, deposited along with this User Guide, provide details of the tasks that were conducted.

6 Weighting the data

6.1 Overview

Two sets of weights have been developed for each cohort:

1. A cross-sectional weight that should be used for any cross-sectional analysis of Sweep 4 data only. All sample members that responded at Sweep 4 have a cross-sectional weight.
2. A longitudinal weight for analysis of more than one wave of data. Sample members that have responded at every wave of GUS thus far have a longitudinal weight.

6.2 Background

- The sampling frame was the child-level Child Benefit records held by the Inland Revenue. Children were selected from 130 sample points in Scotland. The sample points consist of aggregations of Data Zones².
- There are two cohorts of children: the birth cohort and child cohort. Children in the birth cohort were aged approximately 10 months at the time of first interview whereas children in the child cohort were aged around 34 months. Weights for the birth and child cohorts have been generated separately, since these two groups are always analysed separately.
- The Sweep 4 interview follows up all main carers who responded at the previous interview and gave ScotCen permission to be re-contacted. Some of the respondents of Sweep 1 or 2 who had asked not to take part for a year but were willing to be contacted the following year were also included at Sweep 4.
- At Sweep 4 we used proxy interviews to gather information on the main respondent's resident partner.

6.3 The sweep 4 sample

The Sweep 4 sample can be split into two components; for the purposes of describing the weighting these two components have been named Sample A and Sample B and are defined as follows:

- Sample A – Wave 4 respondents who responded at all waves
- Sample B – Wave 4 respondents who responded at Wave 1 but had missed an intervening wave.

The two samples will be treated separately during the weighting. This is because the Sample B respondents are likely to have different response behaviour to respondents in Sample A, as suggested by their much lower response rates. The birth cohort contained 542 families in Sample B, 150 (28%) of which responded at Sweep 4. The child cohort contained 189 families in Sample B,

² Further information on the sample design and the weighting process at sweeps 1, 2 and 3 can be found in the User Guides for those sweeps which are available from the Data Archive or the 'using GUS data' section

of which 100 (53%) responded. The response rates for Sample B were much lower than the response rates for Sample A, which was 94% for the birth cohort and 93% for the child cohort. The issued and responding sample sizes are given in Table 6.1.

Table 6.1 Response rates for different samples

	Issued	Responding	Response rate
Baby cohort			
Sample A	4105	3844	94%
Sample B	289	150	28%
Combined (A+B)	4394	3994	91%
Toddler cohort			
Sample A	2270	2100	93%
Sample B	190	100	53%
Combined (A+B)	2460	2200	90%

There will be two sets of weights developed for each cohort; a cross-sectional weight and a longitudinal weight.

The longitudinal weight will be used for any analysis that includes more than one wave of data. Only members of Sample A (who have responded at every wave of GUS) will have a longitudinal weight. This weight is described in more detail in Section 6.4.

The cross-sectional weight will be used for any cross-sectional analysis of Wave 4 data. All Sweep 4 respondents will have a cross-sectional weight (Sample A + B). These are described in more detail in section 6.5.

6.4 Longitudinal weights

Longitudinal weights were only generated for respondents in Sample A. A model-based weighting technique was used to develop the Sweep 4 longitudinal weights, where response behaviour is modelled using data from previous sweeps. This is the same method used to generate weights at Sweeps 2 and 3. Ineligible households (deadwood) were not included in the non-response modelling³.

Response behaviour was modelled using logistic regression. A logistic regression models the relationship between an outcome variable (in this case response to the Sweep 4 interview) and a set of predictor variables. The predictor variables were a set of socio-demographic respondent and household characteristics collected from the previous sweeps.

The model generated a predicted probability for each respondent. This is the probability the respondent would take part in the interview, given the characteristics of the respondent and the household. Respondents with characteristics associated with non-response (such as being a private tenant) are under-represented in the sample and will receive a low predicted probability.

of the Growing Up in Scotland website www.growingupinScotland.org.uk

³ There were 11 individuals with ineligible outcome codes; these individuals were dropped from the analysis. Ineligible outcome codes include households that were vacant, demolished or derelict and non-residential addresses, where no follow up address could be found.

The non-response weights are then generated as the inverse of the predicted probabilities; hence respondents who had a low predicted probability get a larger weight, increasing their representation in the sample.

The birth and child cohorts were modelled separately, although there were similarities between the two models. A summary of the characteristics related to response behaviour at Sweep 4 are given in Table 6.2. The full models are given in Tables A1 and A2 in the Appendix.

Table 6.2 Characteristics associated with response behaviour in Sample A

Characteristics associated with response	Characteristics associated with non-response
Birth cohort	
Owner occupiers	Rent from a private landlord; rent from Housing Association
From a white ethnic background	From any other ethnic background
At least one parent/carer in full-time employment	No parent/carer working; at least one parent/carer in part-time employment
Mother aged 25 or over	Younger mother aged under 20
One child in the household	Four or more children in household
Mother has a degree	Mother has no qualifications
Lives in other urban areas; small towns	Lives in large urban areas
Child cohort	
Owner occupiers	Rent from a private landlord, rent from Housing Association
Lives in less deprived areas (not in the 15% most deprived Data Zones)	Lives in deprived areas (in the 15% most deprived Data Zones)
One child in the household	Three children in the household
Respondent in managerial and professional occupations	Respondent in lower supervisory and technical occupations; does not work.
From a white ethnic background	From any other ethnic background

6.4.1 Final Sweep 4 longitudinal weights

The final W4 weight is the product of the Sweep 4 non-response weight and the Sweep 3 interview weight. For each cohort the final weights were scaled to the responding Sweep 4 sample size, this makes the weighted sample size match the unweighted sample size. Table A3 in the Appendix shows the distribution of the sample weighted by the Sweep 4 and Sweep 3 weights, showing the reduction in bias caused by the Sweep 4 weights.

6.5 Cross-sectional weights

Cross-sectional weights were generated for all respondents at Sweep 4 (the combined A and B samples) and should be used for any cross-sectional analysis of Sweep 4 data.

Calibration weighting methods were used to create the cross-sectional weights. This method takes the pre-calibrated weighted combined sample and adjusts the weights using an iterative procedure. The resulting weighting factors, when applied to the combined data, will make the survey estimates

match a set of population estimates for a set of key variables. The population estimates in this instance are survey estimates from Sample A, weighted by the longitudinal weight. Since the longitudinal weight corrects for sampling error and non-response bias at each stage of GUS, the weighted Sample A estimates are the best population estimates available. The key variables used in the weighting were; Area level deprivation indicator (measured using the Scottish Index of Multiple Deprivation), respondent employment status, respondent age at interview, household income and whether the respondent was a lone parent.

The pre-calibration weights were the Sweep 4 longitudinal weight for Sample A and the weight from the last completed sweep for Sample B. Prior to calibration these weights were scaled to the achieved sample size, giving a mean weight of one. This was done separately for each sample.

The calibration corrects for any differences due to differential non-response between Sample A and Sample B. The weighted distribution of Sample A and the weighted distribution of the combined sample, pre and post-calibration, are given in Tables A4 and A5 for the birth and child cohorts, respectively.

6.6 Sample efficiency

Adding weights to a sample can affect the sample efficiency. If the weights are very variable (i.e. they have very high and/or very low values) the weighted estimates will have a larger variance. More variance means standard errors are larger and confidence intervals are wider, so there is less certainty over how close the estimates are to the true population value.

The affect of the sample design on the precision of survey estimates is indicated by the effective sample size (neff). The effective sample size measures the size of an (unweighted) simple random sample that would have provided the same precision (standard error) as the design being implemented. If the effective sample size is close to the actual sample size then we have an efficient design with a good level of precision. The lower the effective sample size, the lower the level of precision. The efficiency of a sample is given by the ratio of the effective sample size to the actual sample size. The range of the weights, the effective sample size and sample efficiency for both sets of weights are given in Table 6.3.

Table 6.3 Range of weights and sample efficiency

	Weight values			Unweighted achieved sample size	Neff	Sample Efficiency
	Minimum	Maximum	Mean			
Birth cohort						
Longitudinal weight	0.65	2.83	1.00	3844	3484	90.6%
Cross-sectional weight	0.67	2.85	1.00	3994	3648	91.3%
Child cohort						
Longitudinal weight	0.61	2.90	1.00	2100	1941	92.4%
Cross-sectional weight	0.63	2.82	1.00	2200	2048	93.1%

6.7 Applying the weights

The cross-sectional weights should be used for any cross-sectional analysis, i.e. any analysis of Sweep 4 data only. All sample members that responded at Sweep 4 have a cross-sectional weight.

The longitudinal weight should be used for any analyses of more than one wave of data. Sample members that have responded at every wave of GUS have a longitudinal weight.

