Growing Up in Scotland
Sweep 8: 2014-15

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

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 8 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').

Table 1.1	Study design: ages and stages									
Sweep	Cohort and age at interview									
Launch year	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 the time they are in Primary 6 ?

1.2 Sample design[1]

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

[1] Note that the sample design for BC2 varies slightly. Information is provided in the user guide accompanying the BC2 Sweep 1 dataset.

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. 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 Authority[2] 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 8 adult and child questionnaires were initially discussed and agreed by the study's Scottish Government Project Manager 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 of selected items in the child questionnaire was carried out in March 2014. A full CAPI/CASI instrument with both adult and child questionnaires was piloted across May/June 2014.

1.4 Timing of fieldwork

Sweep 8 fieldwork was issued differently to previous sweeps. In previous sweeps, fieldwork was conducted over a 14 month period with cases issued to field according to the child's age and interviews taking place as around a specified date calculated according to the child's birthday (the 'target interview date'). At Sweep 8 there was interest in interviewing families when the child was in their first term of Primary 6. Therefore, at Sweep 8 fieldwork moved from an 'ages' to a 'stages' approach. This means that the age gap between children at the time of interview is larger at Sweep 8 than at previous sweeps. Conversely,

at Sweep 8, almost all children were in the same school year at the time of interview (i.e. in their first term of Primary 6).

Because of how children were initially sampled, the children in BC1 span two different school years. Therefore, Sweep 8 fieldwork was split into two phases:

Phase 1 fieldwork took place between September 2014 and end of December 2014. 2815 cases were issued for Phase 1 fieldwork (77% of the total sample issued at Sweep 8); 2402 cases were achieved as part of Phase 1 fieldwork (76% of the total number of cases achieved at Sweep 8).

Phase 2 fieldwork took place between September 2015 and end of November 2015. 858 cases were issued for Phase 2 fieldwork (23% of the total sample issued at Sweep 8); 748 cases were achieved as part of Phase 2 fieldwork (24% of the total number of cases achieved at Sweep 8).

1.5 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 1
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
Achieved interviews at Sweep 7	3456
Cases to field at Sweep 8:	
All issued to field*	3673
Total cases achieved at Sweep 8	3150
Main carer interviews achieved at Sweep 8**	3148
Child interviews achieved at Sweep 8	3087
Response rate	
As % of all issued cases at sweep 8	86%
As % of all Sweep 1 cases	60%

* The number of cases issued to the field at Sweep 8 is higher than the number of Interviews achieved at Sweep 7 because some of the Sweeps 1 to 7 respondents missed at Sweep 7 were re-issued at Sweep 8.
** In 2 cases only the child completed an interview.

2 Sweep 8 data collection elements

2.1 Interview with child's main carer

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. The children completed a short self-complete questionnaire using an audio-CASI approach (see section 2.2 below) and also undertook cognitive assessments (described in more detail 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 8, this means the same respondent as Sweep 7 (or Sweep 6 / Sweep 5 / Sweep 4 / Sweep 3 / Sweep 2 / Sweep 1 if the household skipped one or more 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 second time at Sweep 8. 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 child[3].

The child questionnaire consisted of the following sections:

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

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, relationship with parents/carers, and life satisfaction.

2.3 Cognitive assessments

Cognitive assessments were previously carried out with the children in BC1 at sweeps 3 and 5. Cognitive assessments were also carried out at Sweep 8. At Sweep 8, children were assessed using the 'Listening Comprehension' subtest of the Weschler Individual Achievement Tests, 2nd Edition (WIAT-II).

WIAT-II is an educational assessment tool which is widely used by educational psychologists to examine cognitive development and educational ability. The assessments carried out with the GUS children were adapted for use in a survey setting, and modified to be administered in CAPI.

The Listening Comprehension subtest is designed to measure the ability to listen for detail by selecting the picture that matches a word or sentence (e.g. 'point to the dog'), and generating a word that matches a picture and an oral description (e.g. 'what is this?'). There are strict protocols which must be adhered to when administering assessments. These ensure that the resultant data can be confidently compared with the normative data used to produce the various derived scores necessary for analysis.

The Listening Comprehension test includes three sub-assessments: Receptive Vocabulary, Sentence Comprehension and Expressive Vocabulary (see table 2.1 below).

