

Growing Up in Scotland Sweep 7: 2012-13

User Guide

ScotCen Social Research
Scotiabank House
6 South Charlotte Street
Edinburgh, EH2 4AW
T 0131 240 0210
www.scotcen.org.uk

A Company Limited by Guarantee. Registered in England No.4392418.
A Charity registered in England and Wales (1091768) and Scotland (SC038454)

Contents

1	Overview of the survey	1
1.1	Study design	1
1.2	Sample Design	2
1.3	Developing and Piloting	4
2	Data Collection Methods	5
2.1	Mode of data collection	5
2.2	Child Interview	5
2.3	Length of Interview	6
2.4	Timing of Fieldwork	6
3	Height and Weight Measurements	7
4	Response Rates	8
5	Coding and editing	9
6	Weighting the data	10
6.1	Background	10
6.1.1	Weights developed for SW7	10
6.2	Weights for main carer interview data	10
6.2.1	Main carer sample	10
6.2.2	Longitudinal weights for main carer interview data	11
6.2.3	Cross-sectional weights for main carer interview data	12
6.2.4	Sample efficiency of main carer interview data	13
6.3	Weights for child interview data	13
6.3.1	Weighting the child interview (ACASI) data	13
6.3.2	Sample efficiency of the child interview data	14
6.4	Applying the weights	14
7	Using the data	15
7.1	Variables on the data file	15
7.2	Variable naming convention	15
7.3	Variable labels	16
7.4	Derived variables	16
7.5	Multicoded questions	17
7.6	Indicators and summary	17
7.6.1	Household data	17

7.6.2	Socio-economic characteristics: National Statistics Socio-economic Classification (NS-SEC).....	18
7.6.3	Socio-economic characteristics: Equivalised household annual income	18
7.6.4	Area-level variables: Scottish Government Urban/Rural Classification ..	19
7.6.5	Area-level variables: Scottish Index of Multiple Deprivation	20
7.6.6	Area-level variables: Carstairs Index	20
7.6.7	Area-level variables: Scottish Health Board indicator.....	20
7.6.8	Child Development: Strengths and Difficulties Questionnaire	21
7.6.9	Parenting: Alabama Parenting Questionnaire	22
7.6.10	Parenting: Pianta Child-Parent Relationship Scale.....	24
7.6.11	Height and weight measurements: Body Mass Index (BMI) scores	24
7.7	Dropped variables.....	27
7.8	Missing values conventions	27
8	Documentation	28
9	References	29
10	Related publications.....	30
11	Contact details	31
12	Appendix	32

1 Overview of the survey

The overarching aim of the Growing Up in Scotland study is set out in its purpose, which is:

“To generate, through robust methods, specifically Scottish data about outcomes throughout childhood and into adulthood for children growing up in Scotland across a range of key domains:

- *Cognitive, social, emotional and behavioural development*
- *Physical and mental health and wellbeing*
- *Childcare, education and employment*
- *Home, family, community and social networks*
- *Involvement in offending and risky behaviour*

Such data will encompass, in particular, topics where Scottish evidence is lacking and policy areas where Scotland differs from the rest of the UK.”

At sweep 7 data collection for the study included three main elements:

1. A face-to-face CAPI (Computer Assisted Personal Interview) interview with the cohort child’s main carer
2. A self-complete Audio-CASI (Computer Assisted Self-Complete Interview) interview with the cohort child
3. Height and weight measurement of the cohort child

1.1 Study design

The survey was initially based on two cohorts of children: the first aged approximately 10 months at the time of first interview (involving around 5217 children at the first sweep) and the second aged approximately 34 months (involving around 2800 children at the first sweep). A second birth cohort of 6127 children aged around 10 months at was recruited in 2011 with children. All cohorts were named samples drawn from Child Benefit records.

The configuration of cohorts and sweeps for all sweeps of data collection launched to date is summarised below. BC1 refers to the younger of the two cohorts (‘birth cohort 1’), CC1 to the slightly older cohort (‘child cohort’) and BC2 to the most recent birth cohort (‘birth cohort 2’).

Sweep Launch year	Cohort and age at interview									
	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
1 2005/06	BC1		CC1							
2 2006/07		BC1		CC1						
3 2007/08			BC1		CC1					
4 2008/09				BC1		CC1				
5 2009/10					BC1		-			
6 2010/11						BC1		-		
1 (BC2) 2011/12	BC2						-		-	
7 2012/13		-						BC1		-
2 (BC2)/7.5 (BC1) 2013/14			BC2						BC1 w-c*	
2.5 (BC2)/8 BC1 2014/15				BC2 w-c*						BC1
3 (BC2) 2015/16					BC2					

*'w-c' indicates 'web-CATI' data collection. These sweeps involved shorter questionnaires issued initially as web surveys. Participants who did not respond to the web survey were then contacted by telephone and invited to complete the questionnaire with a telephone interviewer.

