Scottish Centre *for* Social Research

Growing Up In Scotland Sweep 6: 2010-2011 *User Guide*

Authors: Paul Bradshaw, Joan Corbett and Sarah Tipping

1 Overview of the survey

The data files contain data from **birth cohort one** of Growing Up in Scotland (GUS) Sweep 6, the sixth year of a longitudinal research study aimed at tracking the lives of three cohorts of Scottish children from the early years, through childhood and beyond. Funded by the Scottish Government Education Directorate, its principal aim is to provide information to support policy making, but it is also intended to be a broader resource for secondary analysis.

The overarching aim of the study is set out in it's 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 6, data collection for the study included two main elements:

1. A face-to-face CAPI interview with the cohort child's main carer
2. Height and weight measurement of the cohort child and his/her natural mother

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 one'), CC1 to the slightly older cohort ('child cohort') and BC2 to the most recent birth cohort ('birth cohort two'). Note that at sweep six, data was collected from BC1 only.

Table 1.1 Sample design: sweeps 1 to 7

Sweep *Launch year*	Cohort and age at interview							
	0-1	**1-2**	**2-3**	**3-4**	**4-5**	**5-6**	**6-7**	**7-8**
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		-
1a ***2011/12***	*BC2*						-	
7 ***2012/13***								BC1
2a ***2013/14***			*BC2*					

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

1. *Cross-sectional time specific data* – e.g. what proportion of 2-3 year-olds are living in single parent families in 2005?
2. *Cross-sectional time series data* – e.g. is there any change in the proportion of 10 month old children living in single parent families between 2005 and 2011?
3. *Longitudinal cohort data* – e.g. what proportion of children who were living in single parent households aged 0-1 are living in different family circumstances at age 5-6?

1.2 Sample Design[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

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

[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.

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.

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 Development and Piloting

Policy priorities and key topics of interest for the sweep 6 questionnaire were initially discussed and agreed by the study's Scottish Government Project Managers, internal policy and external academic stakeholders. The questionnaire was 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. A full instrument was initially piloted in CAPI in November 2009. This instrument was revised for the second CAPI 'Dress Rehearsal' Pilot in January 2010.

2 Data collection methods

2.1 Mode of data collection

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

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

At the following sweeps, interviewers were instructed to undertake the interview with the same respondent as in the previous sweep. At Sweep 6, this means the same respondent as Sweep 5 (or Sweep 4 / Sweep 3 / Sweep 2 / Sweep 1 if the household skipped some of the sweeps). Where this was not possible or appropriate, interviews were conducted with the child's main carer. In practice, most interviews were undertaken with the previous sweep respondent (98.4% of interviews were with the previous respondent) and this was usually the child's mother (97.5% of interviews were with the child's mother).

2.2 Length of Interview

Overall, the average interview (including height and weight measurements) lasted around 70 minutes. The median interview length was 62 minutes.

2.3 Timing of fieldwork

Fieldwork was undertaken over a fourteen month period commencing in April 2010. The sample was issued in twelve monthly sweeps at the beginning of each month and each month's sample was in field for a maximum period of two and a half months. For example, sample 2 was issued at the beginning of May 2010 and remained in field until mid-July 2010.

To ensure that respondents were interviewed when their children were approximately the same age, each case was assigned a 'target interview date'. This was identified as the date on which the child turned 58.5 months old. Interviewers were allotted a four-week period based on this date (two weeks either side) in which to secure the interview. In difficult cases, this period was extended up to and including the child's subsequent birthday which allowed a further four weeks. The vast majority of interviews were achieved within the four-week target period.

3 Height and weight measurements

Child's height and weight measurements were previously taken in SW2 and SW4 and repeated in SW6. Natural mother's height and weight measurements were taken for the first time in SW6. The relationship between general build and health is of great interest to the Scottish Government,

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

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

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

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

For the weight measurements, there was an option to weigh the child whilst being held by an adult. In this case, the adult was weighed on his/her own first and then the adult and the child were weighed together. Both weights were entered in the computer, which calculated the child's weight.
The interviewer was asked to code whether they experienced problems with the height and/or weight measurements and, if they did, to indicate whether they felt the end result was reliable or unreliable at WfXhei14 and WfXwei19. As a rough guide, if the measurement was likely to be more than 2 cms (3/4 inch) from the true figure for height or 1 kg (2 lbs) from the true figure for weight, it was coded as unreliable.

If the respondent was not willing to allow the sample child to have his/her height or weight measured, for example saying that they were too busy or already knew their measurements, a Refusal code was entered for the measurements variables WdXhei01 and WdXwei01, with the reason for refusal at WfXhei02 or WfXwei02.

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

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

The data has been used to estimate an approximate BMI (Body Mass Index) score for each child. Further details on the data and variables associated with the height and weight measurements can be found in section 7.

4 Response rates

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

Table 4.1 Number of issued and achieved cases and response rates

	Birth Cohort
Achieved interviews at sweep 1	5217
Achieved interviews at sweep 2	4512
Achieved interviews at sweep 3	4193
Achieved interviews at sweep 4	3994
Achieved interviews at sweep 5	3833
Cases to field at sweep 6:	
All issued to field*	4247
Cases achieved at sweep 6	3657
Response rate	
As % of all issued cases at sweep 6	86%
As % of all sweep 1 cases	70%

* The number of cases issued to the field at sweep 6 is higher than the number of Interviews achieved at sweep 5 because some of the sweeps 1 to 5 respondents missed at sweep 5 were re-issued at sweep 6.

5 Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted. The GUS Sweep 6 Coding Instructions, deposited along with this User Guide, provide details of the tasks that were conducted.