Table 2.1 Child cognitive assessments: WIATT-II Listening Comprehension

Assessment name	Assesses	Method	Max no. of items
Receptive vocabulary	Ability to listen for details and knowledge of words	Child is asked to select a picture that matches a word	16
Sentence comprehension	Ability to listen for details and knowledge of words	Child is asked to select a picture that matches a sentence	10
Expressive vocabulary	Knowledge of words	Child is asked to generate a word that matches a picture and oral description	15

For each assessment, the starting point is determined by the child's age. The assessment continues until the last item or until six consecutive incorrect

responses are given.[4] At GUS sweep 8, all children started at the beginning of each assessment.

The following scores are available in the dataset:

- ***Receptive Vocabulary Raw Score:*** A count of all the items on Receptive Vocabulary the child answered correctly.
- ***Sentence Comprehension Raw Score:*** A count of all the items on Sentence Comprehension the child answered correctly.
- ***Expressive Vocabulary Raw Score:*** A count of all the items on Expressive Vocabulary the child answered correctly.
- ***Listening Comprehension Raw score:*** The raw score is a count of the number of items the child answered correctly. The total raw score for the Listening Comprehension subtest is derived by adding up the raw scores for each of the three sub-assessments (Receptive vocabulary; Sentence comprehension and Expressive vocabulary).
- ***Listening Comprehension Standard Score***: A normalised transformation of the raw score which uses an external standard or 'norming' sample and takes into account the child's age in months at the time the assessment was undertaken. The standard score can be used as a measure of how far a child's score from the mean (and median) score for a child their age, measured in standard deviations. The Listening Comprehension standard score can also be compared to other types of normalised derived scores, like subtest scaled scores from the Wechsler intelligence scales.

For each raw score outlined above it is possible to derive *within-sample standardised z scores* which allow for comparisons to be made across sub-assessments (measures in standard deviations from the mean).

Note that the exercises are designed to provide a picture of the range of skills across a number of children, not to give a clinical assessment of an individual child.

Further information about the WIAT-II measures is available online, at: http://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildAchievementMeasures/WechslerIndividualAchievementTest-SecondUKEdition(WIAT-IIUK).

2.4 Height and weight measurements

Child's height and weight measurements were previously taken in sweeps 2, 4 6 and 7, and were also included at Sweep 8. The main carer's height and weight measurements were taken at Sweep 6 and again at Sweep 8.

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,

[4] Further details are available in the cognitive exercise instructions.

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).

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 (WhXhei14 and WhXwei19. 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 (WhXhei01 and WhXwei01, with the reason for refusal at WhXhei02 or WhXwei02.

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.

2.5 Consent to teacher follow-up

As part of the sweep 8 interview the adult respondent and the cohort child were asked if they were happy to provide details for the child's the child's Primary 6 teacher and for him/her to be invited to complete a paper or online questionnaire about the child. Consent was obtained for 96% of cases.

Data from the GUS Primary 6 Teacher Questionnaire will be made available through the UK Data Service alongside other GUS data.

2.6 Length of interview

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

3 Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted. The GUS Sweep 8 Coding and Editing Instructions, deposited along with this User Guide, provide details of the tasks that were conducted.

4 Weighting the data

4.1 Background

4.1.1 Weights developed for Sweep 8

Four weights were developed for Sweep 8 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. Sweep 8 was the second sweep to collect information directly from the study child (the first one was Sweep 7).

The four weights were:

- A cross-sectional weight for adults that should be used for any cross-sectional analysis of data collected in the Sweep 8 main carer interview. All main carers who responded at Sweep 8 have a cross-sectional adult weight.
- A longitudinal weight for analysis of main carers who have responded at every sweep of GUS.
- A cross-sectional weight that should be used for any cross-sectional analysis of the Sweep 8 child ACASI data (i.e. data collected from the child). All children that completed the ACASI interview at this sweep 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 had responded at every sweep of GUS up to and including Sweep 7.

The Sweep 8 interview follows up all mothers who responded at the Sweep 7 interview and gave NatCen permission to be re-contacted. In addition, mothers who had refused the Sweep 7 interview but had responded at any previous Sweep were contacted if they had given a 'sweep only' refusal at Sweep 7.

4.2 Weights for main carer interview data

4.2.1 Main carer sample

The Sweep 8 sample of adult respondents can be split into two groups. 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 2-7.

The two samples were 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 569 individuals in Sample B, 334 (59%) of which responded at Sweep 8. The response rate for Sample A (3,102) was much higher at 91%. The issued and responding sample sizes are given in Table 6.1.