6.8 Weighting variables

The final interview sweep 4 weights are described in Table 6.4.

Table 6.4 Description of weight variables in the data file

Variable name	Label
DdWTbrth	Dd Birth cohort Sw4 weight (cross sectional sample)
DdWTchld	Dd Child cohort Sw4 weight (cross sectional sample)
DdWTbth2	Dd Birth cohort Sw4 weight - longitudinal
DdWTchd2	Dd Child cohort Sw4 weight - longitudinal

Separate weights are provided for each cohort because analysis should always treat each cohort as a distinct population. However, key analysis using this data may involve comparison between the cohorts. It is usually more convenient to undertake this analysis by combining the two cohort datasets into a single dataset and then ensuring that subsequent analysis is either filtered to select a single cohort, or that output is nested by cohort type (variable name = 'SampType'). On merging the datasets it is possible to create a combined weight variable in order that nested analysis uses just a single weight variable. The value of the combined weight is equal to the value of the corresponding cohort weight variable for that child. Syntax to create the combined main interview weight is included below:

Compute DdWTbrch = DdWTbrth.

If (SampType = 2) DdWTbrch = DdWTchld.

7 Using the data

The GUS Sweep 4 data consists of two SPSS files

GUS_SW4_B.sav	3994 cases	Birth cohort
---------------	------------	--------------

GUS_SW4_C.sav	2200 cases	Child cohort
---------------	------------	--------------

7.1 Variables on the files

Each of the data files contain questionnaire variables (excluding variables used for administrative purposes) and derived variables. The variables included in the file are detailed in the “**Variable List**” document in the data section of the documentation. As far as possible they are grouped in the order they were asked in the interview.

For variables with answers following a scale, such as ‘Strongly agree’ to ‘Strongly disagree’ for instance, it must be noted that the order of the answer categories may not follow systematically an ascending or descending scale throughout the list of variables. Also the answers may equally refer to positive or negative statements as in the Strength and Difficulties questions MdSDQ01 to 25. The phrasing of the question and the list of answers provided on the showcards - if any - shape the variables. The user must therefore take these variations into account when creating derived variables.

The large number of checks undertaken on the data ahead of its deposit occasionally brings to light quality or validity issues which should be taken into account when analysis is being undertaken on the related variables. These issues are listed in Appendix B.

7.2 Variable naming convention

Variables names are made up of 8 characters, the first indicates the source of the variable, the second the year of collection and the rest is an indication of the question topic. Therefore where the same question was asked in the different sweeps the names will usually be the same apart from the second character. If a variable name has changed substantially between sweeps this is marked in the variable list. The naming convention is summarised in Table 7.1

Table 7.1 GUS variable naming conventions

Character No:						
1		2		3	4, 5 & 6	7&8
Source of data		Sweep/Wave		Key theme prefix	Sub theme stem	Question/Variable number
Non- sequential Capitals: D,M, P, S		Sequential lower case: a, b, c..		Non-sequential Capitals: C, P, N...	Abbreviated lower case: e.g. hea,	01 - 99
Source code	Details	Sweep code	Details			
AL	Area Level variable					
D	Derived variable	a	Sweep 1 (2005/06)			
DP	Derived variable from partner int	b	Sweep 2 (2006/07)			
DWP	DWP variable	c	Sweep 3 (2007/08)			

M	Main carer/adult interview	
P	Partner's interview	
W	Weights and Heights	

7.3 Variable labels

In the Sweep 4 dataset the variable labels are restricted to 40 characters as far as possible; the first 2 show the source and year of the data (as in the variable name). Although the labels give an indication of the topic of the question it is essential to refer to the questionnaire to see the full text of the question and the routing applied to that variable. The variable list shows the page numbers of the relevant questionnaire section.

7.4 Derived variables

Derived variables included in the dataset are listed with the questionnaire variables for the same topic. The SPSS syntax used to create them can be found in the “**Derived Variables**” section of the documentation.

7.5 Household data

In addition to the questions asked about the child and parents, the respondent was also asked about each household member. The gender, age and marital status of each household member was collected along with their relationship to each other and the cohort child. Each person was identified by their person number, which they will retain through each sweep of the survey. The variable MdHGSI(n) can be used to see whether a person who was in the household at sweep 1, 2 or 3 is still in the household at sweep 4.

A set of derived summary household variables is also included in the data. Amongst other things these detail the number of adults, number of children or number of natural parents in the household. A list of these variables is included in Table 7.2. A set of variables which allow identification of the respondent and their partner (if present) in the household grid are also included. These permit easier analysis of respondent and partner age, marital status and relationship to other people in the household. The age variables have been banded for all persons in the household except the study child.

Table 7.2 Key household derived variables

DdHGnmad	Dd - Number of adults (16 or over) in household
DdHGnmkd	Dd - Number of children in household
DdHGnmsb	Dd - Number of siblings in household
DdHGnp01	Dd - Number of natural parents in household
DdHGrsp01	Dd - Whether respondent is natural mother
DdHGrsp02	Dd - Whether respondent is natural father
DdHGnp02	Dd - Natural mother in household
DdHGnp03	Dd - Natural father in household
DdHGnp04	Dd - Respondent living with spouse/partner
DdMothID	Dd – Mother’s ID (= Person number in household)
DdFathID	Dd - Father’s ID

DdRespID	Dd – Respondent’s ID
DdPartID	Dd – Respondent’s partner’s ID
DdRPAge	Dd – Respondent’s partner’s age (banded)
DdRPsex	Dd - Respondent partners sex

7.6 Childcare data

The childcare section of the CAPI questionnaire utilises feed-forward data. This technique allows information collected at the previous sweeps to be ‘fed-forward’ into the current sweep’s CAPI questionnaire for the respondent to confirm or change rather than such information being completely re-collected. This reduces respondent burden and allows for the saved time to be used elsewhere in the interview.

At sweep 4, for those cases where childcare had been used at the previous sweep, details of the previous sweep arrangements – including the provider name, provider type, the number of hours they looked after the child per week and the number of days over which those hours were spread – were fed-forward. The respondent could confirm whether all details were still correct, change the number of hours or days, or indicate that the arrangement was no longer being used. All respondents could also provide details of new arrangements which were in place at sweep 4 but had not been in place at the previous sweep. The multiple sets of information collected create a particularly complex data structure.

To make this complex picture more comprehensible, the childcare data can be usefully separated into three sections suitable for different types of analysis. The first is concerned with **continuity of provision** from sweep to sweep. The relevant variables include those which contain the details of the childcare arrangements of the previous sweep, and those which confirm whether or not the arrangement is still in place, and for those arrangements which have been ceased, the reasons why. These variables are detailed in Table 7.3.

Table 7.3 Childcare variables for exploring continuity of provision

Variable name	Description
MaCtya01/DbCtya01/DcCtya01	Sw1 / Sw2 / Sw3 1st childcare provider type
MaCtma01/DbCtma01/ DcCtma01	Sw1 / Sw2 / Sw3 1st childcare provider - no of hours per week
MaCdya01/DbCdya01/ DcCdya01	Sw1 / Sw2 / Sw3 1st childcare provider - no of days per week
MaCtyb01/DbCtyb01/ DcCtyb01	Sw1 / Sw2 / Sw3 2nd childcare provider type
MaCtmb01/DbCtmb01/ DcCtmb01	Sw1 / Sw2 / Sw3 2nd childcare provider - no of hours per week
MaCdyb01/DbCdyb01/ DcCdyb01	Sw1 / Sw2 / Sw3 2nd childcare provider - no of days per week
MaCtyc01/DbCtyc01/ DcCtyc01	Sw1 / Sw2 / Sw3 3rd childcare provider type
MaCtmc01/DbCtmc01/ DcCtmc01	Sw1 / Sw2 / Sw3 3rd childcare provider - no of hours per week
MaCdyc01/DbCdyc01/ DcCdyc01	Sw1 / Sw2 / Sw3 3rd childcare provider - no of days per week
MaCtyd01/DbCtyd01/ DcCtyd01	Sw1 / Sw2 / Sw3 4th childcare provider type
MaCtmd01/DbCtmd01/ DcCtmd01	Sw1 / Sw2 / Sw3 4th childcare provider - no of hours per week
MaCdyd01/DbCdyd01/ DcCdyd01	Sw1 / Sw2 / Sw3 4th childcare provider - no of days per week
MaCtye01/DbCtye01/ DcCtye01	Sw1 / Sw2 / Sw3 5th childcare provider type
MaCtme01/DbCtme01/ DcCtme01	Sw1 / Sw2 / Sw3 5th childcare provider - no of hours per week
MaCdye01/DbCdye01/ DcCdye01	Sw1 / Sw2 / Sw3 5th childcare provider - no of days per week
MdCsta01	Md Whether still using 1st provider from last sweep
MdCcta01	Md Previous 1st ccare provider - revised hrs at Sw4
MdCcda01	Md Previous 1st ccare provider - revised days at Sw4

