Growing Up in Scotland Sweep 9: 2017-18

User Guide

Contents

1 Survey details

1.1 Study aims and objectives

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."

1.2 Sweep 9 overview

At Sweep 9 data collection included five main elements:

1. A face-to-face CAPI (Computer Assisted Personal Interview) interview with the cohort child's main carer. This includes a self-completion element.

2. A self-complete CASI (Computer Assisted Self-Complete Interview) interview with the cohort child

3. Height and weight measurement of the cohort child

4. Cognitive assessments of the cohort child

5. A self-complete PAPI (Pen and Paper Interview) questionnaire with any resident partner of the main adult respondent

This user guide accompanies data collected from the main carer and the cohort child. The data collected via paper questionnaires with resident partners will be deposited separately.

1.3 Study design

GUS 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). In 2018, an additional 502 families were recruited to the study. These families took part in interviews alongside families in the original birth cohort. Further details are provided in section 1.4.2.

A second birth cohort of 6127 children aged around 10 months at the first interview was recruited in 2011.

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 original 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 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	10-11	12-13
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) 8.5 (BC1) 2015/16					BC2						BC1 w-c*	
9 2016/17												BC1

*'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 12-year-old children were living in single parent families in 2017/18?

- 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 2011?

- 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 aged 12?

1.4 Sample design[1]

1.4.1 BC1 Main sample (the original birth cohort)

The original or 'main' BC1 sample was recruited at Sweep 1.

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 were born between 1st June 2004 and 31st May 2005. 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

---

[1] This section focusses on sample design for the sample interviewed at Sweep 9 – i.e. BC1. Information about the sample design for BC2 is provided in the user guide accompanying the BC2 Sweep 1 dataset which is available through the UK Data Service website.

[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.

considered 'sensitive' and children that had been sampled for research by the DWP in the last 3 years.

1.4.2 BC1 Boost sample (refreshment sample recruited in 2018)

Whilst the overall levels of attrition seen in GUS are typical for a cohort/longitudinal study, the effects of attrition are spread unevenly over the sample, with some sub-groups affected more than others. This limits the scope of sub-group analysis as it reduces the number of available cases in specific sub-groups. Analysis of the achieved sample from Birth Cohort 1 after Sweep 8 revealed that two groups in particular had become under-represented since the beginning of the study: children born to mothers aged 16-24 at time of birth and children living in the 15% most deprived areas (according to the Scottish Index of Multiple Deprivation). Children in both groups were affected the most, with less than 100 cases available for analysis children of young mothers in deprived areas, as shown in table 1.3 below.

Table 1.3. Attrition rates for key sub-groups in the GUS BC1 sample			
	Sample size at sw1	Sample size at sw8	% interviewed at sw8
Group 1: Age 16-24 in most deprived 15% data zones	400	95	24%
Group 2: Age 25+ in most deprived 15% data zones	471	226	48%
Group 3: Age 16-24 in remaining 85% data zones	828	380	46%
Group 4: Age 25+ in remaining 85% data zones	3518	2450	70%
All families	5217	3151	60%

To resolve this under-representation, it was agreed that a boost sample for BC1 would be selected and issued as part of the phase 2 fieldwork for sweep 9. The boost sample specifically targeted those in the under-represented groups (i.e. groups 1 to 3 in the table above).

Sample design and selection

Sampling frame (children)

The additional sample was drawn from Child Benefit records held by HMRC. Child Benefit records were the original sampling frame from which the sweep 1 sample was drawn. They include information about parental demographics, such as age and sex, and geographical location.

Child Benefit is no longer a universal benefit but those ineligible for the boost sample were parents on higher incomes who are already well represented in GUS. Only a very small proportion of the GUS target groups was excluded from the Child Benefit database. Child Benefit continues to have an exceptionally high take up rate – around 96% of those eligible claimed the benefit - meaning that amongst those who are in our target groups, the records still offer very good coverage; that is, they capture almost all families in our population of interest.

Children born between 1st June 2004 and 31st May 2005 and in at least one of the two key sub-groups (i.e. mother aged 16-24 at birth; living in 15% most deprived data zones in Scotland) were eligible for the boost. Following liaison with HMRC, eligible children from across Scotland were extracted from Child Benefit records. An anonymised version of the file was sent to NatCen with unique identifiers for child and household.

Target numbers

The aim was to sample 1,500 children from those eligible, broken down as follows amongst the three groups of interest:

Table 1.4. Target numbers by group	
Group of interest	**Number to sample**
Group 1: Age 16-24 in most deprived 15% data zones	700
Group 2: Age 25+ in most deprived 15% data zones	550
Group 3: Age 16-24 in remaining 85% data zones	250

Sampling frame (PSUs)

An un-clustered sample of eligible children would be highly dispersed geographically and impractical for face-to-face interviewing. Therefore, eligible children were clustered into Primary Sampling Units (PSUs) which were based on data zones.

Eligible children were first allocated to a data zone based on their postcode. Data zones with eligible children were then grouped into PSUs comprising several geographically contiguous data zones. To control the numbers of children sampled in the three groups of interest, the clusters were designed to include either deprived[3] or non-deprived data zones (but not both) and to contain an average of approximately 16 children.

---

[3] 15% most deprived data zones

PSUs constructed from deprived data zones were sampled separately from those containing non-deprived data zones. Thus, children in groups 1 and 2 were sampled separately from children in group 3.

Sampling children in group 3

A total of 16 PSUs was selected, with equal probability, from non-deprived data zones (children from group 3 only) using systematic random sampling, with Local Authority (LA) (grouped into four broad categories based on NUTS2) and the 2016 Scottish Index for Multiple Deprivation (SIMD) as implicit stratification variables. No further sampling of children within these PSUs took place. This resulted in selection of 247 children from group 3.