Table 6.1 Response rates for the two groups of main interview respondents

	Issued	Responding	Response rate
Sample A	3,102	2,815	91%
Sample B	569	334	59%
Combined (A+B)	3,671	3,149	86%

Two sets of weights were developed for the responding adults: a cross-sectional weight and a longitudinal weight. Only members of Sample A (who have responded at every sweep of GUS) received a longitudinal weight. This weight is described in more detail in Section 6.2.1.

All Sweep 8 respondents will have a cross-sectional weight (Sample A + B). These are described in more detail in Section 6.2.2.

4.2.2 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 Sweep 8 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 7. Ineligible households (deadwood) were not included in the non-response modelling.

Response behaviour was modelled using logistic regression. This models the relationship between an outcome variable (in this case response to the Sweep 8 interview) and a set of predictor variables. The predictor variables were a set of socio-demographic individual and household characteristics collected from the previous sweeps of the study (mainly from Sweep 7).

The model generated a predicted probability of response for each individual. A set of non-response weights were generated equal to the inverse of these predicted probabilities; hence respondents who had a lower than average predicted probability received a higher than average weight, increasing their representation in the sample.

Variables found to predict response at Sweep 8 are shown in Table 6.2. All of them were entered in the non-response model which was used to calculate the non-response weights.

Table 6.2 Variables used in adult non-response weighting (longitudinal sample)
Mother's age at cohort child's birth
Highest education level of respondent
NS-SEC of respondent
Whether respondent has a job
Household employment status
Whether respondent has had any accidents/injuries
Self-reported general health of respondent
Whether respondent lives in a deprived area (15% most deprived datazones)
Total number of calls

The final Sweep 8 main carer longitudinal weight was calculated as the product of the non-response weight and the Sweep 7 longitudinal interview weight. The final weights were scaled to the responding Sweep 8 sample size, so that the weighted sample size matches the unweighted sample size.

4.2.3 Cross-sectional weights for main carer interview data

Cross-sectional weights were generated for all respondents at Sweep 8 (the combined A and B samples) and should be used for any cross-sectional analysis of Sweep 8 data.

Calibration weighting was applied to the combined sample (A+B) to create the cross-sectional weights. This method adjusts a set of starting weights using an iterative procedure so that they match pre-defined population totals. The resulting weights, when applied to the combined data, will make the survey estimates match the population estimates which in this instance were calculated from Sample A, weighted by the Sweep 8 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 choice of the variables used in the calibration was dictated by the bias remaining after the Sweep 8 longitudinal weights were applied to Sample A and the cross-sectional weight from the last completed sweep were applied for Sample B. The variables used in the weighting are listed in Table 6.3.

Table 6.3 Variables used in calibration of the adult cross-sectional sample
Mother's age at cohort child's birth
Respondent age
Highest education level of respondent
NS-SEC of respondent
Respondent employment status
Whether respondent has a disability/limiting illness
Household employment status
Household income (incl. missing category)
Last known tenure
Urban/rural classification
SIMD 2012 quintile

The calibration adjusts for any differences due to differential non-response between Sample A and Sample B.

4.2.4 Sample efficiency of main carer interview data

Weighting affects the statistical efficiency of a sample: the more variable the weights the larger the variance of the (weighted) survey estimates. More variable weights will result in larger standard errors and wider confidence intervals, so there is less certainty over where the "true" population values lie.

The precision of weighted survey estimates is indicated by the effective sample size (neff) which measures the size of an (unweighted) simple random sample that would provide the same precision (standard error) as the weighted sample. The efficiency of the weights 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 6.4.

Table 6.4 Range of adult weights and sample efficiency

	Min	Max	Mean	N	Neff	Efficiency
Main carer longitudinal weight	0.56	5.09	1	2,815	2,200	78%
Main carer cross-sectional weight	0.54	4.92	1	3,149	2,541	81%

4.3 Weights for child interview data

4.3.1 Weighting the child interview (ACASI) data

For the second time in GUS, children in Sweep 8 were asked to fill in a short self-completion questionnaire. This was done using ACASI (Audio Computer Assisted Self Interviewing). A large proportion of children completed the questionnaire; 98% of children 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 there were a small number of households (two in total) where the child completed the interview but the adult did not (therefore adult weights were not available); otherwise we could 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 children whose parents had not completed the Sweep 8 adult interview were given a weight from the last interview completed as an entry weight to calibration.