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

- Cross-sectional time specific data – e.g. what proportion of 2-3 year-olds were living in single parent families in 2005?
- Cross-sectional time series data – e.g. is there any change in the proportion of 10 month old children living in single parent families between 2005 and 2012?
- 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 7-8?

1.2 Sample Design¹

The sample for all cohorts was recruited at sweep 1. There has been no sample refreshment.

The initial 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.

¹ Note that the sample design for BC2 varies slightly. Information is provided in the user guide accompanying the BC2 sweep one dataset.

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 Authority² 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.

² 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 BC1	
Sample Number	Dates of Birth required Birth Cohort 1
1	01-June-2004 - 30-Jun-2004
2	01-Jul-2004 - 31-Jul-2004
3	01-Aug-2004 - 31-Aug-2004
4	01-Sep-2004 - 30-Sep-2004
5	01-Oct-2004 - 31-Oct-2004
6	01-Nov-2004 - 30-Nov-2004
7	01-Dec-2004 - 31-Dec-2004
8	01-Jan-2005 - 31-Jan-2005
9	01-Feb-2005 - 28-Feb-2005
10	01-Mar-2005 - 31 Mar-2005
11	01-Apr-2005 - 30-Apr-2005
12	01-May-2005 - 31-May-2005

1.3 Developing and Piloting

Policy priorities and key topics of interest for the sweep 7 adult and child questionnaires were initially discussed and agreed by the study's Scottish Government Project Managers and a number of internal and external stakeholders. The questionnaires were then developed by the GUS team at ScotCen with input from colleagues at the MRC Social and Public Health Sciences Unit and Centre for Research on Families and Relationships (CRFR) in reference to these priorities and topics. Cognitive testing was carried out for the child questionnaire and mode in October 2011 and January 2012. A full instrument with both adult and child questionnaires was initially piloted in CAPI/CASI in November 2011. This instrument was revised for the second CAPI/CASI 'Dress Rehearsal' Pilot in March 2012.

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 adult respondent, using the laptop, input their responses directly into the questionnaire program. Also, for the first time, the cohort children were interviewed directly. The children completed a short self-complete questionnaire using an audio-CASI approach (see section 2.2 below).

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 7, this means the same respondent as Sweep 6 (or Sweep 5 / Sweep 4 / Sweep 3 / 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% of interviews were with the previous respondent) and this was usually the child's mother (97% of interviews were with the child's mother).

2.2 Child Interview

As noted above the cohort children were interviewed directly for the first time at sweep 7. The children participated by answering questions themselves on the interviewer's laptop using an audio-CASI (A-CASI) approach. In this approach, as well as the questions and response options being displayed on screen, a recording of an interviewer reading them is also available. The children wore headphones whilst completing the questionnaire so that they could listen to the recordings. Informed consent was gained from both the main carer and from the child³.

The child questionnaire consisted of the following sections:

- 1) A short interviewer-led section including an introduction, consent procedures and practice questions.
- 2) An audio-CASI section with questions on the following topics: school, friends, parents/family life, materialism, and wellbeing.

³ Further information about consent procedures and administration of the Audio-CASI program can be found in the Project Instructions .

The child data collection was funded by the Medical Research Council Social and Public Health Sciences Unit in Glasgow and Education Scotland.

2.3 Length of Interview

Overall, the average interview (including adult and child interviews and height and weight measurements) lasted around 70 minutes. The median interview length was 73 minutes.

2.4 Timing of Fieldwork

Fieldwork was undertaken between May 2012 and May 2013. The sample was issued in eleven monthly sweeps 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 June 2012 and remained in field until mid-August 2012. For previous GUS sweeps, each monthly sample contained only children born in a specific month. However, a delay to fieldwork launch meant that three monthly samples were combined into two at the beginning of fieldwork. The first sample issued contained all children born in June and those born in the first half of July; the second contained those born in the second half of July and all of those in August. The rest of the samples were issued as in previous sweeps.

To ensure that interviews took place when the cohort children were approximately the same age, each case was assigned a 'target interview date'. This was identified as the date on which the child turned 94.5 months old (around 6 weeks before their 8th birthday). 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. As such, almost all children were interviewed just before their 8th birthday.

3 Height and Weight Measurements

Child's height and weight measurements were previously taken in SW2, SW4, SW6, and were also included at SW7. The main carer's height and weight measurements were also taken at SW6, but were not taken at SW7.

The interviewers were asked to measure the height and weight of all children. 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 WgXhei14 and WgXwei19. 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 WgXhei01 and WgXwei01, with the reason for refusal at WgXhei02 or WgXwei02.