Sampling children in groups 1 & 2

Sampling of PSUs containing deprived data zones (children in groups 1 and 2) was slightly more complex. The ratio of children in groups 1 and 2 was calculated (for each PSU) and PSUs were divided into two groups depending on this ratio[4]. In other words, a group of PSUs with a relatively high proportion of children with young mothers was created and separated from PSUs containing lower proportions of children with young mothers. This was done to achieve the desired number of children in each category of the key-subgroups defined above, whilst minimising the design effect due to differential sampling fractions.

A total of 36 PSUs was selected from the set of PSUs with relatively high proportions of children with young mothers, whilst 60 PSUs were selected from those with lower proportions of these children. Systematic random sampling was used, with LA and the total number of children in each PSU as implicit stratification variables.

The number of PSUs sampled in each of the three groups, along with the total number of children in these PSUs, is summarised in Table 1.5 below.

Table 1.5. Number of PSUs and children sampled by group		
PSU group	No. of sampled PSUs	No. of children
Group 1: Age 16-24 in most deprived 15% data zones	16	709
Group 2: Age 25+ in most deprived 15% data zones	36	793
Group 3: Age 16-24 in remaining 85% data zones	60	247

Sub-sampling children in group 2

---

[4] The threshold used was a ratio equal to 0.8 of children with mothers aged 16-24 (at birth of child) to children with mothers aged 25 or over.

As can be seen from table 1.6, the number of children in group 2 was somewhat higher than the target number (550). This was unavoidable given the overall ratio of eligible children in groups 1 and 2. Therefore, a random sub-sample of 550 children in group 2 was taken across the deprived PSUs (using systematic sampling). Following this, one child was selected at random from the small number of households with two children.

The result of these two stages of sub-sampling was to reduce the numbers in groups 1 and 2 to 699 and 550; respectively.

Final steps

The final sample of children was sent to HMRC who returned the records with full names and addresses attached. The final numbers of children, following removal of duplicates, are shown in table 1.6 (below).

Table 1.6. Final numbers of children by group	
PSU group	No of children
Group 1: Age 16-24 in most deprived 15% data zones	664
Group 2: Age 25+ in most deprived 15% data zones	543
Group 3: Age 16-24 in remaining 85% data zones	236

1.5 Developing and piloting

Policy priorities and key topics of interest for the Sweep 9 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 the study's Questionnaire Advisory Group and policy teams across the Scottish Government.

Cognitive testing of selected items in the child questionnaire was carried out in August 2016. A full CAPI/CASI instrument, with both adult and child questionnaires, was piloted in October/November 2016.

1.6 Timing of fieldwork

In sweeps 1-7, 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').

Ahead of Sweep 8 there was interest in interviewing families according to the child's school year. Therefore, from 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 Sweeps 8 and 9 than at previous sweeps. Conversely, at Sweeps 8 and 9, almost all children were in the same school year at the time of interview (i.e. at sweep 9 most children were in their second term of their first year at secondary school - Secondary 1/S1).

Because of how children were initially sampled, the children in BC1 span two different school years. Therefore, Sweep 9 fieldwork was split into two phases:

Phase 1 fieldwork took place between January and July 2017. 2718 cases were issued for Phase 1 fieldwork (55% of the total sample issued at Sweep 9); 2219 cases were achieved as part of Phase 1 fieldwork (65% of the total number of cases achieved at Sweep 9).

Phase 2 fieldwork took place between January and July 2018. 2197 cases were issued for Phase 2 fieldwork (45% of the total sample issued at Sweep 9); 1200 cases were achieved as part of Phase 2 fieldwork (35% of the total number of cases achieved at Sweep 9).

1.7   Response rates

Table 1.7 shows historical response for BC1. It indicates that a total of 2917 interviews with the original sample were achieved at Sweep 9, representing 56% of cases achieved at Sweep 1.

Table 1.7 BC1 historical response		
	Cases achieved	**% of Sweep 1 cases**
Sweep 1	5217	-
Sweep 2	4512	86%
Sweep 3	4193	80%
Sweep 4	3994	77%
Sweep 5	3833	73%
Sweep 6	3657	70%
Sweep 7	3456	66%
Sweep 8	3150	60%
Sweep 9 (Main sample only)	2917	56%

Details of the number of cases issued and achieved at Sweep 9 and the response rates are presented in Table 1.8.

Table 1.8 Sweep 9 response				
	Total cross-sectional sample (Main and Boost samples)		**Longitudinal sample** (Main sample only)	
	No. of cases	% of issued in-scope	No. of cases	% of issued in-scope
Total issued to field	4915	-	3558	-
In-scope issued to field*	4892	-	3540	-
Total achieved	3419	70%	2917	82%
Main carer interviews achieved	3418	70%	2917	82%
Child interviews achieved	3289	67%	2834	80%
Child cognitive assessments achieved	3275	67%	2830	80%
Child height and weight measurements achieved	3220	66%	2794	79%
Partner interviews** achieved	2077	77%(of eligible***)	1923	80%(of eligible***)

*'In scope': cases with confirmed or unknown eligibility (excl. cases where family had moved out of Scotland or where cohort child has died).

**Data from the resident partner interviews (PAPI) are deposited separately from the CAPI/CASI data.

*** % of productive cases where a partner was resident in the household (Main and Boost sample: n=2687 and Main sample only: n=2393).

1.8 Length of household interview

Overall, the average interview (including adult and child interviews and height and weight measurements) lasted around 82 minutes. The median length was 79 minutes.

2 Sweep 9 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 9, this means the same respondent as Sweep 8 (or Sweep 7 / 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 (84% of interviews were with the previous respondent) and this was usually the child's mother (95% of interviews were with the child's mother).

2.2 Child interview

As noted above, the cohort children were interviewed directly for the third time at Sweep 9. The children participated by answering questions themselves on the interviewer's laptop using a CASI) approach. In this approach the questions and response options are displayed on screen. Informed consent was gained from both the main carer and from the child[5].