Two sets of weights were generated:

i) a set of longitudinal weights: these are weights for children who completed the ACASI and whose parents had completed every wave of GUS up to and including Sweep 7, and

ii) a set of cross-sectional weights: these are weights for children who completed ACASI but whose parents had missed one or more waves prior to Sweep 8.

As with the adult cross-sectional weights, the choice of variables used in the calibration was dictated by the bias remaining after the appropriate Sweep 8 weights were applied. The variables used in calibration are listed in Table 6.5 below.

Table 6.5 Variables used in calibration of child interview data

Longitudinal sample	**Cross-sectional sample**
Mother's employment status	Mother's employment status
Self-reported general health of respondent	Whether respondent has a job
Whether respondent has a disability/limiting illness	Whether respondent has a disability/limiting illness
SIMD 2012 quintile	Highest education level of respondent
	Respondent ethnicity
	Whether any book/stories read in last week
	Family type

The final weights were scaled to the responding Sweep 8 ACASI sample size, so that the weighted sample size matches the unweighted sample size.

4.3.2 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 6.6.

Table 6.6 Range of weights and sample efficiency

	Min	Max	Mean	N	Neff	Efficiency
ACASI longitudinal weight	0.56	5.38	1	2,765	2,142	77%
ACASI cross-sectional weight	0.53	5.21	1	3,088	2,468	80%

4.4 Applying the weights

The cross-sectional weights should be used for any cross-sectional analysis of Sweep 8 data. All sample members that responded at Sweep 8 have a cross-sectional weight therefore the base for analysis is maximised when using the cross-sectional weights. When the base for analysis is responding adults the adult weight should be used. When the base for analysis is responding children, the child weight should be used.

The longitudinal weight should be used for any longitudinal analysis - analysis that looks at change over time at the level of individual sample members – of those taking part in Sweep 8. Sample members that have responded at every sweep of GUS have a longitudinal weight. Therefore it can be used for any longitudinal analysis of data involving Sweep 8 and any one or more of the previous Sweep(s). For example, the Sweep 8 longitudinal weight could be used for analysis of changes between Sweeps 7 and 8 or for analysis of transitions between successive sweeps starting with Sweep 1 and ending with Sweep 8.

Table 6.7 Description of weight variables in the data file

Variable name	Label
DhWTbrth	Dh Birth cohort Sweep 8 weight
DhWTbth2	Dh Birth cohort Sweep 8 weight - longitudinal
DhWTchld	Dh Child ACASI Sw8 weight
DhWTchd2	Dh Child ACASI Sw8 weight - longitudinal

5 Using the data

The GUS Sweep 8 data consists of the following SPSS file:

GUS_SW8_B.sav	3150 cases	Birth cohort 1

5.1 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**". 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 MhSDQ01 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.

Please also see Appendix B for any further issues to take into account when working with the data.

5.2 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
		h	Almost 10 years (in Primary 6)

5.3 Variable labels

In the Sweep 8 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.

5.4 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.

5.5 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.

5.6 Indicators and summary

5.6.1 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 MhHGSl(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 8.

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
Variable name	**Description**
DhHGnmad	Dh - Number of adults (16 or over) in household
DhHGnmkd	Dh - Number of children in household
DhHGrsp05	Dh Resp is childs mother? (incl. adopt./foster/step-mothers)
DhHGrsp06	Dh Resp is childs father? (incl. adopt./foster/step-fathers)
DhHGrsp01	Dh - Whether respondent is natural mother
DhHGrsp02	Dh - Whether respondent is natural father
DhHGrsp07	Dh Who is the respondent in relation to the child
DhHGnp02	Dh - Natural mother in household
DhHGnp03	Dh - Natural father in household
DhHGnp04	Dh - Respondent living with spouse/partner
DhHGrsp08	Dh Resps partner relation to the child

Table 7.2 continued	
DhMothID	Dh – Mother's ID (= Person number in household)
DhFathID	Dh – Father's ID
DhRespID	Dh – Respondent's ID
DhPartID	Dh – Respondent's partner's ID
DhHGmag5	Dh Age of natural mother at birth of cohort child (banded)
DhHGagC	Dh Study childs age at interview (months)

5.6.2 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.

5.6.3 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 (DhEqvinc) and quintiles of the scale (DhEqv5) are available in the datasets.

5.6.4 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 the Scottish Government's website: http://www.gov.scot/Topics/Statistics/About/Methodology/UrbanRuralClassification?utm_source=website&utm_medium=navigation&utm_campaign=statistics-evaluation-tools.