MdCrSa01	Md - Why not using prev provider 1 at Sw4
MdCStb01	Md Whether still using 2nd provider from last sweep
MdCctb01	Md Previous 2nd ccare provider - revised hrs at Sw4
MdCcdb01	Md Previous 2nd ccare provider - revised days at Sw4
MdCrSb01	Md - Why not using prev provider 2 at Sw4
MdCStc01	Md Whether still using 3rd provider from last sweep
MdCctc01	Md Previous 3rd ccare provider - revised hrs at Sw4
MdCcdc01	Md Previous 3rd ccare provider - revised days at Sw4
MdCrSc01	Md - Why not using prev provider 3 at Sw4
MdCStd01	Md Whether still using 4th provider from last sweep
MdCctd01	Md Previous 4th ccare provider - revised hrs at Sw4
MdCcdd01	Md Previous 4th ccare provider - revised days at Sw4
MdCrSd01	Md - Why not using prev provider 4 at Sw4
MdCSte01	Md Whether still using 5th provider from last sweep
MdCcte01	Md Previous 5th ccare provider - revised hrs at Sw4
MdCcde01	Md Previous 5th ccare provider - revised days at Sw4
MdCrSe01	Md - Why not using prev provider 5 at Sw4
DdCstp01	Dd Whether any of the previous ccare arrgmts stopped
DdCstp02	Dd No of previous sweep providers stopped
DdCnpv01	Dd No of ccare provs from last sweep still being used
DdCapv01	Dd Whether resp still uses a previous ccare provider

The second section is concerned with the details of **new arrangements** which were in place at sweep 4. These variables include details of the provider type, the number of hours and days per week they look after the child, the child's age when the arrangement commenced and the reasons given for using the provision. Details of the variables are listed in Table 7.4.

Table 7.4 Variables for exploring new childcare arrangements at sweep 4

Variable name	Description
MdCany02	Md If no ccare at last sweep whether using ccare at Sw4
MdCany03	Md If ccare at last sweep - any new prov at Sw4
MdCtya01	Md New provider 1 - type
MdCtma01	Md 1st new ccare provider - hours per week
MdCdya01	Md 1st new ccare provider - number of days per week
MdCaga01	Md Age (months) started new provider 1
MdCwya01 – MdCwya18	Md Reasons for using 1 st new provider
MdCtyb01	Md New provider 2 - type
MdCtmb01	Md 2nd new ccare provider - hours per week
MdCdyb01	Md 2nd new ccare provider - number of days per week
MdCagb01	Md Age (months) started new provider 2
MdCwyb01 – MdCwyb18	Md Reasons for using 2 nd new provider
MdCtyc01	Md New provider 3 - type
MdCtMd01	Md 3rd new ccare provider - hours per week
MdCdyc01	Md 3rd new ccare provider - number of days per week
MdCagc01	Md Age (months) started new provider 3
MdCwyc01 – MdCwyc18	Md Reasons for using 3 rd new provider
MdCtyd01	Md New provider 4 - type

MdCtmd01	Md 4th new ccare provider - hours per week
MdCdyd01	Md 4th new ccare provider - number of days per week
MdCagd01	Md Age (months) started new provider 4
MdCwyd01 – MdCwyd18	Md Reasons for using 4 th new provider
DdCnnp01	Dd No of new childcare arrangements at Sweep 4

Information from the first two sections was used to derive a set of variables forming the third section – **current arrangements**. These derived variables indicate - for all childcare arrangements in place at the time of the sweep 4 interview - the provider type, number of hours and days of the arrangement, and whether or not it is a new arrangement at sweep 4. A range of summary variables indicating, for example, use of any childcare, total number of providers, total hours looked after by all providers and use of different provision are also included. These variables are detailed in Table 7.5.

Table 7.5 Variables for exploring current childcare arrangements at sweep 4

Variable name	Description
DdCtya01	Dd - Childcare prov A: provider type
DdCnwa	Dd - Provider A: new or existing
DdCtma01	Dd Provider A: No of hours per week
DdCdya01	Dd Provider A: No of days per week
DdCtyb01	Dd - Childcare prov B: provider type
DdCnwb	Dd - Provider B: new or existing
DdCtmb01	Dd Provider B: No of hours per week
DdCdyb01	Dd Provider B: No of days per week
DdCtyc01	Dd - Childcare prov C: provider type
DdCnwc	Dd - Provider C: new or existing
DdCtmc01	Dd Provider C: No of hours per week
DdCdyc01	Dd Provider C: No of days per week
DdCtyd01	Dd - Childcare prov D: provider type
DdCnwd	Dd - Provider D: new or existing
DdCtmd01	Dd Provider D: No of hours per week
DdCdyd01	Dd Provider D: No of days per week
DdCtye01	Dd - Childcare prov E: provider type
DdCnwe	Dd - Provider E: new or existing
DdCtme01	Dd Provider E: No of hours per week
DdCdye01	Dd Provider E: No of days per week
DdCtyf01	Dd - Childcare prov F: provider type
DdCnwf	Dd - Provider F: new or existing
DdCtmf01	Dd Provider F: No of hours per week
DdCdyf01	Dd Provider F: No of days per week
DdCtyg01	Dd - Childcare prov G: provider type
DdCnwg	Dd - Provider G: new or existing
DdCtmg01	Dd Provider G: No of hours per week
DdCdyg01	Dd Provider G: No of days per week
DdCany01	Dd Whether resp uses regular CCare at Sw4 (not including the excluded pre-school cases – see 7.6.1)
DdCtot01	Dd Number of ccare providers at Sw4 (not including the excluded pre-school

	cases – see 7.6.1)
--	--------------------

Although not listed in Table 7.5, this section also covers variables associated with cost, availability, choice and preferences. Details of these questions and the corresponding variables are available in the sweep 4 questionnaire which accompanies this user guide.

7.6.1 Childcare and Pre-school arrangements

Children in the child cohort at the interviews for sweeps 2 and 3, and children in the birth cohort at Sweep 4, were aged between 3 and 5 years old. At this age, children in Scotland are eligible for funded pre-school places in private and education authority run nursery classes, nursery schools, and playgroups. It became clear on analysis of data from sweep 2 that a number of parents whose children were attending pre-school had not provided those pre-school details in the childcare section. The exclusion of these pre-school arrangements from the childcare data meant that data on the proportion of parents using childcare, the number of providers being used, the mix of provision and the total number of hours, was inaccurate in that it missed the pre-school arrangement.

To resolve this, a number of derived variables were created which incorporated information from the pre-school module and created a more accurate picture of current childcare use amongst parents.

Similar corrections for Pre-school omissions were carried out at Sweep 4. The individual details of the pre-school place as childcare at sweep 4 have been derived for the birth cohort using the information from Sweep 4 variables *MdPRyn01* 'Md Child currently attends pre-school' and *MdPRty01* 'Md - Type of pre-school place'. For the one case in the child cohort which did not mention any childcare arrangement at Sweep 4 and had not started primary school yet, it was assumed that the child was still at Pre-school and the pre-school information was taken from Sweep 2 variables *MbPRyn01* 'Mb Child attends pre-school at Sw2' and *MbPRty01* 'Mb Pre-school provider type'.

The final set of derived variables detailing the *current* childcare arrangements for all cases at sweep 4, including the pre-school omissions, are listed in Table 7.6.

Table 7.6 Childcare variables including a correction for the excluded pre-school cases

Variable name	Description
DdCany02	Whether or not using childcare (including those who had excluded pre-school arrangements)
DdCtot02	Number of childcare providers being used at sw3 (including previously excluded pre-school arrgts)
DdCPrSpv	Sw4 Pre-School provider type for those who did not provide pre-school details in childcare section
DdCPrSHr	No of hours looked after per week by the missed Pre-School childcare provider
DdCPrSDy	No of days looked after per week by the missed Pre-School childcare provider
DdCtmi01	No of hrs child looked after by someone else (average week)
DdCtmi02	No of hrs child looked after by someone else in an average week - BANDED
DdCday01	Highest number of days per week in any one childcare arrangement
DdCtyp01	Does respondent use grandparents for childcare?