The child questionnaire consisted of the following sections:

- A short interviewer-led section including an introduction, consent procedures and practice questions.
- A CASI section with questions on topics including school, friends and peer relationships, health and wellbeing, online activities and

---

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

experiences, relationship with resident and non-resident parents/carers, smoking and alcohol, and anti-social behaviour.

2.3 Cognitive assessments

Cognitive assessments were previously carried out with the children in BC1 at sweeps 3, 5 and 8. Cognitive assessments were also carried out at Sweep 9. At Sweep 9, children were assessed using the 'Listening Comprehension' subtest of the Weschler Individual Achievement Tests, 2nd Edition (WIAT-II). This is the same assessment which was administered at Sweep 8.

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.[6] At GUS sweep 9, all children started at the same point (note that at sweep 9 this was not the first item in each sub-test) however, some children may have subsequently been asked earlier items depending on their progress through the assessment. Where children were not asked those earlier

---

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

items, they were scored positively. Understanding which set of items were administered to the child is important when analysing the results.

The following scores are available in the dataset:

- ***Receptive Vocabulary Adjusted Raw Score:*** A count of all the items on Receptive Vocabulary the child answered correctly (including where early items were automatically scored).

- ***Sentence Comprehension Adjusted Raw Score:*** A count of all the items on Sentence Comprehension the child answered correctly (including where early items were automatically scored).

- ***Expressive Vocabulary Adjusted Raw Score:*** A count of all the items on Expressive Vocabulary the child answered correctly (including where early items were automatically scored).

- ***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 adjusted 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, 7, 8 and were also included at Sweep 9.

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).

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 (WiXhei14 and WiXwei19). 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 (WiXhei01 and WiXwei01), with the reason for refusal at WiXhei021-8 or WiXwei021-7.

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 5.6.20.

3 Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted. The GUS Sweep 9 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 9

Five weights were developed for Sweep 9 of BC1. Two weights were generated for analysis of information collected during the main interview with the main carer, two weights for analysis of data collected from the child using ACASI, and one additional weight for the PAPI responses from the main carer's resident partner questionnaire. Sweep 9 was the third sweep to collect information directly from the study child (the first one was Sweep 7) and the first to gather information from the main carer's partner.

The five weights were:

- A cross-sectional weight for adults that should be used for any cross-sectional analysis of data collected in the Sweep 9 main carer interview. All main carers that responded at Sweep 9 have a cross-sectional adult weight.

- A longitudinal weight for analysis of main carers that have responded at every sweep of GUS including SW9[7].

- A cross-sectional weight that should be used for any cross-sectional analysis of the Sweep 9 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 8.

- A cross-sectional weight for analysis of the Sweep 9 PAPI data (i.e. data collected from the main carer's resident partner). All children that completed the PAPI interview at this sweep have a cross-sectional child weight.

- The Sweep 9 interview follows up all main carers who responded at the Sweep 8 interview and gave NatCen permission to be re-contacted. In addition, mothers who had refused the Sweep 8 interview but had responded at any previous Sweep were contacted if they had given a 'sweep only' refusal at Sweep 8.

- Sweep 9 is also the first BC1 Sweep to include a Boost Sample. The boost sample was drawn from Child Benefit records held by HMRC and was added to improve the under-representation of children living in deprived areas and/or born to young mothers aged 16-24.

---

[7] This excludes the web-CATI Sweep 8.5.

4.2 Weights for main carer interview data

4.2.1 Main carer sample

The Sweep 9 sample of adult respondents can be split into three groups. For the purposes of describing the weighting these have been named Sample A, Sample B and Boost Sample. These 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 2-8.

- Boost Sample – adults from the refreshment sample added at Sweep 9.

The three samples were treated separately during the weighting. This is because respondents in the Sample B and the Boost Sample are likely to have different response behaviour to those in Sample A, as demonstrated by the difference in their response rates.

- There were 764 individuals in Sample B, 389 (51%) of which responded at Sweep 9.

- The response rate for Sample A (n=2,793) was much higher (91%) with 2,528 responding.

The issued and responding sample sizes for the three groups are given in Table 4.1.

Table 4.1 Response rates for the two groups of main interview respondents			
	Issued	**Responding**	**Response rate**
Sample A	2,793	2,528	91%
Sample B	764	389	51%
Boost	1,357	501	37%
Combined (A+B)	4,914	3,418	70%

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 previous sweep of GUS[8]) received a longitudinal weight. This weight is described in more detail in Section 6.2.1.

All Sweep 9 respondents will have a cross-sectional weight (Sample A + B + Boost). These are described in more detail in Section 4.2.2.

---

[8] Excluding the web-CATI Sweep 8.5.

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 9 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 8. 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 9 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 8).

Table 4.2 Variables used in adult non-response weighting (longitudinal sample)
Number of persons 16 and under
Mother's age at cohort child's birth
Highest Education level of Respondent
Ethnicity of Respondent
Household employment status (at least one carer in full-time, part time employment)
General child health
Number of books/stories read to the child in previous week
Total number of visits made to the address
Respondent's current tenure

The final Sweep 9 weight was calculated as the product of the non-response weight and the Sweep 8 interview weight. The final weights were scaled to the responding Sweep 9 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 9 (Sample A + Sample B + the Boost sample) and should be used for any cross-sectional analysis of Sweep 9 data.

Calibration weighting was applied to the combined sample 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, make the survey estimates match the population estimates which in this instance were calculated from Sample A, weighted by the longitudinal weight. Since the longitudinal weight corrects for non-response bias at each stage of GUS, the weighted

Sample A estimates are the best estimates available for children from the cohort from which Sweep 1 was sampled who remain in Scotland.

The choice of the variables used in the calibration was decided upon by comparing (*Sample A weighted by the Sweep 9 longitudinal weights)* with the combination of {*Sample B weighted by the cross-sectional weight from the last completed sweep + the Boost sample weighted by the interim non-response weight[9]}*. This was done using a logistic regression model where the dependent variable was equal to 1 if the case was a member of sample A and 0 if not.