5.6.5 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 8 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

5.6.6 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

5.6.7 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 'ALhHBdBc' 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
Borders	Aggregated
Dumfries and Galloway	Aggregated
Fife	Identified
Forth Valley	Identified
Grampian	Identified
Greater Glasgow and Clyde	Identified
Highland	Identified

Table 7.9 continued	
Lanarkshire	Identified
Lothian	Identified
Orkney	Aggregated
Shetland	Aggregated
Tayside	Identified
Western Isles	Aggregated

5.6.8 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-olds[5]. 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.

Table 7.10	Derived variables associated with the Strengths and Difficulties Questionnaire
Variable name	**Description**
DhDsdem1	Dh SDQ: Emotional symptoms score
DhDsdco1	Dh SDQ: Conduct problems score
DhDsdhy1	Dh SDQ: Hyper-activity or inattention score
DhDsdpr1	Dh SDQ: Peer problems score
DhDsdps1	Dh SDQ: Pro-social score
DhDsdto1	Dh SDQ: Total difficulties score

Further details on the SDQ can be found at: http://www.sdqinfo.com/

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

5.6.9 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:

BMI range	**Description**
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/m 2 (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**
DhBMI	Dh BMI (reliable child weight measurements only)
DhUKbmi	Dh UK BMI national classification standards
DhINTbmi	Dh International BMI cut-offs
DhINTbmi2	Dh BMI status (ovrwt inc. obese) - international cut-offs
DhINTbmi3	Dh BMI status (non-obese vs obese) - international cut-offs
DhISDbmi	Dh ISD BMI 5 group classification
DhISDHWt	Dh Study child weight within/outwith ISD healthy range
DhISDovW	Dh Study child overweight, including obese (ISD)

5.6.10 Parental physical and mental health

At Sweep 8, health-related quality of life was measured by the Medical Outcomes Study 12-Item Short Form (SF-12). This measure was previously used at sweeps 1, 3 and 5 with BC1. It has also been used in the Scottish Health Survey and in other population surveys (for example, the Health Survey for England and the National Survey of NHS Patients). The SF-12 is a widely used self-reported generic measure of health status, yielding both a physical component (PCS) and a mental health component (MCS) summary scale score. It is tailored for use in large health surveys of general populations. Higher scores on both the physical and mental health component scales are indicative of better health-related quality of life, the indicator is based on informants' self-reports of their own physical and mental functioning and as such are subjective. This may lead to differential reporting between informants with equivalent status.

Table 7.14 Constituent and derived variables associated with the SF-12	
Variable name	**Description**
MeHpgn01	Me - How is resp health in general
MeHlmt01	Me - Resp health limits moderate activities
MeHlmt02	Me - Resp health limits climbing stairs
MeHlmt03	Me - Resp health limited accomplishments past 4 wks
MeHlmt04	Me - Resp health limited reg activities past 4 wks
MeHlmt05	Me - Resp mental health limited accomplishments past 4 wks

Table 7.14 continued	
MeHlmt06	Me - Resp mental health limited quality of accomplishments past 4 wks
MeHlmt07	Me - Resp physical pain limited normal work past 4 wks
MeHpgn02	Me - Time resp felt calm in past 4 wks
MeHpgn03	Me - Time resp felt energetic in past 4 wks
MeHpgn04	Me - Time resp felt down in past 4 wks
MeHpgn05	Me - Time resp health interfered socially in past 4 wks
DeSF12ph	De - Physical PCS - 12 Scale
DeSF12mn	De - Mental MCS - 12 Scale

5.6.11 Pragmatics subscale of the Children's Communication Checklist

Sweep 8 includes a parent report measure of children's communication which uses selected items from the 'Pragmatics' subscale of the Children's Communications Checklist (CCC) (Bishop, 1998). Items from the CCC have also previously been used on the Avon Longitudinal Study of Parents And Children (ALSPAC).

The CCC consists of nine subscales to measure children's communicative ability: Speech, Syntax, Initiation, Coherence, Conversation, Context, Rapport, Social Behaviour, and Restricted interests. Of these, Initiation, Coherence, Conversation, Context and Rapport can then be combined to form a 'pragmatics' scale (Botting, 2004).

25 items were asked of the cohort child's main carer as part of the self-completion section. Due to copyright, these items have not been included in the dataset but are available on request.