DdCtyp02	Does respondent use another relative for childcare?
DdCtyp03	Does respondent use private creche/nursery for childcare?
DdCtyp04	Does respondent use a childminder for childcare?
DdCtyp05	Does respondent use a local authority playgroup for childcare?
DdCtyp06	Does respondent use a local authority nursery for childcare?
DdCtyp07	Does respondent use a private playgroup for childcare?
DdCtyp08	Does respondent use a community/voluntary playgroup for childcare?
DdCtyp09	Does respondent use an ex-spouse or partner for childcare?
DdCtyp10	Does respondent use the child's older sibling for childcare?
DdCtyp11	Does respondent use a friend or neighbour for childcare?
DdCtyp12	Does respondent use a daily visiting nanny for childcare?
DdCtyp13	Does respondent use a live-in nanny for childcare?
DdCtyp14	Does respondent use a babysitter for childcare?
DdCtyp15	Does respondent use a workplace creche or nursery for childcare?
DdCtyp16	Does respondent use a family centre for childcare?
DdCtyp17	Does respondent use a nursery class attached to a primary school for childcare?
DdCtyp18	Does respondent use an agency carer?
DdCtyp19	Does respondent use another type of childcare provider for childcare?
DdCtyp20	Does respondent currently use OTHER INFORMAL childcare?
DdCtyp21	Does respondent currently use NURSERY OR CRECHE for childcare?
DdCtyp22	Does respondent currently use PLAYGROUP for childcare?
DdCtyp33	Does respondent currently use CHILDREN S CLUB for childcare?
DdCtyp23	Does respondent currently use OTHER PROVIDERS for childcare?
DdCtyp30	Does respondent currently use informal childcare?
DdCtyp31	Does respondent currently use formal childcare?
DdCtyp32	Current use of formal and informal childcare

7.7 Indicators and summary variables

7.7.1 Socio-economic characteristics: National Statistics Socio-economic Classification (NS-SEC)

The National Statistics Socio-economic Classification (NS-SEC) is a social classification system that attempts to classify groups on the basis of employment relations, based on characteristics such as career prospects, autonomy, mode of payment and period of notice. There are fourteen operational categories representing different groups of occupations (for example higher and lower managerial, higher and lower professional) and a further three 'residual' categories for full-time students, occupations that cannot be classified due to a lack of information or other reasons. The operational categories may be collapsed to form a nine, eight, five or three category system.

The Growing Up in Scotland dataset includes the five category system in which respondents and their partner, where applicable, are classified as managerial and professional, intermediate, small employers and own account workers, lower supervisory and technical, and semi-routine and routine occupations. A sixth category 'never worked' is also coded on this variable. The decision on whether or not this category should be included as a separate category, incorporated with

category 5 'Semi-routine or routine' or set to 'missing' is dependent on the particular analysis to which it is being applied.

Further information on NS-SEC is available from the National Statistics website at:

<http://www.ons.gov.uk/about-statistics/classifications/current/ns-sec/index.html>

7.7.2 Socio-economic characteristics: Equivalised household annual income

The income that a household needs to attain a given standard of living will depend on its size and composition. For example, a couple with dependent children will need a higher income than a single person with no children to attain the same material living standards. "Equivalisation" means adjusting a household's income for size and composition so that we can look at the incomes of all households on a comparable basis. Official income statistics use the 'Modified OECD' equivalence scale, in which an adult couple with no dependent children is taken as the benchmark with an equivalence scale of one. The equivalence scales for other types of households can be calculated by adding together the implied contributions of each household member from the table below.

Table 7.7 Income equivalence scales for household members

Household member	Equivalence scale
Head	0.67
Subsequent adults	0.33
Each child aged 0-13	0.20
Each child aged 14-18	0.33

For example, a household consisting of a single adult will have an equivalence scale of 0.67 - in other words he or she can typically attain the same standard of living as a childless couple on only 67 percent of its income. In a household consisting of a couple with one child aged three, the head of the household would contribute 0.67, the spouse 0.33, and the child 0.20, giving a total equivalence scale of 1.20. In other words this household would need an income 20 percent higher than a childless couple to attain the same standard of living.

The distribution of income for the population of the United Kingdom as a whole is taken from the most recent available data from the Family Resources Survey. The data and methodology are the same as those used by the Government in its annual Households Below Average Income publication.

GUS collects a banded version of total net household income from all sources in the main CAPI interview. This income data is adjusted, using the above equivalence scale, according to the characteristics of the household, to produce an equivalised annual household income value. Variables with the full equivalised income scale (DdEqvinc) and quintiles of the scale (DdEqv5) are available in the datasets.

7.7.3 Socio-economic characteristics: Index of Material Deprivation

The full suite of items from the DWP Index of Material Deprivation, as first used in the 2004/2005 Family Resources Survey (McKay and Collard, 2004), was included in the sweep four questionnaire. A similar index has also been used in the Families and Children Study (Willits, 2006).

Most material deprivation measures generally ask respondents about the ownership of items regarded as 'necessities' by a majority of the population. People are then classified as 'deprived' if they go without some of these items. Poverty measures based on this type of information are also

known as consensual poverty measures. Essentially, the absence of items is taken to reflect deprivation and the greater the number of items absent, the greater the degree of deprivation. In most research nowadays, respondents are asked to clarify if they do not have or consume an item whether this is because: a) they do not 'need' it, or b) they 'cannot afford' it. It is, therefore, possible to distinguish between 'unenforced' and 'enforced' hardship.

Goodman and Myck (2005) highlight the benefit of a material deprivation measure over income, they state: 'we can use material deprivation as a proxy for long-term financial status...' and conclude 'material deprivation seems to contain some additional information about a family's financial well-being, over and above the information summarised in the level of current disposable income.'

Table 7.8 Variables associated with the Index of Material Deprivation

Variable name	Description
Family:	
MdEmd01	One-week holiday
MdEmd02	Friends or family for a meal once a month
MdEmd03	Two pairs of all weather shoes
MdEmd04	Home in decent state of decoration
MdEmd05	Household contents insurance
MdEmd06	Regular savings
MdEmd07	Replace worn-out furniture
MdEmd08	Replace/repair major electrical goods
MdEmd09	Money to spend on self
MdEmd10	Hobby/leisure activity
MdEmd11	Keep home warm
Child	
MdEmd12	One-week holiday
MdEmd13	Enough bedrooms
MdEmd14	Leisure equipment
MdEmd15	Celebrations
MdEmd16	Swimming once a month
MdEmd17	Hobby/leisure activity
MdEmd18	Friends round for tea
MdEmd19	Toddler group/nursery/playgroup once a week
MdEmd20	School trips
MdEmd21	Access to safe outdoor space nearby

7.7.4 Area-level variables: Scottish Government Urban/Rural Classification

The Scottish Government Urban Rural Classification was first released in 2000 and is consistent with the Government's core definition of rurality which defines settlements of 3,000 or less people to be rural. It also classifies areas as remote based on drive times from settlements of 10,000 or more people. The definitions of urban and rural areas underlying the classification are unchanged.

The classification has been designed to be simple and easy to understand and apply. It distinguishes between urban, rural and remote areas within Scotland and includes the following categories:

Table 7.9 Scottish Government Urban Rural Classification

Classification	Description
1. Large Urban Areas	Settlements of over 125,000 people
2. Other Urban Areas	Settlements of 10,000 to 125,000 people
3. Accessible Small Towns	Settlements of between 3,000 and 10,000 people and within 30 minutes drive of a settlement of 10,000 or more
4. Remote Small Towns	Settlements of between 3,000 and 10,000 people and with a drive time of over 30 minutes to a settlement of 10,000 or more
5. Accessible Rural	Settlements of less than 3,000 people and within 30 minutes drive of a settlement of 10,000 or more
6. Remote Rural	Settlements of less than 3,000 people and with a drive time of over 30 minutes to a settlement of 10,000 or more

For further details on the classification see Scottish Government (2008) *Scottish Government Urban Rural Classification 2007 – 2008*. This document is available online at <http://www.scotland.gov.uk/Publications/2008/07/29152642/0>

7.7.5 Area-level variables: Scottish Index of Multiple Deprivation

The Scottish Index of Multiple Deprivation (SIMD) 2006 identifies small area concentrations of multiple deprivation across Scotland. It is based on 37 indicators in the seven individual domains of Current Income, Employment, Health, Education Skills and Training, Geographic Access to Services (including public transport travel times for the first time), Housing and a new Crime Domain. SIMD 2006 is presented at data zone level, enabling small pockets of deprivation to be identified. The data zones, which have a median population size of 769, are ranked from most deprived (1) to least deprived (6,505) on the overall SIMD and on each of the individual domains. The result is a comprehensive picture of relative area deprivation across Scotland. The classificatory variable contained in the GUS Sweep 4 datasets is based on the 2006 version of SIMD. It should be noted that the analyses in the GUS Sweep 1 report are based on the 2004 version of SIMD as the 2006 version had not been published at the time the report was being written.