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.

4 Response Rates

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

Table 4.1 Number of issued and achieved cases and response rates	
	Birth Cohort
Achieved interviews at sweep 1	5217
Achieved interviews at sweep 2	4512
Achieved interviews at sweep 3	4193
Achieved interviews at sweep 4	3994
Achieved interviews at sweep 5	3833
Achieved interviews at sweep 6	3657
Cases to field at sweep 7:	
All issued to field*	4159
Total cases achieved at sweep 7	3456
Main carer interviews achieved at sweep 7**	3453
Child interviews achieved at sweep 7	3371
Response rate	
As % of all issued cases at sweep 7	83%
As % of all sweep 1 cases	66%

* The number of cases issued to the field at sweep 7 is higher than the number of Interviews achieved at sweep 6 because some of the sweeps 1 to 6 respondents missed at sweep 6 were re-issued at sweep 7.

** In 3 cases only the child completed an interview.

5 Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted. The GUS Sweep 7 Coding Instructions, deposited along with this User Guide, provide details of the tasks that were conducted.

6 Weighting the data

6.1 Background

6.1.1 Weights developed for SW7

Four weights were developed for Sweep 7 of BC1. Two weights were generated for analysis of information collected during the main interview with the main carer, plus two additional weights for analysis of data collected from the child using ACASI. SW7 was the first sweep to collect information directly from the study child.

The four weights were thus:

- A cross-sectional weight for adults that should be used for any cross-sectional analysis of data collected in the Sweep 7 main carer interview. All main carers that responded at SW7 have a cross-sectional adult weight.
- A longitudinal weight for analysis of main carers with data at all waves. Main carers that have responded at every sweep of GUS have a longitudinal adult weight.
- A cross-sectional weight that should be used for any cross-sectional analysis of the Sweep 7 ACASI data (i.e. data collected from the child). All children that completed the ACASI interview have a cross-sectional child weight.
- A longitudinal weight for analysis of ACASI data for children who responded to the ACASI and whose main carer has responded at every sweep.

The SW7 interview followed up all main carers who responded at the SW6 interview and gave NatCen permission to be re-contacted. In addition, main carers who had no interview at SW6 but had responded previously were also contacted if they had given permission to do so.

6.2 Weights for main carer interview data

6.2.1 Main carer sample

The SW7 sample of adult respondents can be split into two components. For the purposes of describing the weighting these have been named Sample A and Sample B and are defined as follows:

- Sample A – adults who had responded at all previous sweeps
- Sample B – adults who had responded at Sweep 1 but had missed one or more interviews in Sweeps 2-6

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 those in Sample A, as suggested by their much lower response rates. There were 789 individuals in Sample B, 334 (42%) of which responded at SW7. The response rates for Sample A were much higher at 94%. The issued and responding sample sizes are given in Table 1.

Table 6.1 Response rates for different samples of main interview respondents

	Issued	Responding	Response rate
Sample A	3331	3119	94%
Sample B	789	334	42%
Combined (A+B)	4120	3453	84%

Two sets of weights were developed for the responding adults: a cross-sectional weight and a longitudinal weight. The longitudinal weight will be used for any analysis that includes more than one sweep of data. Only members of Sample A (who have responded at every sweep of GUS) will have a longitudinal weight. This weight is described in more detail in Section 2.1.

6.2.2 Longitudinal weights for main carer interview data

Longitudinal weights were only generated for respondents in Sample A. A model-based weighting technique was used to develop the SW7 longitudinal weights, where response behaviour is modelled using data from previous sweeps. This is the same method used to generate weights for adults who completed the main interview at sweeps 2 to 6. Ineligible households (deadwood) were not included in the non-response modelling.

Response behaviour was modelled using logistic regression. A logistic regression models the relationship between an outcome variable (in this case response to the SW7 interview) and a set of predictor variables. The predictor variables were a set of socio-demographic respondent and household characteristics collected from the previous two 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. 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.

A summary of the characteristics related to response behaviour for the birth cohort at SW7 are given in Table 6.2. The full model is given in Table A1 in the Appendix.

Table 6.2 Characteristics associated with response and non-response

Characteristics associated with response	Characteristics associated with non-response
Owner occupiers	Family rents accommodation
Mother aged 20 or over at time of birth	Mother aged under 20 at time of birth
Only one child in the household	More than one child in the household
Couple household	Lone parent household
Degree or equivalent qualifications	Standard level or no qualifications
Child had no accidents in the past 12 months	Child had at least one accident in the past 12 months
Child's general health is good	Child's general health is very good or poor
Child was read stories in the past week	Child was not read stories in the past week
Family residence is not in one of the 15% most deprived data zones	Family residence is in one of the 15% most deprived data zones

The final SW7 weight is the product of the SW7 non-response weight and the SW7 interview weight. For each cohort the final weights were scaled to the responding SW7 sample size, this makes the weighted sample size match the unweighted sample size.