5.6.12 Parent-Child Communication (selected items from the People In My Life (PIML) scale)

The People in My Life measure is a self-report instrument designed to measure attachment to parents and peers in middle childhood. The GUS Sweep 7 child questionnaire included selected items from the Parent Attachment scale. Further information about the PIML scale can be found on the Fast Track Project website: http://fasttrackproject.org/techrept/p/pml/

Resident carers who are not either the child's biological or adoptive parents are referred to as 'resident mother figures' and 'resident father figures'. Note that responses to questions asked about a child's biological or adoptive mother and father are stored separately from questions about mother and father figures who are not the child's biological or adoptive parents. A 'parent figure' was defined

as someone who is resident with the child and who is 'a main carer to the child and is involved in their day-to-day care'.

An overview of the relevant variable names is given below, with details about individual variables provided in Table 7.15.

- **ChMum1-ChMum9**: ask about the child's relationship with their *biological or adoptive mother* (where she is resident with the child).
- **ChMumAl1-ChMumAl9**: ask about the child's relationship with a resident mother figure in cases where the child's biological or adoptive mother does not live with the child.
- **ChDad1-ChDad9**: ask about the child's relationship with their *biological or adoptive father* (where he is resident with the child).

  ChDadAl1-ChDadAl9: ask about the child's relationship with a *resident father figure* in cases where the child's biological or adoptive father does not live with the child.

Table 7.15	Selected items from People In My Life scale (child questionnaire)
Variable name	**Description**
ChMum1/ ChMumAl1	My Mum listens to what I have to say
ChMum2/ ChMumAl2	My Mum cares about me
ChMum3/ ChMumAl3	I can count on my Mum to help me when I have a problem
ChMum4/ ChMumAl4	My Mum can tell when I'm upset about something
ChMum5/ ChMumAl5	I talk to my Mum when I'm having a problem
ChMum6/ ChMumAl6	If my Mum knows something is bothering me, she asks me about it
ChMum7/ ChMumAl7	I share my thoughts and feelings with my Mum
ChMum8/ ChMumAl8	My Mum pays attention to me
ChMum9/ ChMumAl9	My Mum is proud of the things I do
ChDad1/ ChDadAl1	My Dad listens to what I have to say
ChDad2/ ChDadAl2	My Dad cares about me
ChDad3/ ChDadAl3	I can count on my Dad to help me when I have a problem
ChDad4/ ChDadAl4	My Dad can tell when I'm upset about something

Table 7.15 continued	
ChDad5/ ChDadAl5	I talk to my Dad when I'm having a problem
ChDad6/ ChDadAl6	If my Dad knows something is bothering me, he asks me about it
ChDad7/ ChDadAl7	I share my thoughts and feelings with my Dad
ChDad8/ ChDadAl8	My Dad pays attention to me

5.7 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.

5.8 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.

6 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.

7 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.

Bishop, D.V.M. (1998) Development of the Children's Communication Checklist (CCC): A method for assessing qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry*, 39(6), 879-893.

Bishop, D.V.M. (2003) *Children's Communications Checklist (CCC-2)*. London: Pearson Education, Inc.

Botting, N. (2004). Children's Communication Checklist (CCC) scores in 11-year old children with communication impairments. *International Journal of Language and Communication Disorders*, 39(2), 215 - 227. doi: 10.1080/13682820310001617001<http://dx.doi.org/10.1080/136828203100016 17001>

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

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

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

8 Related publications

To date, two Scottish Government reports have been published which utilises Sweep 8 data. The reports can be found on the Scottish Government website and links are also available from the Growing Up in Scotland website. http://www.growingupinscotland.org.uk/.

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.

McCrorie, P and Ellaway, A (2017) *Objectively measured physical activity levels of Scottish children: Analysis from a sub-sample of 10-11 year olds in the Growing Up in Scotland study*. Edinburgh: Scottish Government.

Parkes, A; Riddell, J; Wight, D and Buston, K (2017) Growing Up in Scotland: *Father-child relationships and child socio-emotional wellbeing*. Edinburgh: Scottish Government.

9 Contact details

Queries should be directed to the GUS team at ScotCen Social Research:

Line Knudsen (Senior Researcher): line.knudsen@scotcen.org.uk

Jackie Palmer (Data Manager): jackie.palmer@scotcen.org.uk

Paul Bradshaw (Project Director): paul.bradshaw@scotcen.org.uk

ScotCen Social Research

Scotiabank House,

6 South Charlotte St,

Edinburgh, EH2 4AW.