In the dataset, the data zones are grouped into quintiles. Quintiles are percentiles which divide a distribution into fifths, i.e., the 20th, 40th, 60th, and 80th percentiles. Those respondents whose postcode falls into the first quintile are said to live in one of the 20% least deprived areas in Scotland. Those whose postcode falls into the fifth quintile are said to live in one of the 20% most deprived areas in Scotland.

Further details on SIMD can be found on the Scottish Government Website <http://www.scotland.gov.uk/Topics/Statistics/SIMD/Overview>

7.7.6 Area-level variables: Carstairs Index

The Carstairs and Morris index was originally developed in the 1980s using 1981 census data. It is composed of four indicators at postcode sector level that were judged to represent material disadvantage in the population (Lack of car ownership, Registrar General Social Class, Overcrowded households and male unemployment). The index has also been calculated based on 1991 and 2001 census data. It is often used in health-related research. Further information can be found on the website of the NHS Information Services Division here:

http://www.show.scot.nhs.uk/publications/isd/deprivation_and_health/background.HTM

7.7.7 Height and weight measurements: Body Mass Index (BMI) scores

Body Mass Index (BMI), i.e. weight divided by height squared, is a score that adjusts your weight for your height. Taken as a number in isolation, the BMI it does not actually represent anything medically. It is only meaningful in the context of a distribution of values for a population. Individuals are then placed into centile bands to show where they stand in relation to the rest of the population, in particular whether they have unusually high or low BMI.

It was initially developed as an index using adult data and in adults BMI stays fairly constant on average as people get older. Therefore BMI centiles for adults ignore age and calculate the same BMI for two people with the same weight and height regardless of the differences in their ages.

BMI is now being used more commonly among children. However among young children in particular, BMI changes quite significantly as the child ages. Since to have a certain BMI at one age may be the norm but be unusually high or low at another age, different centiles are calculated for different ages.

While the BMI measure has come under some scrutiny for not always being accurate, it remains the best non-invasive measure for obesity. Furthermore, a review of the measure by (Reilly *et al.*, 1999) in the *British Medical Journal* suggests that the BMI is more likely to understate, rather than overstate, the true levels of obesity, as has been discussed by Prentice (Prentice, 1998) and Barlow and Dietz (Barlow & Dietz, 1998).

The main child overweight and obesity variables have been produced using the International Obesity Taskforce cut-offs. These cut-offs are based on BMI reference data from six different countries around the world (over 190,000 subjects in total aged 0 to 25 from UK, Brazil, Hong Kong, the Netherlands, Singapore, and the United States). In summary, the BMI percentile curves that pass through the values of 25 and 30 kg/m² (standard adult cut-off points for overweight and obesity, respectively) at age 18 were smoothed for each national dataset and then averaged. The averaged curves were then used to provide age and sex-specific BMI cut-off points for children and adolescents aged 2 to 18. By averaging the distribution curves from each reference country, the international cut-offs for children purport to be representative of the countries but independent of the overweight or obesity level in each country. One of the benefits of using these international standards is the possibility of making international comparisons. However, the international classification is not without problems: international reference data differ from those for the UK population, and this is reflected in the sex-specific overweight and obesity estimates produced by the International classification.

In light of this lack of consensus on its use, variables have also been produced using the 85th (overweight cut-off) / 95th (obesity cut-off) BMI percentiles of the UK reference curves (referred to as the National BMI percentiles classification). The National BMI percentiles classification has been used in the past to describe childhood overweight and obesity prevalence trends in the UK and the 85th / 95th cut-off points are commonly accepted thresholds used to analyse overweight and obesity in children (detail on relevant cut-offs and their descriptions are included below). The National BMI percentiles classification has been shown to be reasonably sensitive (i.e. not classifying obese children as non-obese) and specific (i.e. not classifying non-obese children as obese). A key issue to bear in mind however is that the National BMI percentiles classification are based on the arbitrary assumption that the prevalence of overweight and obesity at the point when the reference data was compiled was 15% and 5%, respectively. Furthermore, there seems to be no indication that these cut-off points relate directly or indirectly to any physiological outcomes or health or disease risks. It is worth noting that the UK component of the international classification used the same sample as that used to construct the UK reference BMI data.

In addition to these International and National BMI classifications, the Information Services Division (ISD) at the Scottish Government uses an alternative method to produce BMI centiles, (Cole's LMS method) which takes into account the fact that BMI data does not follow a normal distribution. Further information can be found at <http://www.isdscotland.org/isd/3640.html>

Note that only those height and weight measurements considered by the interviewer to be reliable were used to calculate the BMIs.

Percentile cut-off	Description
At or below 5th percentile	Underweight
Above 5th percentile and below 85th percentile	Healthy weight
At or above 85th percentile and below 95th percentile	Overweight
At or above 95th percentile and below 98th percentile	Obese
At or above 98th percentile	Morbidly obese

Table 7.10 Derived BMI variables

Variable name	Description
DdBMI	Dd BMI (reliable child weight measurements only)
DdUKbmi	Dd UK BMI national classification standards
DdINTbmi	Dd International BMI cut-offs
DdINTbmi2	Dd BMI status (ovrwt inc. obese) - international cut-offs
DdINTbmi3	Dd BMI status (non-obese vs obese) - international cut-offs
DdISDbmi	Dd ISD BMI 5 group classification
DdISDbcBMI	Dd ISD BMI 5 group classification (excl. far outliers)
DdISDHWT	Dd Study child weight within/outwith ISD healthy range
DdISDcHWT	Dd Study child weight within/outwith ISD healthy range (excl. far outliers)
DdISDovW	Dd Study child overweight, including obese (ISD)
DdISDcOvW	Dd Study child overweight, including obese (ISD excl. far outliers)

7.7.8 Child Development: Strengths and Difficulties Questionnaire

Parents in the child cohort completed the Strengths and Difficulties Questionnaire (SDQ). The SDQ is a brief behavioural screening questionnaire designed for use with 3-16 year olds (Goodman, 1997). The scale includes 25 questions which are used to measure five aspects of the child's development – emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and pro-social behaviour. A score is calculated for each aspect, as well as an overall 'difficulties' score which is generated by summing the scores from all the scales except pro-social. For all scales, except pro-social where the reverse is true, a higher score indicates greater evidence of difficulties. The dataset includes the constituent items, and the derived variables including the various composite scores and total score. Details of these variables are included in Table 7.11 with syntax illustrated in the derived variables documentation.

Table 7.11 Derived variables associated with the Strengths and Difficulties Questionnaire

Variable name	Description
DdDsdem1	Dd SDQ: Emotional symptoms score
DdDsDdo1	Dd SDQ: Conduct problems score
DdDsdhy1	Dd SDQ: Hyper-activity or inattention score
DdDsdpr1	Dd SDQ: Peer problems score
DdDsdps1	Dd SDQ: Pro-social score
DdDsdto1	Dd SDQ: Total difficulties score

Further details on the SDQ can be found at:

<http://www.sdqinfo.com/>

7.7.9 Parental Health: Depression, Anxiety and Stress Scale

Six items from the Depression, Anxiety and Stress (DASS) scale (Lovibond & Lovibond, 1995) were included in the self-completion section of the interview. DASS is available in a 42-item, or 21-item scale in full. We took 6 items: 3 measuring stress, and 3 measuring depression. These items can be combined to create a stress scale and depression scale. Standardized versions of the scales (z-scores) can be combined to produce a single scale measuring evidence of negative emotional symptoms in the respondent. The constituent items and the derived scale variables are detailed in Table 7.12 below. Syntax for compiling the derived variables is detailed in the derived variables documentation.

Table 7.12 Constituent and derived variables associated with the Depression, Anxiety and Stress scale

Variable name	Description
MdHdas01	I found myself getting upset by quite trivial things (stress)
MdHdas02	I found it difficult to relax (stress)
MdHdas03	I felt that I had nothing to look forward to (depression)
MdHdas04	I felt sad and depressed (depression)
MdHdas05	I found that I was very irritable (stress)
MdHdas06	I was unable to become enthusiastic about anything (depression)
DdHdas01	DASS Raw Stress Score
DdHdas02	DASS Raw Depression Score (0-9)
ZDdHdas01	Standardised DASS Stress Score
ZDdHdas02	Standardised DASS Depression Score
DdHdas03	Composite DASS score

Further information on DASS is available at:

<http://www2.psy.unsw.edu.au/groups/dass/>

7.7.10 Risk and protectiveness: the Parent Supervision Attributes Profile Questionnaire

To explore parental perceptions of risk and protective parenting behaviour, six items from the Parent Supervision Attributes Profile (PSAPQ) questionnaire were included (Morrongiello and Corbett, 2006). Initially developed as a measure for assessing injury risk due to inadequate supervision, the full questionnaire comprises 29 items tapping protectiveness, supervision beliefs), tolerance for children's risk taking, and extent of belief in fate as the primary determinant of children's safety. Three items were drawn from the protectiveness sub-scale, and three from the supervision sub-scale. These items were selected on the basis that they captured more general attitudes to protectiveness and supervision less directed at the specific risk of physical injury.