6.2.3 Cross-sectional weights for main carer interview data

Cross-sectional weights were generated for all respondents at Sweep 7 (the combined A and B samples) and should be used for any cross-sectional analysis of Sweep 7 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; respondent employment status; respondent age; household income and whether the respondent was a lone parent.

The pre-calibration weights were the Sweep 7 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 Table A3 in the appendix.

6.2.4 Sample efficiency of main carer interview data

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 effect 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.

	Minimum	Maximum	Mean	N	Neff	Efficiency
Main carer longitudinal weight	0.60	3.83	1.00	3119	2594	83.2%
Main carer cross-sectional weight	0.60	3.75	1.00	3453	2955	85.6%

6.3 Weights for child interview data

6.3.1 Weighting the child interview (ACASI) data

For the first time in GUS, in Sweep 7 children were asked to fill in a short self-completion questionnaire. This was done using ACASI. A large proportion of children completed the questionnaire - 97% of those whose main carer had completed the main CASI interview.

Calibration methods were used to generate non-response weights for the children. Non-response modelling was not used because the child responses did not nest well within the adult responses (there were a small number of households where the child completed the interview but the adult did not), otherwise we would have used adult interview data to model non-response to the ACASI interview. The high response rate would also have made it difficult to generate a robust model.

The profile of the responding children was calibrated to a set of key respondent and household characteristics that were significantly related to non-response behaviour. These were: mother's age, household income, respondent's work status, tenure and local urban/rural indicators. The calibration totals were weighted survey estimates taken from the Sweep 7 adult interview data.

Two sets of weights were generated: a set of longitudinal weights, these are weights for children who completed the ACASI and whose parents had completed every wave of GUS so far, and a set of cross-sectional weights, these are weights for children who completed ACASI but whose parents had missed one or more waves of interviewing.

There were two children in the longitudinal sample and three children in the cross-sectional sample whose parents had not completed the Sweep 7 adult interview. These children were calibrated using information on mother's age and urban/rural only, since they were missing up to date information on the other measurements.

For each cohort the final weights were scaled to the responding SW7 ACASI sample size, this makes the weighted sample size match the unweighted sample size.

6.3.2 Sample efficiency of the child interview data

The range of the weights, the effective sample size and sample efficiency for both sets of ACASI weights are given in Table 6.4.

	Minimum	Maximum	Mean	N	Neff	Efficiency
ACASI longitudinal weight	0.60	4.01	1.00	3057	2865	92%
ACASI cross-sectional weight	0.61	3.84	1.00	3374	2523	81%

6.4 Applying the weights

For each sample, the cross-sectional weights should be used for any cross-sectional analysis, i.e. any analysis of Sweep 7 data only. All sample members that responded at SW7 have a cross-sectional weight.

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

Variable name	Label
DgWTbrth	Dg Birth cohort Sweep 7 weight
DgWTbth2	Dg Birth cohort Sweep 7 weight - longitudinal
DgWTchld	Dg Child ACASI Sw7 weight
DgWTchd2	Dg Child ACASI Sw7 weight - longitudinal

7 Using the data

The GUS Sweep 7 data consists of the following SPSS file:

GUS_SW7_B.sav	3456 cases	Birth cohort
---------------	------------	--------------

7.1 Variables on the data file

The data file contains 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. Please note that variable descriptions in the variable list cannot be relied upon to capture the detail of the question wording, or the answer categories used. For the precise question wording, please refer to the interview documentation.

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 MgSDQ01 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 normally 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 – BC1

Character no.:			
1		2	
Source of data		Sweep/Sweep	
Non-sequential Capitals: D,M, P, S		Sequential lower case: a, b, c..	
Source code	Details	Sweep code	Child's age
AL	Area Level variable	a	10 months
D	Derived variable	b	Almost 2 years
DP	Derived variable from partner int	c	Almost 3 years
DWP	DWP variable	d	Almost 4 years
M	Main carer/adult interview	e	Almost 5 years
P	Partner interview	f	Almost 6 years
C	Child interview	g	Almost 8 years

7.3 Variable labels

In the Sweep 7 dataset the variable labels have been shortened 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 Multicoded questions

Some questions in the survey enabled participants to give more than one answer. In the dataset each of the answer options has been converted into a binary variable with the people who selected that option coded 1 and the rest coded 0.

7.6 Indicators and summary

7.6.1 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 MgHGSi(n) can be used to see whether a person who was in the household at a previous sweep is still in the household at sweep 7.