T: 0131 240 0210

10 Appendices

10.1 Appendix A

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

	B	S.E.	Wald	df	Sig.	Exp(B)
Mother's age at birth						
<20	0.00		9.34	4	0.05	1
20-24	0.03	0.22	0.02	1	0.88	1.03
25-29	0.22	0.22	1.04	1	0.31	1.25
30-34	0.13	0.22	0.35	1	0.55	1.14
5 35+	0.63	0.26	6.08	1	0.01	1.88
Highest Education level of Respondent						
Degree or equivalent	0.00		12.19	4	0.02	1
Vocational qualification below degree	-0.22	0.19	1.29	1	0.26	0.81
Higher Grade or equivalent	-0.49	0.27	3.36	1	0.07	0.61
Standard Grade or equivalent	-0.66	0.22	9.08	1	0.00	0.52
6 No Qualifications	-0.24	0.26	0.83	1	0.36	0.79
Respondent NSSEC - 6 Category						
Managerial and professional occupations	0.00		9.34	4	0.05	1
Intermediate occupations	-0.17	0.22	0.64	1	0.43	0.84
Small employers and own account workers	-0.30	0.28	1.13	1	0.29	0.74
Lower supervisory and technical occupations	-0.58	0.28	4.33	1	0.04	0.56
Semi-routine and routine occupations/ Never worked	-0.53	0.20	7.22	1	0.01	0.59
Respondent currently has job						
Not applicable	0.00		8.73	2	0.01	1
Yes	0.38	0.16	5.27	1	0.02	1.46
No	0.74	0.34	4.80	1	0.03	2.10
Household employment status						
At least one parent/carer in full-time employment	0.00		10.74	2	0.00	1
At least one parent/carer in part-time employment	-0.49	0.15	10.28	1	0.00	0.62
No parent/carer working	0.02	0.20	0.01	1	0.91	1.02
Number of accidents/injuries						
0 No	0.00					1
Yes	-0.44	0.13	10.80	1	0.00	0.65

Table A.1 continued						
	B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**
Total number of calls						
1	0.00		26.82	4	0.00	1
2	-0.63	0.18	12.00	1	0.00	0.54
3	-0.45	0.21	4.64	1	0.03	0.64
4	-0.74	0.23	10.31	1	0.00	0.48
5	-1.00	0.20	24.82	1	0.00	0.37
General health of respondent						
Excellent	0.00		8.94	4	0.06	1
Very good	0.35	0.18	3.82	1	0.05	1.42
Good	0.21	0.18	1.37	1	0.24	1.24
Fair	0.28	0.22	1.62	1	0.20	1.32
5 Poor	-0.32	0.28	1.27	1	0.26	0.73
SIMD in 15% most deprived datazones						
Not in 15% most deprived	0.00					1
In 15% most deprived	-0.31	0.15	4.47	1	0.03	0.73

Notes:
1. Response is 1 = sample A response to Sweep 8, 0 = sample A non-response..
2. Model weighted by the Sweep 7 longitudinal weight.
3. B is the estimate coefficient with standard error S.E.
4. 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.
5. 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 7 weighted by Sweep 7 (longitudinal) weight	Sweep 8 weighted by Sweep 7 (longitudinal) weight	Sweep 8 weighted by Sweep 8 (longitudinal) weight
	%	%	%
Mother's age at birth			
<20	6.9	6.1	6.8
20-24	16.9	16.0	16.9
25-29	23.5	23.6	23.4
30-34	31.9	32.3	31.9
5 35+	20.8	22.0	20.9
Highest Education level of Respondent			
Degree or equivalent	29.7	31.4	29.7
Vocational qualification below degree	40.7	40.8	40.7
Higher Grade or equivalent	6.7	6.5	6.7
Standard Grade or equivalent	14.6	13.5	14.6
6 No Qualifications	8.4	7.8	8.4
Respondent NSSEC - 6 Category			
Managerial and professional occupations	32.7	34.7	32.7
Intermediate occupations	18.2	18.6	18.2
Small employers and own account workers	7.3	7.4	7.2
Lower supervisory and technical occupations	5.4	5.2	5.4
Semi-routine and routine occupations/ Never worked	36.5	34.1	36.4
Respondent currently has job			
Not applicable	25.1	23.6	25.1
Yes	70.9	72.2	70.8
No	4.1	4.2	4.1
Household employment status			
At least one parent/carer in full-time employment	68.2	70.1	68.3
At least one parent/carer in part-time employment	19.6	18.5	19.5
No parent/carer working	12.2	11.4	12.2
Whether any accidents/injuries			
No	77.3	78.1	77.4
Yes	22.7	21.9	22.6