Table 7.13 Constituent variables associated with the PSAPQ

Variable name	Description
MdPsup01	I can trust my child to play by <i>^himself</i> without constant supervision
MdPsup02	I stay close enough to my child so that I can get to him/her quickly
MdPsup03	I think of all the dangerous things that could happen
MdPsup04	I make sure I know where my child is and what he/she is doing
MdPsup05	I keep my child from playing rough games or doing things where <i>^he might get hurt</i> "
MdPsup06	I feel very protective of my child

7.8 Dropped Variables

All variables in the questionnaire documentation with '[not in dataset]' next to their name have been deleted from the archived dataset (or have been transformed into derived variables instead).

The following types of variables have been deleted or replaced with a derived variable coded into broader categories in order to reduce the potential to identify individuals:

1. Those containing text
2. Those which contained a personal identifier (e.g. name/address)
3. Those considered to be disclosive, such as:
 - Detailed ethnicity
 - Detailed religion
 - Language spoken at home
 - Full interview date
 - Full date of birth
 - Timing variables

There are no geographical variables in the archived dataset beyond area urban-rural classification and Scottish index of multiple deprivation summary variable.

7.9 Missing values conventions

- 1 Not applicable: Used to signify that a particular variable did not apply to a given respondent usually because of internal routing.
- 8 Don't know, Can't say.
- 9 No answer/ Refused

These conventions have also been applied to most of the derived variables. The derived variable specifications should be consulted for details.

8 Documentation

The documentation has been organised into the following sections:

- Survey materials containing interviewer and coding instructions.
- Data documentation containing the questionnaire with variable names added, the list of variables in the dataset (including derived variables), a separate list of derived variables with their SPSS syntax and the show cards.

9 References

Barlow, S.E., and Dietz W.H. (1998) "Obesity evaluation and treatment: Expert Committee recommendations, The Maternal and Child Health Bureau, Health Resources and Services Administration and the Department of Health and Human Services", *Pediatrics*, 102(3): E29

Goodman, R. (1997) "The Strengths and Difficulties Questionnaire: a research note", *Journal of Child Psychology and Psychiatry*, 38, pp581-586

Goodman, A. and Myck, M. (2005). *Parental income and child outcomes: what can we learn from material deprivation?*, Mimeos, Institute for Fiscal Studies.

Lovibond, S.H. & Lovibond, P.F. (1995). *Manual for the Depression Anxiety Stress Scales*. (2nd. Ed.) Sydney: Psychology Foundation

McKay, S. and Collard, S. (2004) *Developing deprivation questions for the Family Resources Survey*, Department for Work and Pensions Working Paper Number 13. HMSO.

Morrongiello, B.A. and Corbett, M. (2006) "The Parent Supervision Attributes Profile Questionnaire: a measure of supervision relevant to children's risk of unintentional injury", *Injury Prevention*, 12, pp19–23

Prentice, A.M. (1998) 'Body mass index standards for children' *British Medical Journal*, 317, pp1401-2.

Willits, M. (2006) *Measuring child poverty using material deprivation: possible approaches*, Department for Work and Pensions Working Paper No.28, HMSO

Reilly, JJ., Dorosty, A.R., and Emmett, P.M. (1999) "Prevalence of overweight and obesity in British children: cohort study", *BMJ*, 319, pp1039-1039

10 Related publications

Further information about GUS Sweep 4 is available in:

Barnes, M., Chanfreau, J. and Tomaszewski, W. (2010) *Growing Up in Scotland: The circumstances of persistently poor children*, Edinburgh: Scottish Government
<http://www.scotland.gov.uk/Publications/2010/04/26095519/0>

Bradshaw, P. and Tipping, S. (2010) *Growing Up in Scotland: Children's social, emotional and behavioural characteristics at entry to primary school*, Edinburgh: Scottish Government
<http://www.scotland.gov.uk/Publications/2010/04/26102809/0>

Bromley, C. and Cunningham-Burley, S. (2010) *Growing Up in Scotland: Health inequalities in the early years*, Edinburgh: Scottish Government
<http://www.scotland.gov.uk/Publications/2010/04/26103009/0>

Marryat, L. and Martin, C. (2010) *Growing Up in Scotland: Maternal mental health and its impact on child behaviour and development*, Edinburgh: Scottish Government
<http://www.scotland.gov.uk/Publications/2010/04/26102536/0>

Other publications which include the use of GUS data include:

Bradshaw, P., Sharp, C, Webster, C. and Jamieson, L. (2009) *Growing Up in Scotland: Parenting and the Neighbourhood Context*, Edinburgh: Scottish Government
<http://www.scotland.gov.uk/Publications/2009/03/13143448/11>

Bradshaw, P. and Wasoff, F. (2009) *Growing Up in Scotland: Multiple Childcare Provision and its Effects on Child Outcomes*, Edinburgh: Scottish Government
<http://www.scotland.gov.uk/Resource/Doc/263884/0079032.pdf>

Bromley, C. (2009) *Growing Up in Scotland: the Impact of Children's Early Activities on Cognitive Development*, Edinburgh: Scottish Government
<http://www.scotland.gov.uk/Resource/Doc/263956/0079071.pdf>

Marryat, L., Reid, S. and Wasoff, F. (2009) *Growing Up in Scotland Sweep 3 Non-resident Parent Report*, Edinburgh: Scottish Government
<http://www.scotland.gov.uk/Publications/2009/01/21085002/0>

Marryat, L., Skafida, V. and Webster, C. (2009) *Growing Up in Scotland Sweep 3 Food and Activity Report*, Edinburgh: Scottish Government
<http://www.scotland.gov.uk/Publications/2009/01/21085143/0>

Bradshaw P, Cunningham-Burley S, Dobbie F, McGregor A, Marryat L, Ormston, R. and Wasoff F. (2008) *Growing Up in Scotland: Sweep 2 Overview Report*, Edinburgh: The Scottish Government

Anderson S, Bradshaw P, Cunningham-Burley S, Hayes F, Jamieson L, McGregor A, Marryat L and Wasoff F. (2007) *Growing Up in Scotland: Sweep 1 Overview Report*, Edinburgh: The Scottish Executive

Bradshaw, P. with Jamieson, L. and Wasoff, F. (2008) *Use of informal support by families with young children*, Edinburgh: Scottish Government

Bradshaw, P. and Martin, C. with Cunningham-Burley, S. (2008) *Exploring the experience and outcomes for advantaged and disadvantaged families* Edinburgh: Scottish Government

Jamieson, L. with Ormston, R. and Bradshaw, P. (2008) *Growing Up in Rural Scotland*, Edinburgh: Scottish Government

Skafida, V. (2008) "Breastfeeding in Scotland: The impact of advice for mothers", *Centre for Research on Families and Relationships, Briefing 36, February 2008*, Edinburgh: Centre for Research on Families and Relationships

Skafida, V. (2009) "The relative importance of social class and maternal education for breast-feeding initiation", *Public Health Nutrition*, First View Article, published 26 Feb 2009

Links to these reports and others, along with additional related information are available on the GUS website: <http://www.growingupinScotland.org.uk/>

11 Contact details

Contacts at the Scottish Centre for Social Research, 73 Lothian Road, Edinburgh, EH3 9AW

GUS Project Manager	Paul Bradshaw	0131 221 2564	paul.bradshaw@scotcen.org.uk
GUS Senior Researcher	Louise Marryat	0131 221 2557	louise.marryat@scotcen.org.uk
GUS Data Manager	Mireille Ferrandon	0131 221 2578	mireille.ferrandon@scotcen.org.uk

Appendix A: Full non-response models

Table A1 Non-response model for birth cohort (Sample A)