A variable representing the four combinations of 'household in top 15% most deprived areas' and 'Young mother at birth of child' was included in the model and the calibration itself. These two variables were the ones used when sampling children for the Boost Sample. Including them in the calibration weighting forced the (weighted) cross-sectional sample to have the same proportions of children in these four categories as our best population estimate coming from (weighted) Sample A.

The variables used in the calibration weighting are listed in Table 4.3 below.

Table 4.3 Variables used in calibration of the adult cross-sectional sample
Household is in top 15% most deprived areas (SIMD) * Young mother at birth of child (i.e. 24 years or younger) [10]
Family type (lone parent, couple family)
Household income (incl. missing category)
Respondent age
NS-SEC of respondent
Respondent's self-reported general health
Ethnicity of Respondent
Mother's age at cohort child's birth
Respondent's current tenure
Urban/Rural Classification
SIMD 2016 Access domain quintile
SIMD 2016 Housing domain quintile
SIMD 2016 Employment domain quintile
SIMD 2016 Index quintile

The variables included in the final model are summarised in Table 4.4 below.

Table 4.4 Variables used in Boost Sample non-response weighting
Household is in top 15% most deprived areas (SIMD) * Young mother at birth of child (i.e. 24 years or younger)
SIMD 2016 Access domain quintile
SIMD 2016 Housing domain quintile

[9] See section 2.3 for a description of the boost sample non-response weighting.
[10] Four category combination of these two measures

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 4.5.

Table 4.5 Range of adult weights and sample efficiency						
	Min	**Max**	**Mean**	**N**	**Neff**	**Efficiency**
Main carer longitudinal weight	0.54	7.13	1	2,528	1,883	74%
Main carer cross-sectional weight	0.29	5.86	1	3,418	2,833	83%

4.3 Weights for child interview data

4.3.1 Weighting the child interview (ACASI) data

For the third time in GUS, children in Sweep 9 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; 96% of children whose main carer had completed the main CASI interview.

Calibration methods were used to generate non-response weights for the children.

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 9, 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 9.

Children in the Boost Sample received only the cross-sectional weights as SW9 was the first wave to which they were invited and thus they couldn't have completed any of the previous Sweeps.

One child whose main carer had not completed the Sweep 9 adult interview was given a weight from the last interview completed as an entry weight to

calibration for the cross-sectional sample. The child did not receive an ACASI longitudinal weight as the main carers had not completed every previous GUS wave up to SW9.

As with the adult cross-sectional weights, the choice of variables used in the calibration was dictated by the small bias remaining after the appropriate (longitudinal or cross-sectional) Sweep 9 weights were applied. The variables used in calibration are listed in Table 4.6.

Table 4.6 Variables used in calibration of child interview data	
Family type (lone parent, couple family)	Household is in top 15% most deprived areas (SIMD)
Main carer's current tenure	Young mother at birth of child (i.e. 24 years or younger)
Urban/Rural Classification	Urban/Rural Classification
Highest Education level of main carer	Highest Education level of main carer
Self-reported general child health	Mothers employment status
Main carer's self-reported general health	Self-reported general child health
Frequency main carer sits to eat with child	Main carer's self-reported general health
	Whether main carer has a disability/limiting illness
	Whether main carer currently has a job

The final weights were scaled to the responding Sweep 9 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 4.7.

Table 4.7 Range of weights and sample efficiency						
	Min	**Max**	**Mean**	**N**	**Neff**	**Efficiency**
ACASI longitudinal weight	0.53	7.25	1	2,469	1,822	74%
ACASI cross-sectional weight	0.25	4.27	1	3,289	2,724	83%

4.4 Weighting the partner interview (PAPI) data

For the second time in GUS, partner interviews (PAPI questionnaires) were attempted in any household with live-in partners. Partner interviews were first carried out at sweep 2 of the survey. The chances are that some partners have changed over time. Therefore, only cross-sectional weights have been computed for sweep 9 data which align the profile of the achieved sample of partners with the profile of all existing partners in the responding households.

77% of partners completed the questionnaire. A bivariate analysis suggested that the responding sample is systematically different from those that did not respond. Non-response behaviour was modelled using logistic regression. This is a method of analysing the relationship between an outcome variable (in this case response to the sweep 9 interview) using a set of predictor variables. The model takes account of the relationship of the predictor variables to the outcome and the relationships of the predictor variables to each other. Weighting the model by the cross-sectional weights for main carer interview data allows to identify bias remaining only due to non-response of the partners.

The variables identified as significantly predicting the non-response behaviour are listed in Table 4.8.

Table 4.8 Description of weight variables in the data file
Household is in top 15% most deprived areas (SIMD) * Young mother at birth of child (i.e. 24 years or younger) [11]
Household income (incl. missing category)
Respondent age (main carer)
NS-SEC of respondent (main carer)
Respondent's self-reported general health (main carer)
SIMD 2016 Housing domain quintile
Urban/Rural Classification
Partner's relation to the child
Partner's highest education level
Quantiles of OECD-Modified McClements household score for equivalent income
Partner's ethnicity

The model generated a predicted probability for each respondent. This is the probability the respondent would take part in the sweep 2 interview, given their characteristics, and those of the household, collected at sweep 1. Respondents with characteristics associated with non-response (such as being a private tenant) are under-represented in the final sweep 2 sample and will thus receive

---

[11] Four category combination of these two measures

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.

4.5 Applying the weights

For each sample, the cross-sectional weights should be used for any cross-sectional analysis, i.e. any analysis of Sweep 9 data only. All sample members, including those from the Boost Sample, that responded at Sweep 9 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 previous sweep of GUS have a longitudinal weight.