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's and partner's 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	
DgHGnmad	Dg - Number of adults (16 or over) in household
DgHGnmkd	Dg - Number of children in household
DgHGnmsb	Dg - Number of siblings in household
DgHGnp01	Dg - Number of natural parents in household
DgHGrsp01	Dg - Whether respondent is natural mother
DgHGrsp02	Dg - Whether respondent is natural father
DgHGnp02	Dg - Natural mother in household
DgHGnp03	Dg - Natural father in household
DgHGnp04	Dg - Respondent living with spouse/partner
DgMothID	Dg – Mother's ID (= Person number in household)
DgFathID	Dg – Father's ID
DgRespID	Dg – Respondent's ID
DgPartID	Dg – Respondent's partner's ID
DgRPAge	Dg – Respondent's partner's age (banded)
DgRPsex	Dg - Respondent partners sex

7.6.2 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/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html>.

7.6.3 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 (DgEqvinc) and quintiles of the scale (DgEqv5) are available in the datasets.

7.6.4 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 7.8 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>

7.6.5 Area-level variables: Scottish Index of Multiple Deprivation

The Scottish Index of Multiple Deprivation (SIMD) 2009 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 2009 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 7 datasets is based on the 2009 version of SIMD. It should be noted that analyses in various GUS reports may be based on earlier versions of SIMD.

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>

7.6.6 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

7.6.7 Area-level variables: Scottish Health Board indicator

To provide some geographic information which would allow comparison across the sweeps for the Birth Cohort, a Scottish Health Boards derived variable 'ALgHBdBc' has been added to the dataset. In order to reduce the risk of potential disclosure, only those Health Boards which had 250 cases or more in the Birth Cohort at Sweep 1 were identified, the rest being aggregated into a single category called 'Other'. The 9 Health Boards identified, out of the original 14 Scottish Health Boards, are listed in table 7.9 below.

Table 7.9 Scottish Health Boards identified in the dataset

Scottish Health Board (in alphabetical order)	Identified or Aggregated in the dataset
Ayrshire and Arran	Identified
<i>Borders</i>	<i>Aggregated</i>
<i>Dumfries and Galloway</i>	<i>Aggregated</i>
Fife	Identified
Forth Valley	Identified
Grampian	Identified
Greater Glasgow and Clyde	Identified
Highland	Identified
Lanarkshire	Identified
Lothian	Identified
<i>Orkney</i>	<i>Aggregated</i>
<i>Shetland</i>	<i>Aggregated</i>
Tayside	Identified
<i>Western Isles</i>	<i>Aggregated</i>

7.6.8 Child Development: Strengths and Difficulties Questionnaire

The Strengths and Difficulties Questionnaire (SDQ) is a brief behavioural screening questionnaire designed for use with 3-16-year-olds⁴. 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 7.10 with syntax illustrated in the derived variables documentation.

⁴ Goodman, R. (1997) "The Strengths and Difficulties Questionnaire: a research note", *Journal of Child Psychology and Psychiatry*, 38, pp581-586

Variable name	Description
DgDsdem1	Dg SDQ: Emotional symptoms score
DgDsdco1	Dg SDQ: Conduct problems score
DgDsdhy1	Dg SDQ: Hyper-activity or inattention score
DgDsdpr1	Dg SDQ: Peer problems score
DgDsdps1	Dg SDQ: Pro-social score
DgDsdto1	Dg SDQ: Total difficulties score

Further details on the SDQ can be found at:

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

7.6.9 Parenting: Alabama Parenting Questionnaire

The Alabama Parenting Questionnaire (APQ) is used to measure parenting practices amongst parents of 6-18-year-old children (e.g. Frick *et al*, 1999; Shelton *et al*, 1996). The full APQ consists of both an adult and a child questionnaire. Each questionnaire contains 42 statements about parenting practices. For each statement, the parent or child is asked to indicate on a five-point scale from ‘never’ to ‘always’ to what extent the statement is reflective of their own practices (for the parent) or experience (for the child)⁵.

For sweep 7 of GUS, a subset of 14 items were included in the main adult questionnaire (in the self-complete section), and 13 items were included in the child questionnaire. The items have been modified to suit the GUS survey context. For example, the number of answer categories for the child questionnaire has been reduced from five to four, and the wording of a number of items have been changed as a result of testing.

The APQ can be used to construct measures of parenting practices across five dimensions: involvement, positive parenting, supervision and monitoring, use of discipline, and corporal punishment.

⁵ A short 9-item version of the APQ has also been developed (Elgar *et al*, 2006).