	B	S.E.	Wald	df	Sig.	Exp(B)
Age of mother (grouped)			22.8	4	0.000	
<20					(baseline)	
20-24	0.36	0.21	3.0	1	0.083	1.44
25-29	0.65	0.22	8.3	1	0.004	1.91
30-34	0.87	0.23	13.8	1	0.000	2.40
35+	1.14	0.26	18.5	1	0.000	3.12
Number of children in the household			7.2	3	0.067	
1					(baseline)	
2	0.13	0.15	0.7	1	0.389	1.14
3	-0.09	0.19	0.2	1	0.618	0.91
4+	-0.49	0.24	4.1	1	0.042	0.61
Highest education level of respondent			9.2	4	0.055	
Degree or equivalent					(baseline)	
Vocational qualification below degree	-0.04	0.19	0.1	1	0.822	0.96
Higher Grade or equivalent	0.09	0.29	0.1	1	0.761	1.09
Standard Grade or equivalent	-0.16	0.22	0.5	1	0.461	0.85
No Qualifications	-0.57	0.24	5.8	1	0.016	0.57
Ethnicity of mother			5.0	1	0.026	
White					(baseline)	
Other ethnic background	-0.58	0.26	5.0	1	0.026	0.56
Household employment			28.6	2	0.000	
At least one parent/carer in full-time employment					(baseline)	
At least one parent/carer in part-time employment	-0.40	0.19	4.6	1	0.033	0.67
Household employment	-0.87	0.16	28.6	1	0.000	0.42
Urban/rural indicator (Scotland)			10.2	4	0.037	
1 Large urban area (125,000+)					(baseline)	
2 Other urban area (10,000-125,000)	0.35	0.15	5.5	1	0.019	1.42
3,4,5 All small town (3,000-10,000)	0.61	0.23	7.0	1	0.008	1.84
6 Accessible rural (<3,000)	0.17	0.23	0.6	1	0.445	1.19
7, 8 very remote and Remote rural (<3,000)	0.42	0.36	1.3	1	0.250	1.52
Constant	2.15	0.27	65.4	1	0.000	8.58

Notes:

1. The response is 1 = sample A response to wave 4, 0 = sample A non-response.
2. Model is weighted by wave 2 baby weight
3. The model $R^2 = 0.04$ (Cox and Snells).
4. B is the estimate coefficient with standard error S.E.

5. The Wald-test measures the impact of the categorical variable on the model with the appropriate number of degrees of freedom df. If the test is significant ($\text{sig} < 0.05$) then the categorical variable is considered to be 'significantly associated' with the response variable and therefore included in the model.

6. The Wald test for each level of the categorical variable is also shown. This tests the difference between that level and the baseline category.

Table A2 Non-response model for child cohort (Sample A)

	B	S.E.	Wald	df	Sig.	Exp(B)
Tenure			6.7	2	0.036	
Owner occupiers					(baseline)	
Rents HA/council	-0.53	0.206	6.7	1	0.010	0.59
Rents private	-0.20	0.299	0.4	1	0.505	0.82
Scottish Index of Multiple Deprivation			10.6	1	0.001	
Does not live in 15% most deprived					(baseline)	
Data Zone					(baseline)	
Lives in 15% most deprived						
Data Zones	-0.66	0.203	10.6	1	0.001	0.52
Ethnicity of mother			8.5	1	0.004	
White					(baseline)	
Other ethnic background	-0.93	0.320	8.5	1	0.004	0.39
Respondent NS-SEC (5 groups)			15.2	5	0.010	
Managerial and professional occupations					(baseline)	
Intermediate occupations	0.20	0.283	0.5	1	0.479	1.22
Small employers and own account workers	0.15	0.447	0.1	1	0.732	1.17
Lower supervisory and technical occupations	-0.33	0.352	0.9	1	0.345	0.72
Semi-routine and routine occupations	-0.45	0.224	4.0	1	0.044	0.64
Missing/never worked	-1.02	0.332	9.5	1	0.002	0.36
Number of children in the household			7.9	3	0.047	
1					(baseline)	
2	0.30	0.187	2.6	1	0.104	1.35
3	0.67	0.247	7.4	1	0.007	1.96
4+	0.48	0.322	2.2	1	0.140	1.61
Urban/rural indicator (Scotland)			7.9	4	0.095	
1 Large urban area (125,000+)					(baseline)	
2 Other urban area (10,000-125,000)	-0.40	0.190	4.4	1	0.035	0.67
3,4,5 All small town (3,000-10,000)	0.20	0.294	0.5	1	0.500	1.22
6 Accessible rural (<3,000)	-0.34	0.281	1.5	1	0.228	0.71
7, 8 very remote and Remote rural (<3,000)	0.14	0.495	0.1	1	0.782	1.15
Constant	2.92	0.242	145.2	1	0.000	18.58

Notes:

1. The response is 1 = sample A response to wave 4, 0 = sample A non-response.
2. Model is weighted by wave 2 toddler weight
3. The model $R^2 = 0.04$ (Cox and Snells).
4. B is the estimate coefficient with standard error S.E.
5. The Wald-test measures the impact of the categorical variable on the model with the appropriate number of degrees of freedom df. If the test is significant (sig < 0.05) then the categorical variable is considered to be 'significantly associated' with the response variable and therefore included in the model.
6. The Wald test for each level of the categorical variable is also shown. This tests the difference between that level and the baseline category.

Table A3 Distribution of sample A

	BIRTH COHORT			CHILD COHORT		
	Wave 3	Wave 4	Wave 4	Wave 3	Wave 4	Wave 4
	weighted by W3 weight	weighted by W3 weight	weighted by W4 weight	weighted by W3 weight	weighted by W3 weight	weighted by W4 weight
	%	%	%	%	%	%
Tenure						
Owner occupier	62.8	64.6	63.0	63.1	65.0	63.1
Rents HA/council	28.4	27.0	28.4	28.2	26.3	28.2
Rents private	8.8	8.4	8.6	8.7	8.7	8.7
family status						
Lone parent	19.6	18.2	19.4	22.1	21.1	22.1
Couple parent	80.4	81.8	80.6	77.9	78.9	77.9
Whether child was mother's first-born						
First born	49.8	49.8	50.0	47.6	47.2	47.6
Other children	50.2	50.2	50.0	52.4	52.8	52.4
Mother's age						
<20	7.6	6.9	7.6	6.9	6.3	6.6
20-24	17.3	16.7	17.3	18.4	17.4	18.1
25-29	23.6	23.6	23.6	22.9	22.9	22.9
30-34	31.0	31.7	31.0	32.5	33.4	32.9
35+	20.5	21.1	20.5	19.3	20.0	19.6
Highest education level of respondent						
Degree or equivalent	27.4	28.2	27.5	28.2	29.2	28.3
Vocational qualification below degree	37.9	38.2	37.9	38.9	39.1	39.3
Higher Grade or equivalent	7.8	7.9	7.8	7.1	7.3	7.2
Standard Grade or equivalent	17.9	17.4	17.9	15.9	15.4	15.7
No Qualifications	9.0	8.2	8.9	10.0	9.1	9.6
Household income						
<£10,000	12.8	11.9	12.9	12.3	11.4	12.1
£10,000-£19,999	19.8	19.3	19.8	18.3	17.8	18.4
£20,000-£31,999	22.4	22.8	22.4	22.3	22.6	22.5
£32,000+	38.6	39.7	38.6	39.2	40.7	39.5
Missing	6.3	6.3	6.3	7.8	7.5	7.6

Table A3 Distribution of sample A (continued)

	BIRTH COHORT			CHILD COHORT		
	Wave 3 weighted by W3 weight	Wave 4 weighted by W3 weight	Wave 4 weighted by W4 weight	Wave 3 weighted by W3 weight	Wave 4 weighted by W3 weight	Wave 4 weighted by W4 weight
	%	%	%	%	%	%
Respondent NSSEC						
Managerial and professional occupations	34.7	35.6	34.7	34.4	35.5	34.4
Intermediate occupations	19.4	20.0	19.8	17.5	18.1	17.5
Small employers and own account workers	5.5	5.6	5.5	5.8	6.0	5.8
Lower supervisory and technical occupations	5.8	5.6	5.6	6.0	6.0	6.0
Semi-routine and routine occupations	30.5	29.7	30.5	32.1	30.9	32.2
Missing/never worked	4.1	3.5	3.9	4.1	3.5	4.0
Ethnicity of respondent						
White	96.2	96.4	96.2	96.3	96.7	96.3
Other ethnic background	3.8	3.6	3.8	3.7	3.3	3.7
Household employment						
At least one parent/carer in full-time employment	72.7	74.5	72.8	73.0	74.6	73.3
At least one parent/carer in part-time employment	12.1	11.8	12.1	12.7	12.4	12.8
No parent/carer working	15.2	13.7	15.1	14.3	13.0	13.9
Mother's employment status						
Childs mother working - full-time	14.6	14.9	14.6	16.3	16.6	16.3
Childs mother working - part-time	48.2	48.9	48.0	48.6	49.7	49.1
Childs mother not working	37.3	36.2	37.4	35.1	33.7	34.6
Number of children in the household						
1	32.3	32.1	32.3	23.0	22.3	23.0
2	44.6	45.3	44.6	50.9	51.3	50.9
3	17.1	17.0	17.1	19.6	19.9	19.6
4+	6.0	5.7	6.0	6.6	6.4	6.5