Table 4.9 Description of weight variables in the data file	
Variable name	**Label**
DiWTbrth	Dh Birth cohort Sweep 9 weight
DiWTbth2	Dh Birth cohort Sweep 9 weight - longitudinal
DiWTchld	Dh Child ACASI Sw9 weight
DiWTchd2	Dh Child ACASI Sw9 weight - longitudinal

ScotCen Social Research | Sweep 9: 2017-18

5   Using the data

The GUS Sweep 9 data consists of the following SPSS file:

GUS_SW9_B.sav	3419 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 MiSDQ01 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 A 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 5.1

Table 5.1	GUS variable naming conventions – BC1		
Character no.:			
1		2	
Source of data		**Sweep/Sweep**	
Non-sequential Capitals: D,M, P, C		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	Around10-11 years (in Primary 6)
		i	Around 12-13 years (in Secondary 1)

5.3   Variable labels

In the Sweep 9 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 variables

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 MiHGSI(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 9.

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 5.2. 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 5.2	Key household derived variables
Variable name	**Description**
DiHGnmad	Di Number of adults (16 or over) in household
DiHGnmkd	Di Number of children in household
DiHGrsp05	Di Resp is childs mother? (incl. adopt./foster/step-mothers)
DiHGrsp06	Di Resp is childs father? (incl. adopt./foster/step-fathers)
DiHGrsp01	Di Is respondent natural mother
DiHGrsp02	Di Whether respondent is natural father
DiHGrsp07	Di Who is the respondent in relation to the child
DiHGnp02	Di Is Natural mother in household
DiHGnp03	Di Natural father in household
DiHGnp04	Di If Respondent is living with spouse/partner

DiHGrsp08	Di Resps partner relation to the child
DiMothID	Di Mother's ID (= Person number in household)
DiFathID	Di Father's ID
DiRespID	Di Respondent's ID
DiPartID	Di Respondent's partner's ID
DiHGmag5	Di Age of natural mother at birth of cohort child (banded)
DiHGagC	Di 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 5.3 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.

GUS collects a banded version of total net household income from all sources in the main CAPI interview. The midpoint of the band is used to calculate equivalised income. This midpoint income value 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 (DiEqvinc) and quintiles of the scale based on within sample distribution (DiEqv5) are available in the datasets[12].

5.6.4 Area-level variables

Scottish Government Urban/Rural Classification

The dataset includes a binary measure of urban/rural location (ALirural). This is based on the Scottish Government's two-fold urban rural classification which is itself derived from the more detailed six-fold classification shown in Table 5.4.

The Scottish Government Urban Rural Classification was first released in 2000 and is consistent with the Government's core definition of rurality which defines

---

[12] Note previous user guides suggested this variable referred to UK wide income distribution using data from the Family Resources Survey. This is not the case for this sweep nor any previous sweep. Income distribution is considered only amongst the GUS sample.

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.

Table 5.4	Scottish Government Six-fold and Two-fold Urban Rural Classifications	
Classification	**Description – six-fold**	**Description – two-fold**
1. Large Urban Areas	Settlements of over 125,000 people	1. Urban
2. Other Urban Areas	Settlements of 10,000 to 125,000 people	1. Urban
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	1. Urban
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	1. Urban
5. Accessible Rural	Settlements of less than 3,000 people and within 30 minutes' drive of a settlement of 10,000 or more	2. Rural
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	2. Rural

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.

A detailed urban/rural variable with all six categories outlined above is available on request under UKDS Secure Licence.

Scottish Index of Multiple Deprivation

The Scottish Index of Multiple Deprivation (SIMD) 2016 identifies small area concentrations of multiple deprivation across Scotland. It is based on 38 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 2016 is presented at data zone level, enabling small pockets of deprivation to be identified. The data zones, which have a median population size of 753, are ranked from most deprived (1) to least deprived (6976) 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 9 datasets is based on the 2016 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

Further area-level variables (available on request)

Further geographical measures have been derived and are available on request through UKDS Secure Licence arrangements.

Data Zones

The data zone is the key small-area statistical geography in Scotland. SNS has introduced, for the first time, a common, stable and consistent, small-area geography called data zones. The data-zone geography covers the whole of Scotland and nests within local authority boundaries. Data zones are groups of 2001 Census output areas and have populations of between 500 and 1,000 household residents. Where possible, they have been made to respect physical boundaries and natural communities. They have a regular shape and, as far as possible, contain households with similar social characteristics[13].

Intermediate Geography

Not all statistics are suitable for release at the data-zone level because of the sensitive nature of the statistics, or for reasons of reliability, and it was apparent that a statistical geography between data zone and local authority was required. The intermediate zones are aggregations of data zones within local authorities and contain between 2,500 and 6,000 people.[14]

---

[13] Further information on data zones is available from the Scottish Government Scottish Neighbourhood Statistics Guide: https://www2.gov.scot/Publications/2005/02/20697/52626
[14] Further information on intermediate geography is available from the Scottish Government Scottish Neighbourhood Statistics Guide: https://www2.gov.scot/Publications/2005/02/20697/52626

5.6.5 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[15]. 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.

At sweep 9, the full list of SDQ items were asked of the child's main carer as part of the self-completion section, while the children themselves were asked questions which form the child-report hyperactivity and inattention subscale.

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 5.5 with syntax illustrated in the derived variables documentation.

Table 5.5	Derived variables associated with the Strengths and Difficulties Questionnaire (from main carer questionnaire only)
Variable name	**Description**
DiDsdem1	Di SDQ: Emotional symptoms score
DiDsdco1	Di SDQ: Conduct problems score
DiDsdhy1	Di SDQ: Hyper-activity or inattention score
DiDsdpr1	Di SDQ: Peer problems score
DiDsdps1	Di SDQ: Pro-social score
DiDsdto1	Di SDQ: Total difficulties score

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

5.6.6 Child mental wellbeing: selected items from the Students' Life Satisfaction Scale (Huebner, 1991)

Life satisfaction is measured through the use of selected items from the Students' Life Satisfaction Scale (Huebner, 1991). These items were previously asked in the sweep 7 child questionnaire. Relevant variables are listed in Table 5.6.