Table 7.11 Variables associated with the Alabama Parenting Questionnaire

Variable name	Adult questionnaire		Child questionnaire
Involvement			
MgAPQ4	You ask [child] about his/her day in school	CgPa2	My parents ask about my day in school
MgAPQ8	You help [child] with his/her homework	CgPa6	My parents help me with my homework
MgAPQ14	You play games or do other fun things with [child]	CgPa4	My parents play games or do other fun things with me
Positive parenting			
MgAPQ1	You let [child] know when he/she is doing a good job with something	CgPa7	My parents tell me when I'm doing a good job with something
MgAPQ5	You compliment [child] after he/she has done something well	CgPa5	My parents tell me when I do something well
MgAPQ10	You praise [child] if he/she behaves well	Cgpa1	My parents tell me if I behave well
Inconsistent discipline			
MgAPQ2	You threaten to punish [child] and then do not actually punish him/her	CgPa9	My parents warn me they will do something if I'm naughty but then they don't do it
MgAPQ6	[Child] talks you out of being punished after he/she has done something wrong	CgPa8	I can make my parents change their mind about telling me off
MgAPQ11	[Child] is not told off when he/she has done something wrong	CgPa10	My parents tell me off when I'm <i>really</i> naughty
Monitoring/supervision			
MgAPQ3	When [child] is not at school, you know what he/she is doing	CgPa11	My parents know what I'm doing
MgAPQ7	[Child] can play outside without you being there with him/her	CgPa12	My parents let me play outside without them being there
MgAPQ12	You check to make sure [child] is doing OK	CgPa13	My parents check to make sure I'm doing ok
Corporal punishment			
MgAPQ9	You smack [child] with your hand when he/she has done something wrong	CgPa3	My parents smack me when I have done something wrong
MgAPQ13	You slap [child] whe he/she has done something wrong		

7.6.10 Parenting: Pianta Child-Parent Relationship Scale

The Pianta scale (Pianta, 1992) is used to measure the mother-child relationship at year 8. The scale is constructed using the responses on the extent to which the respondent feels a series of statements apply to her relationship with her child (such as 'I share an affectionate, warm relationship with [my child]').

The full scale has 30 items and looks at 3 dimensions of the relationship – warmth, conflict and dependency. The 15 items included in the sweep 7 GUS questionnaire are a subset of the full scale that were also used in the Millennium Cohort Study (MCS2; 2004/05) and which relate to warmth and conflict. Measures can be constructed for these two dimensions, with a high score corresponding to a high degree of warmth or conflict. Each measure uses seven items, shown below.

Table 7.12 Constituent and derived variables associated with the Pianta Child-Parent Relationship Scale

Variable name	Description
Warmth	
MgPpia01	<i>I share an affectionate, warm relationship with [Child's name]</i>
MgPpia03	<i>[Child's name] will seek comfort from me</i>
MgPpia04	<i>[Child's name] is uncomfortable with physical affection or touch from me</i>
MgPpia05	<i>[Child's name] values his/her relationship with me</i>
MgPpia06	<i>When I praise [Child's name], he/she beams with pride</i>
MgPpia07	<i>[Child's name] spontaneously shares information about [him/herself]</i>
MgPpia09	<i>It is easy to be in tune with what [Child's name] is feeling</i>
MgPpia15	<i>[Child's name] openly shares his/her feelings and experiences with me</i>
Conflict	
MgPpia02	<i>[Child's name] and I always seem to be struggling with each other</i>
MgPpia08	<i>[Child's name] easily becomes angry at me</i>
MgPpia10	<i>[Child's name] remains angry or is resistant after being disciplined</i>
MgPpia11	<i>Dealing with [Child's name] drains my energy</i>
MgPpia12	<i>When [Child's name] wakes up in a bad mood, I know we're in for a long and difficult day</i>
MgPpia13	<i>[Child's name]'s feelings towards me can be unpredictable or can change suddenly</i>
MgPpia14	<i>[Child's name] is sneaky or manipulative with me</i>

7.6.11 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 a person's weight for their 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 placed into bands to show where they stand in relation to the rest of the population, in particular whether they have unusually high or low BMI.

In adults BMI stays fairly constant on average as people get older. Therefore BMI categories 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.

Natural mother's BMI was grouped as follows:

<u>BMI range</u>	<u>Description</u>
Under 18.5	Underweight
18.5 to less than 25	Healthy weight
25 to less than 30	Overweight
25 to less than 30	Obese
40 and over	Morbidly obese

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 (1998) and Barlow and 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 7.13 Child Derived BMI variables

Variable name	Description
DgBMI	Dg BMI (reliable child weight measurements only)
DgUKbmi	Dg UK BMI national classification standards
DgINTbmi	Dg International BMI cut-offs
DgINTbmi2	Dg BMI status (ovrwt inc. obese) - international cut-offs
DgINTbmi3	Dg BMI status (non-obese vs obese) - international cut-offs
DgISDbmi	Dg ISD BMI 5 group classification
DgISDHWt	Dg Study child weight within/outwith ISD healthy range
DgISDovW	Dg Study child overweight, including obese (ISD)

7.7 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, the Scottish index of multiple deprivation summary variable, and a derived variable identifying some of the Scottish Health Board areas.