Table A.2 Continued

	Sweep 7 weighted by Sweep 7 (longitudinal) weight	Sweep 8 weighted by Sweep 7 (longitudinal) weight	Sweep 8 weighted by Sweep 8 (longitudinal) weight
	%	%	%
Total number of calls			
1	27.2	28.8	27.2
2	31.8	31.7	31.8
3	17.7	17.8	17.7
4	9.7	9.4	9.8
5	13.6	12.3	13.5
General health of respondent			
Excellent	17.3	17.3	17.4
Very good	37.5	38.3	37.4
Good	29.6	29.5	29.6
Fair	11.8	11.6	11.8
Poor	3.9	3.4	3.8
SIMD in 15% most deprived datazones			
Not in 15% most deprived	84.0	85.4	84.1
In 15% most deprived	16.0	14.6	15.9
Base (unweighted): 2,815			

Table A.3 Weighted distribution of key variables for samples A and B (main respondent interview data)			
	Sample A	**Combined Sweep 8 sample (A+B)**	
	Weighted by sweep 8 (longitudinal) weight	**Weighted by pre-calibration weight[1]**	**Calibrated to sample A (Sw8 cross-sectional weight)**
Mothers age at birth	%	%	%
<20	6.9	7.2	6.9
20-24	16.8	17.1	16.8
25-29	23.3	23.3	23.3
30-34	32.1	32.1	32.1
35+	20.8	20.3	20.8
Respondent age			
<30	6.2	6.6	6.2
30-34	16.5	16.9	16.5
35-39	22.9	22.9	22.9
40-43	25.6	25.0	25.6
44+	28.8	28.6	28.8
Highest Education level of Respondent			
Degree or equivalent	29.7	29.0	29.7
Vocational qualification below degree	40.9	41.2	40.9
Higher Grade or equivalent	6.3	6.1	6.3
Standard Grade or equivalent	13.8	14.3	13.8
Other/No Qualifications	9.3	9.4	9.3
Respondent NSSEC - 6 Category			
Managerial and professional occupations	34.2	34.8	34.2
Intermediate occupations	17.7	17.4	17.7
Small employers and own account workers	7.7	7.6	7.7
Lower supervisory and technical occupations	5.0	5.1	5.0
Semi-routine and routine occupations/ Never worked	35.4	35.1	35.4
Sw8 Employee or self-employed			
Not applicable	20.6	20.6	20.6
Yes	75.7	75.4	75.7
No	3.7	3.9	3.7
Respondent disability/illness limiting			
Yes	13.1	13.7	13.1
Not at all	8.4	8.3	8.4
Does not have disability	78.5	78.0	78.5
Household employment status			
At least one parent/carer in full-time employment	71.7	70.7	71.7
At least one parent/carer in part-time employment	18.2	18.5	18.2
No parent/carer working	10.1	10.7	10.1

Table A.3 continued			
	Sample A	**Combined Sweep 8 sample (A+B)**	
	Weighted by sweep 8 (longitudinal) weight	**Weighted by pre-calibration weight[1]**	**Calibrated to sample A (Sw8 cross-sectional weight)**
Household income - grouped			
<15,000	12.1	12.7	12.1
15,000-25,999	22.0	22.0	22.0
26,000-49,999	32.0	31.5	32.0
50,000+	27.1	26.7	27.1
Missing	6.8	7.1	6.8
Last known tenure			
Owner occupied	62.8	62.8	62.8
Rents from HA/council	26.8	27.0	26.8
Rents privately	10.4	10.3	10.4
Urban-rural classification			
Large urban	35.7	36.2	35.7
Other urban	32.6	32.8	32.6
Small, accessible towns	10.2	9.8	10.2
Small remote towns	3.0	2.9	3.0
Accessible rural	12.6	12.6	12.6
Remote rural	6.0	5.7	6.0
Overall SIMD 2012 Quintile			
1 - Least deprived	20.5	20.8	20.5
2	17.9	18.4	17.9
3	19.8	19.8	19.8
4	21.1	20.8	21.1
5 - Most deprived	20.6	20.1	20.6
Base (unweighted): 3,671			

10.2 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.

Self-complete section: 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.

Partial completes and child only interviews: Three cases had a partial interview (code 210 or 211 at variable MhOutcome), 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, two cases had a child interview, but no adult interview (code 212 at MhOutcome).