---

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

Table 5.6	Selected items from the Students' Life Satisfaction Scale (child questionnaire)
Variable name	**Description**
CiWew	Ci Do you feel that your life is going well?
CiWed	Ci Do you wish your life was different?
CiWer	Ci Do you feel that your life is just right?
CiWea	Ci Do you feel you have what you want in life?
CiWeg	Ci Do you feel you have a good life?

5.6.7 Child mental wellbeing: selected items from Kidscreen Health-Related Quality of Life scale

Mental wellbeing was also measured through use of selected items form the Kidscreen Health-Related Quality of Life scale (Ravens-Sieberer, U. et al., 2005; The Kidscreen Group, 2006). Relevant items are listed in Table 5.7.

Table 5.7	Selected items from the Kidscreen Health-Related Quality of Life scale (child questionnaire)
Variable name	**Description**
CiWw	Ci Have you felt fit and well?
CiWe	Ci Have you felt full of energy?
CiWs	Ci Have you felt sad?
CiWl	Ci Have you felt lonely?
CiWt	Ci Have you had enough time for yourself?
CiWp	Ci Have your parent(s) treated you fairly?
CiWf	Ci Have you had fun with your friends?
CiWc	Ci Have you got on well at school?
CiWa	Ci Have you been able to pay attention?

5.6.8 Child health behaviours: alcohol and smoking (selected items from HBSC)

As part of their self-completion questionnaire, children were asked a number of questions about alcohol and smoking. Questions were adapted from the Health Behaviour in School Aged Children Survey (HBSC) and are listed in Table 5.8.

Table 5.8	Selected items from HBSC on child's health behaviours - alcohol and smoking (child questionnaire)

Variable name	Description
CiSm	Ci Whether ever tried a cigarette
CiBSn	Ci How often smokes now
CiBSw	Ci How old when you first smoked a whole cigarette
CiBSe	Ci Whether ever tried e-cigarette or vaping device
CiBSa	Ci Whether ever had alcoholic drink
CiBSd	Ci How old when first had an alcoholic drink
CiBAl	Ci How often drank alcohol in the last 30 days
CiBDr	Ci Whether ever been drunk

5.6.9 Anti-social behaviour (child and main carer)

At sweep 9, for the first time on GUS, the child and the main carer were asked a range of questions about their engagement in anti-social behaviours. Questions were also asked of resident partners who took part in the paper self-completion questionnaire (see separate dataset and documentation).

The questions asked at sweep 9 are adaptations of questions previously asked as part of sweep 3 of the Edinburgh Study of Youth Transitions and Crime (Smith, 2004).

For each of the behaviours listed in Table 5.9 below, children were asked whether they had *ever* engaged in a particular form of behaviour and, if so, how many times in the last year. Main carers were asked if they had *ever* engaged in a particular form of behaviour and, if so, how old they were when they last did this.

Table 5.9	Items adapted from the Edinburgh Study of Youth Transitions (child and main carer questionnaires)
Relevant variable names	**Description**
CiASBs/MiASBs	Ci & Mi Ever taken something from a shop or a store without paying for it
CiASBsy MiASBsa	Ci How many times taken something from a shop/store in the last year Mi How old when last took something from a shop/store without paying for it
CiASBr/MiASBr	Ci & Mi Ever been rowdy or rude in a public place
CiASBry MiASBra	Ci How many times been rowdy/rude in last year Mi How old when last rowdy/rude

CiASBm	Ci Ever stolen money or other things that someone else left lying somewhere
CiASBmy	Ci How many times stolen money or other things in the last year
CiASBk/MiASBw	Ci & Mi Ever carried a knife or weapon
CiASBky MiASBwa	Ci How many times carried a knife/weapon in the last year Mi How old when last carried a knife/weapon
CiASBp/MiASBd	Ci & Mi Ever deliberately damaged or destroyed property
CiASBpy MiASBda	Ci How many times deliberately damaged/destroyed property in the last year Mi How old when last deliberately damaged/destroyed property
CiASBb/MiASBb	Ci & Mi Ever broken into a locked place to steal something
CiASBby MiASBba	Ci How many times broken into a locked place to steal something in the last year Mi How old when last broke into a locked place to steal something
CiASBg CiASBgy	Ci Ever written things or sprayed paint on property Ci How many times written things or sprayed paint on property in the last year
CiASBw	Ci Ever used force, threats or a weapon to get money or something else from somebody
CiASBwy	Ci How many times used force, threats or a weapon etc. in last year
CiASBh/MiASBa	Ci Ever hit, kicked or punched someone with the intention of hurting /injuring them Mi Ever assaulted someone with the intention of hurting/injuring them
CiASBhy MiASBaa	Ci How many times hit, kicked or punched someone in the last year Mi How old when last assaulted someone with intention of hurting/injuring them

5.6.10 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 8 child

questionnaire also 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'.

In addition to questions about resident parents/parent figures, at sweep 9 the cohort children were also asked about their relationship with any parents living elsewhere.

An overview of the relevant variable names is given below, with details about individual variables provided in Table 5.10.

- **CiMum1-CiMum9**: ask about the child's relationship with their *biological or adoptive mother* (where she is resident with the child).

- **CiMumAI1-CiMumAI9**: 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.

- **CiDad1-CiDad9**: ask about the child's relationship with their *biological or adoptive father* (where he is resident with the child).

- **CiDadAI1-CiDadAI9**: 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.

- **CiNRMum1-CiNRMum9:** ask about the child's relationship with their *mother living elsewhere* (this can be a biological or adoptive parent – see main carer questionnaire documentation for details (NRPck)).

- **CiNRDad1-CiNRDad9:** ask about the child's relationship with their *father living elsewhere* (this can be a biological or adoptive parent – see main carer questionnaire documentation for details (NRPck)).