Access to more detailed variables is possible on application. Please contact the GUS research team if you require such data.

7.8 Missing values conventions

The following missing values conventions have been observed:

- 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 used during the interview.

9 References

Elgar, Frank; Daniel Waschbusch; Mark Dadds; and Nadine Sigvaldson (2006):
Development and Validation of a Short Form of the Alabama Parenting Questionnaire.
Journal of Children and Family Studies, 16:243-259

Frick, P.J., Christian, R.E., & Wootton, J.M. (1999). Age trends in the association
between parenting practices and conduct problems. *Behavior Modification*, 23(1), 106-
128.

Pianta RC. (1992) *Child–Parent Relationship Scale*. Charlottesville, VA: University of
Virginia.

Shelton, K.K., Frick, P.J., & Wootton, J. (1996). The assessment of parenting practices
in families of elementary school-aged children. *Journal of Clinical Child Psychology*, 25,
317-327.

10 Related publications

To date, only one Scottish Government report has been published which utilises sweep 7 data. The report can be found on the Scottish Government website and a link is also available from the Growing Up in Scotland website.

<http://www.growingupinScotland.org.uk/>.

Parkes, A., Sweeting, H. and Wight, D. (2014) *Growing Up in Scotland: Family and school influences on children's social and emotional well-being* Edinburgh: Scottish Government

The GUS website also has links to all other Scottish Government reports using GUS data as well as a wide range of other reports and journal articles which have utilized the data.

11 Contact details

Contacts at ScotCen Social Research, Scotiabank House, 6 South Charlotte St, Edinburgh, EH2 4AW

GUS Project Director paul.bradshaw@scotcen.org.uk	Paul Bradshaw	0131 240 0230
GUS Researcher line.knudsen@scotcen.org.uk	Line Knudsen	0131 240 0228
GUS Data Manager jackie.palmer@scotcen.org.uk	Jackie Palmer	0131 240 0226

12 Appendix

Table A.1 Non-response model for main carer interview data (Sample A)

	B	S.E.	Wald	df	Sig.	Exp(B)
Current tenure			8.6	2	.013	
Owner occupier					(baseline)	
Rents HA/council	-.526	.185	8.1	1	.004	.591
Rents private	-.484	.236	4.2	1	.040	.616
Mothers age at birth (grouped)			32.1	4	.000	
<20 years					(baseline)	
20-24 years	.395	.202	3.8	1	.051	1.484
25-29 years	1.006	.226	19.9	1	.000	2.735
30-34 years	1.046	.232	20.3	1	.000	2.847
35+ years	1.127	.267	17.8	1	.000	3.087
Number of children in hh (grp4+)			15.2	3	.002	
1					(baseline)	
2	-.426	.201	4.5	1	.035	.653
3	-.634	.216	8.6	1	.003	.530
4+	-.936	.253	13.7	1	.000	.392
Family Type: Couple family	.276	.156	3.1	1	.077	1.318
Highest Education level of Respondent			19.4	4	.001	
Degree or equivalent					(baseline)	
Vocational qualification below degree	-.571	.240	5.7	1	.017	.565
Higher Grade or equivalent	-.497	.340	2.1	1	.144	.608
Standard Grade or equivalent	-1.055	.259	16.6	1	.000	.348
No Qualifications	-.908	.287	10.0	1	.002	.403
Had one or more accidents in past 12 months	-.292	.165	3.1	1	.077	.747

Notes:

1. The response is 1 = sample A response to Sweep 7, 0 = sample A non-response.
2. Model is weighted by the Sweep 6 longitudinal weight
3. The model $R^2 = 0.073$ (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 A.2 Distribution of Sample A (main respondent interview data)