Table A3 Distribution of sample A (continued)

	BIRTH COHORT			CHILD COHORT		
	Wave 3	Wave 4	Wave 4	Wave 3	Wave 4	Wave 4
	weighted	weighted	weighted	weighted	weighted	weighted
	by W3	by W3	by W4	by W3	by W3	by W4
	weight	weight	weight	weight	weight	weight
	%	%	%	%	%	%
Urban/rural indicator (Scotland)						
1 Large urban area (125,000+)	38.5	36.3	37.1	35.3	35.1	35.3
2 Other urban area (10,000- 125,000)	31.2	32.7	32.5	32.6	32.0	32.5
3 Accessible small town (3,000-10,000)	9.6	10.5	10.3	11.0	11.2	10.9
4 Remote small town (3,000- 10,000)	0.9	1.0	1.0	1.3	1.5	1.4
5 Very remote small town (3,000-10,000)	1.8	1.8	1.8	1.3	1.4	1.4
6 Accessible rural (<3,000)	13.6	12.1	11.9	13.4	13.6	13.5
7 Remote rural (<3,000)	2.1	3.1	3.0	2.8	2.9	2.8
8 Very remote rural (<3,000)	2.1	2.5	2.4	2.2	2.3	2.2
Use regular childcare						
Yes	76.5	76.9	76.4	76.9	77.3	77.2
No	23.5	23.1	23.6	23.1	22.7	22.8
SIMD06 quintiles						
0.95 - 7.75 (least deprived)	18.7	19.3	18.8	19.9	20.9	20.2
7.75 - 13.56	19.9	20.3	19.9	20.9	21.4	20.8
13.56 - 21.04	19.3	19.5	19.2	19.8	20.1	19.7
21.05 - 33.70	18.2	18.3	18.5	17.0	17.0	17.0
33.71 -89.09 (most deprived)	24.0	22.7	23.7	22.3	20.8	22.3
SIMD06 Flag lowest 15%						
Less deprived 85% datazones	82.7	83.7	83.0	83.2	84.7	83.2
Most deprived 15% datazones	17.3	16.3	17.0	16.8	15.3	16.8
Regularly attended toddler groups in the last year						
Yes	39.5	40.1	39.5	43.2	43.8	43.3
No	60.5	59.9	60.5	56.8	56.2	56.7
Was child breastfed						
Yes	60.0	60.7	59.8	59.2	60.7	60.1
No	40.0	39.3	40.2	40.8	39.3	39.9
Base (unweighted)		3844	3844		2100	2100

Table A4 Birth cohort - weighted distribution of key variables for samples A and B

	Sample A	Combined Wave 4 sample (A+B)	
	Weighted by Wave 4 weight	Weighted by pre-calibration weight ¹	Calibrated to sample A
Family type			
Lone parent	18.87	19.32	18.87
Couple family	81.13	80.68	81.13
Household income			
<£10,000	10.32	10.73	10.32
£10,000-£19,999	20.07	20.15	20.07
£20,000-£31,999	22.98	22.96	22.98
£32,000+	41.43	40.88	41.43
Missing	5.20	5.28	5.20
Respondent's age			
<25	10.96	11.04	10.96
25-29	18.66	18.93	18.66
30-34	25.07	24.98	25.07
35-39	29.65	29.45	29.65
40+	15.66	15.60	15.66
Respondent employment status			
Employed	64.91	64.65	64.91
Not employed	35.09	35.35	35.09
SIMD06 quintiles			
0.95 - 7.75 (least deprived)	18.76	18.40	18.76
7.75 - 13.56	19.86	19.60	19.86
13.56 - 21.04	19.22	19.44	19.22
21.05 - 33.70	18.49	18.58	18.49
33.71 -89.09 (most deprived)	23.66	23.97	23.66
Base (unweighted)	3844	3994	3994

¹This is the W4 weight for Sample A and the weight from the last completed wave for Sample B

Table A5 Child cohort - weighted distribution of key variables for samples A and B

	Sample A	Combined Wave 4 sample (A+B)	
	Weighted by Wave 4 weight	Weighted by pre-calibration weight ¹	Calibrated to sample A
Family type			
Lone parent	21.27	21.52	21.27
Couple family	78.73	78.48	78.73
Household income			
<£10,000	10.74	11.02	10.74
£10,000-£19,999	18.53	18.90	18.53
£20,000-£31,999	21.85	21.60	21.85
£32,000+	42.28	41.60	42.28
Missing	6.60	6.87	6.60
Respondent's age			
<25	4.24	4.39	4.24
25-29	16.89	17.33	16.89
30-34	21.48	21.28	21.48
35-39	32.26	32.25	32.27
40+	25.12	24.75	25.12
Respondent employment status			
Employed	66.12	65.94	66.12
Not employed	33.88	34.06	33.88
SIMD06 quintiles			
0.95 - 7.75 (least deprived)	20.19	19.96	20.19
7.75 - 13.56	20.82	20.85	20.82
13.56 - 21.04	19.70	19.73	19.70
21.05 - 33.70	17.02	16.90	17.02
33.71 -89.09 (most deprived)	22.26	22.55	22.26
Base (unweighted)	2100	2200	2200

¹This is the W4 weight for Sample A and the weight from the last completed wave for Sample B

Appendix B: Issues to be aware of when working with the data

The large number of checks undertaken on the data ahead of its deposit occasionally brings to light quality or validity issues which should be taken into account when analysis is being undertaken on the related variables. We have listed these issues below.

- Case IDnumber 1001409 was incorrectly coded as belonging to the Birth cohort for the CAPI interview. As a result there may be some missing information relating to that case whenever the routing in the questionnaire used 'SampType=2'. Alternatively there can be more information than expected if included in questions meant only for the birth cohort, for example at MdObtg01-95.
- The year has been entered incorrectly as 2997 for case IDnumber 1007206 at variable MdPcls01 (in the dataset as Derived Variable DdPcls01 'Year started parenting class/grp').
- Dates have been entered incorrectly in CAPI for ten cases at variable MdPRwn01 (in the dataset as Derived Variable DdPRwnY 'Year child started pre-school'). The incorrect years are 1997, 2200, 2820, 2997, 2998, 4200 and 5200.
- Due to a modification in the CAPI programme during field work the routing for MdPRprv1 ('whether pre-school provider used for childcare before the study child's 3rd birthday') does not always match the routing mentioned in the Questionnaire.
- Cases IDnumber 1001331 and 1005725 show a total number of hours of childcare above 168 hours (= 24 hrs x 7) per week at the Derived Variable DdCtmi01 'Number of hrs child looked after by someone else in an average week'.
- In the Health and development section, the CAPI programme used individual variables for each type of long standing illness, the set of variable being repeated three times since up to three illnesses could be recorded. The total number of illnesses mentioned over the combined 3 sets of CAPI variables was never higher than 3. However there could be more than one illness mentioned in a particular set. In the archived dataset, the individual variables for each set have been recoded into one variable for the 1st long standing illness, one for the 2nd and one for the 3rd. When 2 illnesses had been mentioned in the 1st set of CAPI answers, the 1st answer was kept as the 1st illness and the 2nd answer became the 2nd illness. If there was an answer in the 2nd set of illnesses, it was recoded as being the 3rd illnesses. This recoding means that for the 2nd and 3rd illnesses there will be a higher number of illnesses mentioned in MdHlsb01 and MdHlsc01 than answers at the next variables MdHlsb02 and MdHlsc02 (whether the illness limits the child's activities).
- Case IDnumber 1006203 seems an extreme upper outlier with 30 contacts/visits at MdHcon94 'Contact with someone else re child's health in the last six months'.
- There are some extreme upper outliers at MdAtv09 'Number of hours the child watches TV on a weekday': thirteen cases mention 20 to 24 hours on an average week day. And twenty-nine cases mention 40 to 45 hours watching TV over the two days of the week-end at MdAtv10.

- Although the Self-complete section was asked to all respondents, 22 respondents chose not to complete it and these cases show as missing values ('Not Applicable') in the dataset.
- Seven cases had a partial interview (code 210 at variable MdOutcom), so some information may be missing towards the end of the questionnaire (for religion and ethnicity in particular); those cases show either as -1 'Not Applicable' or as -3 'information not available' in the dataset.
- There were some unlikely extreme lower and upper outliers when the respondent gave an estimated height in meters at WdZeht02, with values such as 0.10 m and 0.11 m, or up to 1.70 m and 2 meters. This is probably due to the lack of familiarity with the metric scale as the estimates in feet/inches look more reliable. An extreme upper outlier was also entered at the estimated weight in kg WdZewt02, with 116.1 kg. These outliers affect the corresponding CAPI derived variables DdZeht05 and DdZewt05.