Table 5.10	Selected items from People In My Life scale (child questionnaire)
Variable name	**Description**
CiMum1/ CiMumAI1/ CiNRMum1	Ci My Mum listens to what I have to say
CiMum3/ CiMumAI3/ CiNRMum3	Ci I can count on my Mum to help me when I have a problem
CiMum5/ CiMumAI5/ CiNRMum5	Ci I talk to my Mum when I'm having a problem
CiMum6/ CiMumAI6/ CiNRMum6	Ci If my Mum knows something is bothering me, she asks me about it

CiMum7/ CiMumAl7/ CiNRMum7	Ci I share my thoughts and feelings with my Mum
CiMum8/ CiMumAl8/ CiNRMum8	Ci My Mum pays attention to me
CiDad1/ CiDadAl1/ CiNRDad1	Ci My Dad listens to what I have to say
CiDad3/ CiDadAl3/ CiNRDad3	Ci I can count on my Dad to help me when I have a problem
CiDad5/ CiDadAl5/ CiNRDad5	Ci I talk to my Dad when I'm having a problem
CiDad6/ CiDadAl6/ CiNRDad6	Ci If my Dad knows something is bothering me, he asks me about it
CiDad7/ CiDadAl7/ CiNRDad7	Ci I share my thoughts and feelings with my Dad
CiDad8/ CiDadAl8/ CiNRDad8	Ci My Dad pays attention to me

5.6.11 Parenting and parent-child relationship: arguments and disagreements

As part of the self-completion module, the cohort child's main carer was asked several questions about arguments and disagreements between them and the cohort child. These questions were adapted from questions previously asked as part of wave 6 of The Longitudinal Study of Australian Children (Growing Up in Australia) *(Department of Social Services, 2018)*.

Table 5.11	Selected items adapted from Growing Up in Australia: Parent-child arguments and disagreements (main carer questionnaire)
Variable name	**Description**
MiPDis1	Mi My child and I get on each other's nerves
MiPDis2	Mi My child and I shout at each other
MiPDis3	Mi When child and I argue we stay angry for a very long time
MiPDis5	Mi When child and I disagree, child storms out of the room

5.6.12 Parenting and parent-child relationship: parent-child activities

Questions on parent-child activities at sweep 9 were adapted from similar questions asked on the Growing Up in Ireland study (Growing Up in Ireland, ESRI). The relevant questions are outlined in Table 5.12. Questions were also asked of resident partners of the main GUS respondent who completed the paper self-completion questionnaire (see separate dataset and documentation).

Table 5.12	Selected items adapted from Growing Up in Ireland: Parent-child activities (main carer questionnaire))
Variable name	**Description**
MiPene	Mi How often do you and child sit down to eat together?
MiPeng	Mi How often do you and child play sports or games together?
MiPent	Mi How often do you and child watch TV together?
MiPenh	Mi How often do you and child do household activities together?
MiPeno	Mi How often do you and child go on an outing together?
MiPens	Mi How often do you and child go shopping together for things that child needs?
MiPend	Mi How often do you take child to places child needs to go?

5.6.13 Parenting: autonomy and control (selected items from Epstein's Mother-Father-Peer Inventory Scale)

As part of the main carer self-completion questionnaire, parents/carers were asked about their parenting practices, drawing on selected items from Epstein's Mother-Father-Peer Inventory Scale (*Epstein, 1983*). These questions were previously asked as part of a between-sweep web-CATI survey with GUS parents around the time the child was in Primary 5 (see separate dataset and documentation, forthcoming). Questions are detailed in Table 5.13.

Table 5.13	Selected items from the Mother-Father-Peer Inventory Scale (main carer questionnaire)
Variable name	**Description**
MiPInd01	Mi I encourage child to take own decisions
MiPInd04	Mi I'm always telling child how to behave
MiPInd05	Mi I often worry that child will be hurt or become ill

MiPInd06	Mi I help child to become an independent person
MiPInd09	Mi I encourage child to express opinion
MiPInd12	Mi I encourage child to do things by themselves
MiPInd13	Mi I'm overprotective of child
MiPInd14	Mi I'm always telling child what to do and how to behave

5.6.14 Co-parenting: selected items from Feinberg's Multi-Domain Self Report Measure of Co-parenting (Feinberg et al., 2012)

In cases where the main carer was living with someone as a couple they were asked a selected questions from Feinberg's Multi-Domain Self Report Measure of Co-parenting (Feinberg et al., 2012). Questions were asked of resident partners who took part in the paper self-completion questionnaire (see separate dataset and documentation). The items are listed in Table 5.14.

Table 5.14	Selected items from the Multi-Domain Self Report Measure of Coparenting (main carer questionnaire)
Variable name	**Description**
MiCprp	Mi I believe my partner is a good parent
MiCrpt	Mi My partner tries to show that they are better than me at caring for child
MiCpra	Mi My partner pays a great deal of attention to child
MiCpru	Mi My partner undermines my parenting
MiCprh	Mi My partner appreciates how hard I work at being a good parent
MiCprb	Mi My partner does not trust my abilities as a parent
MiCprg	Mi My partner and I have the same goals for child
MiCpri	Mi My partner and I have different ideas about how to raise child

5.6.15 Parental physical and mental health

At Sweep 9, 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 5.15 Constituent and derived variables associated with the SF-12	
Variable name	**Description**
MiHpgn01	Mi How is resp health in general
MiHlmt01	Mi Resp health limits moderate activities
MiHlmt02	Mi Resp health limits climbing stairs
MiHlmt03	Mi Resp health limited accomplishments past 4 wks
MiHlmt04	Mi Resp health limited reg activities past 4 wks
MiHlmt05	Mi Resp mental health limited accomplishments past 4 wks
MiHlmt06	Mi Resp mental health limited quality of accomplishments past 4 wks
MiHlmt07	Mi Resp physical pain limited normal work past 4 wks
MiHpgn02	Mi Time resp felt calm in past 4 wks
MiHpgn03	Mi Time resp felt energetic in past 4 wks
MiHpgn04	Mi Time resp felt down in past 4 wks
MiHpgn05	Mi Time resp health interfered socially in past 4 wks
DiSF12ph	Di Physical PCS - 12 Scale
DiSF12mn	Di Mental MCS - 12 Scale