	Sweep 6 weighted by sweep 6 (longitudinal) weight	Sweep 7 weighted by sweep 6 (longitudinal) weight	Sweep 7 weighted by sweep 7 (longitudinal) weight
	%	%	%
Tenure			
Owner occupier	63.6	66.1	63.5
Rents HA/council	27.6	25.0	27.3
Rents private	6.9	7.3	7.7
Other		1.5	
Family status			
Lone parent	19.3	18.7	19.7
Couple parent	80.7	81.3	80.3
Mother's age at child's birth			
<20	7.4	6.1	7.1
20-29	40.7	39.9	40.8
30-39	48.6	50.6	48.8
40+	3.3	3.4	3.3
Highest education level of respondent			
Degree or equivalent	28.3	31.3	29.7
Higher grade or Upper level vocational qualification	32.4	32.6	32.3
Upper level Standard Grades and Intermediate Vocational qualifications	23.6	22.1	22.8
Lower level Standard Grades and Vocational qualifications and Other	7.2	6.3	6.8
No Qualifications	8.6	7.7	8.4
Respondent NSSEC - 5 Category			
Managerial and professional occupations	33.1	34.2	32.6
Intermediate occupations	18.8	18.6	18.2
Small employers and own account workers	6.8	7.4	7.3
Lower supervisory and technical occupations	5.2	5.3	5.4
Semi-routine and routine occupations	32.8	32.1	33.7
Never worked	3.3	2.5	2.9
Ethnicity of respondent			
White	96.5	96.7	96.7
Other ethnic background	3.5	3.3	3.3
(continued...)			

Table A.2 Continued

	Sweep 6 weighted by sweep 6 (longitudinal) weight	Sweep 7 weighted by sweep 6 (longitudinal) weight	Sweep 7 weighted by sweep 7 (longitudinal) weight
	%	%	%
Household employment			
At least one parent/carer in full-time employment	70.9	69.3	67.8
At least one parent/carer in part-time employment	15.0	19.4	19.9
No parent/carer working	14.1	11.2	12.3
Mother's employment status			
Childs mother working - full-time (35+ h/wk)	16.4	18.7	18.1
Childs mother working - part-time	50.2	53.8	53.1
Childs mother not working	33.4	27.5	28.8
Number of children in the household			
1	20.1	19.5	19.3
2	51.7	51.3	50.7
3	21.4	21.2	21.3
4+	6.9	7.9	8.7
Urban/rural indicator (Scotland)			
Large urban area (125,000+)	36.1	35.9	36.3
Other urban area (10,000-125,000)	33.2	33.6	33.6
All small town (3,000-10,000)	12.8	13.0	13.0
Accessible rural (<3,000)	12.3	11.4	11.2
Remote rural (<3,000)	5.6	6.1	5.9
SIMD09 quintiles			
Least deprived	19.2	21.3	20.3
2	20.6	21.1	20.4
3	18.8	19.8	19.7
4	19.3	18.4	18.5
Most deprived	22.0	19.4	21.1
<i>Base (unweighted)</i>	3657	3456	3456

Table A.3 Weighted distribution of key variables for samples A and B (main respondent interview data)

	Sample A	Combined Sweep 7 sample (A+B)	
	Weighted by sweep 7 (longitudinal) weight	Weighted by pre-calibration weight ¹	Calibrated to sample A (Sw7 cross-sectional weight)
Family type			
Lone parent	19.2	20.3	19.7
Couple family	80.8	79.7	80.3
Household income			
<£15,599	14.1	14.8	14.5
£15,600-£25,999	21.3	21.7	21.5
£26,000-£51,999	35.6	33.7	34.7
£52,000+	21.0	20.5	20.7
Missing	8.0	9.2	8.6
Respondent's age at time of interview			
<30	12.8	13.9	13.3
30-34	19.7	20.2	20.0
35-39	27.1	26.7	26.9
40-43	24.2	23.4	23.8
44+	16.2	15.8	16.0
Respondent employment status			
Working	71.3	70.4	70.8
Not working	28.7	29.6	29.2
SIMD09 quintiles			
Least deprived	20.6	19.8	20.1
2	20.6	20.3	20.5
3	19.8	19.6	19.7
4	18.4	18.7	18.6
Most deprived	20.6	21.6	21.1
<i>Base (unweighted)</i>	3119	3453	3453

¹This is the SWEEP 7 weight for Sample A and the weight from the last completed sweep 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.

- In the Health and Development section, the CAPI program 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, etc. At Sweep 7 this recoding means that
 - for the 2nd illness there will be a higher number of illnesses mentioned in MgHlsb01 than answers at the next variables regarding this 2nd illness, MeHlsb02 to MeHntb07, and
 - there can be more than 3 illnesses mentioned in total, hence additional variable MgHlsd01, but no corresponding set of answers regarding whether it limits activities or the type of treatment/advice received.
- Although the Self-complete section was asked to all respondents, some respondents chose not to complete it and these cases show as missing values ('Not Applicable') in the dataset.
- Four cases had a partial interview (code 210 or 211 at variable MgOutcom), so some information may be missing towards the end of the interview. These cases show either as -1 'Not Applicable' or as -3 'information not available' in the dataset. Further to this, three cases had a child interview, but no adult interview (code 212 at MgOutcom).