5.6.16 Parent health behaviours: alcohol and smoking (selected items from the Scottish Health Survey)

As part of their self-completion questionnaire, main carers were asked a number of questions about alcohol and smoking. Questions were adapted from the Scottish Health Survey (SHeS) and are listed in Table 5.16

Table 5.16	Selected items adapted from SHeS – parental alcohol and smoking (main carer questionnaire)
Variable name	**Description**
MiHcig02	Mi Resp smoking habits
MiHcig07	Mi does anyone else in house smoke in the house

MiHcig10	Mi do you smoke in the same room as child
MiHcig11	Mi does anyone else in house smoke in the same room as child
MiHalc05	Mi How often resp drinks alcohol
MiHalc06	Mi How many units of alcohol resp drinks
MiHalc07	Mi How often in last year resp not able to stop drinking once started
MiHalc08	Mi How often in last year resp failed to do what was expected because of drinking
MiHalc09	Mi Relative/friend/doctor/health worker concerned about resp's drinking or advised to cut down
MiHalc10	Mi How often resp has 6 or more units of alcohol on one occasion

5.6.17 Additional measures sourced elsewhere

In addition to the items outlined above, Table 5.17 details items which were either directly sourced elsewhere or adapted from existing sources.

Table 5.17		
Variable name(s)	**Description**	**Source**
CiSchh, CiScho, CiCasE	Items on homework and educational aspirations [child questionnaire]	Adapted from the Millennium Cohort Study (Sweep 5) (CLS: 2017)
CiIPkn	Upsetting experiences online [child questionnaire]	Adapted from the EU Kids Online II Child questionnaire (LSE: April 2010)
MiPConf	Parental confidence in parenting [main carer questionnaire]	From the Maternal Postnatal Attachment Scale (Condon & Corkindale, 1998)

5.6.18 Child 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 5.18	Child derived BMI variables
Variable name	**Description**
DiBMI	Di BMI (reliable child weight measurements only)
DiUKbmi	Di UK BMI national classification standards
DiINTbmi	Di International BMI cut-offs
DiINTbmi2	Di BMI status (ovrwt inc. obese) - international cut-offs
DiINTbmi3	Di BMI status (non-obese vs obese) - international cut-offs

DiISDbmi	Di ISD BMI 5 group classification
DiISDHWt	Di Study child weight within/outwith ISD healthy range
DiISDovW	Di Study child overweight, including obese (ISD)

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:

   o Detailed ethnicity
   o Detailed religion
   o Detailed geography variables
   o Language spoken at home
   o Full interview date
   o Full date of birth
   o Timing variables

There are no geographical variables in the archived dataset beyond a binary area urban-rural classification and the Scottish index of multiple deprivation summary variable. As noted in section 5.6.6, access to more detailed variables is possible via the UKDS Secure Licence facility.

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.

Centre for Longitudinal Studies (2017): University of London, Institute of Education, Millennium Cohort Study: Fifth Survey, 2012. [data collection]. 4th Edition. UK Data Service. SN: 7464, http://doi.org/10.5255/UKDA-SN-7464-4

Condon, J.T. and Corkindale, C.J., (1998). The assessment of parent-to-infant attachment: Development of a self-report questionnaire instrument. Journal of Reproductive and Infant Psychology, 16(1), pp.57-76.

Department of Social Services; Australian Institute of Family Studies; Australian Bureau of Statistics, 2018, "Growing Up in Australia: Longitudinal Study of Australian Children (LSAC) Release 6 (Waves 1-6)", doi:10.26193/JOZW2U, ADA Dataverse, V

Epstein, S. (1983). Scoring and interpretation of the Mother-Father-Peer Scale. Unpublished manuscript, University of Massachusetts, Department of Psychology, Amherst.

Feinberg, M.E., Brown, L.D. and Kan, M.L., (2012). A multi-domain self-report measure of coparenting. Parenting, 12(1), pp.1-21.

The KIDSCREEN Group Europe. (2006). The KIDSCREEN Questionnaires - Quality of life questionnaires for children and adolescents. Handbook. Lengerich: Pabst Science Publishers.

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

Growing up in Ireland: Economic and Social Research Institute (ESRI), Questionnaires Available at: https://www.growingup.ie/questionnaires/

London School of Economics (2010): EU KIDS ONLINE 2 Child Questionnaire, Available at: http://www.lse.ac.uk/media-and-communications/assets/documents/research/eu-kids-online/toolkit/Questionnaire-for-child.pdf

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

Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Duer, W., Auquier, P., Power, M., Abel, T., Czemy, L., Mazur, J., Czimbalmos, A., Tountas, Y., Hagquist, C., Kilroe, J. and the European KIDSCREEN Group. (2005). KIDSCREEN-52 quality-of-life measure for children and adolescents. Expert Review of Pharmacoeconomics & Outcomes Research, 5 (3), 353-364.

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

ScotCen Social Research. Scottish Health Survey, UK Data Service.

Smith, D. J. (2004). Edinburgh Study of Youth Transitions and Crime : Waves One to Four, 1997-2001. [data collection]. UK Data Service. SN: 4800

World Health Organisation (WHO): Health Behaviour in School-aged Children (HBSC) Survey. Available at: http://www.hbsc.org/about/index.html

8 Contact details

Further details about the study and a list of publications using the data can be found on the study website: growingupinscotland.org.uk. There is also a list of current projects using the data.

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

gus@scotcen.org.uk

ScotCen Social Research

Scotiabank House

6 South Charlotte Street

Edinburgh, EH2 4AW

T: 0131 240 0210

9 Appendices

9.1 Appendix A: 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: 3 cases had a partial interview (code 210 or 211 at variable MiOutcome), 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, 1 case had a child interview, but no adult interview (code 212 at MiOutcome).