6 Weighting the data

Two sets of weights have been developed for sweep 6 of the birth cohort:

1. A cross-sectional weight that should be used for any cross-sectional analysis of Sweep 6 data only. All sample members that responded at Sweep 6 have a cross-sectional weight.
2. A longitudinal weight for analysis of more than one sweep of data. Sample members that have responded at every sweep of GUS thus far have a longitudinal weight.

6.1 Background

6.1.1 The Sweep 6 sample

The Sweep 6 sample can be split into two components; for the purposes of describing the weighting these two components have been named Sample A and Sample B and are defined as follows:

- Sample A – individuals who had responded at all previous sweeps
- Sample B – individuals who had responded at Sweep 1 but had missed one or more interviews in sweeps 2-5.

The two samples will be treated separately during the weighting. This is because the Sample B respondents are likely to have different response behaviour to respondents in Sample A, as suggested by their much lower response rates. There were 637 individuals in Sample B, 282 (44%)

of which responded at Sweep 6. The response rates for Sample A were much higher at 94%. The issued and responding sample sizes are given in Table 1.

Table 6.1 Response rates for different samples

	Issued	Responding	Response rate
Sample A	3583	3375	94%
Sample B	637	282	44%
Combined (A+B)	4220[3]	3657	87%

Two sets of weights were developed; a cross-sectional weight and a longitudinal weight. The longitudinal weight will be used for any analysis that includes more than one sweep of data. Only members of Sample A (who have responded at every sweep of GUS) will have a longitudinal weight. This weight is described in more detail in Section 3.

The cross-sectional weight will be used for any cross-sectional analysis of Sweep 6 data. All SWEEP 6 respondents will have a cross-sectional weight (Sample A + B). These are described in more detail in Section 4.

6.1.2 Longitudinal weights

Longitudinal weights were only generated for respondents in Sample A. A model-based weighting technique was used to develop the sweep 6 longitudinal weights, where response behaviour is modelled using data from previous sweeps. This is the same method used to generate weights at sweeps 2 to 5. Ineligible households (deadwood) were not included in the non-response modelling[3].

Response behaviour was modelled using logistic regression. A logistic regression models the relationship between an outcome variable (in this case response to the sweep 6 interview) and a set of predictor variables. The predictor variables were a set of socio-demographic respondent and household characteristics collected from the previous two sweeps.

The model generated a predicted probability for each respondent. This is the probability the respondent would take part in the interview, given the characteristics of the respondent and the household. Respondents with characteristics associated with non-response (such as being a private tenant) are under-represented in the sample and will receive a low predicted probability. The non-response weights are then generated as the inverse of the predicted probabilities; hence respondents who had a low predicted probability get a larger weight, increasing their representation in the sample.

A summary of the characteristics related to response behaviour for the birth cohort at Sweep 6 are given in Table 6.2. The full model is given in Table A1 in the Appendix.

[3] There were 27 individuals with ineligible outcome codes; these individuals were dropped from the analysis. Ineligible outcome codes include households that were vacant, demolished or derelict and non-residential addresses, where no follow up address could be found.

Table 6.2 Characteristics associated with response and non-response

Characteristics associated with response	Characteristics associated with non-response
At least one parent/carer in full-time employment	No parent/carer working
Mother aged 25 or over	Younger mother aged under 20
Gave income	Not gave income at previous sweep
Lives in other urban areas	Lives in large urban areas

Final sweep 6 longitudinal weights

The final Sweep 6 weight is the product of the Sweep 6 non-response weight and the sweep 6 interview weight. For each cohort the final weights were scaled to the responding sweep 6 sample size, this makes the weighted sample size match the unweighted sample size. Table A2 in the Appendix shows the distribution of the sample weighted by the Sweep 5 and Sweep 6 weights; showing the reduction in bias caused by the Sweep 6 weights.

6.1.3 Cross-sectional weights

Cross-sectional weights were generated for all respondents at Sweep 6 (the combined A and B samples) and should be used for any cross-sectional analysis of Sweep 6 data.

Calibration weighting methods were used to create the cross-sectional weights. This method takes the pre-calibrated weighted combined sample and adjusts the weights using an iterative procedure. The resulting weighting factors, when applied to the combined data, will make the survey estimates match a set of population estimates for a set of key variables. The population estimates in this instance are survey estimates from Sample A, weighted by the longitudinal weight. Since the longitudinal weight corrects for sampling error and non-response bias at each stage of GUS, the weighted Sample A estimates are the best population estimates available. The key variables used in the weighting were: area level deprivation indicator; respondent employment status; respondent age; household income and whether the respondent was a lone parent.

The pre-calibration weights were the Sweep 6 longitudinal weight for Sample A and the weight from the last completed Sweep for Sample B. Prior to calibration these weights were scaled to the achieved sample size, giving a mean weight of one. This was done separately for each sample.

The calibration corrects for any differences due to differential non-response between Sample A and Sample B. The weighted distribution of Sample A and the weighted distribution of the combined sample, pre and post-calibration, are given in Table A3 in the appendix.

6.2 Sample efficiency

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

The affect of the sample design on the precision of survey estimates is indicated by the effective sample size (neff). The effective sample size measures the size of an (unweighted) simple random sample that would have provided the same precision (standard error) as the design being

implemented. If the effective sample size is close to the actual sample size then we have an efficient design with a good level of precision. The lower the effective sample size, the lower the level of precision. The efficiency of a sample is given by the ratio of the effective sample size to the actual sample size. The range of the weights, the effective sample size and sample efficiency for both sets of weights are given in Table 3.

Table 6.3 Range of weights and sample efficiency

	Minimum	Maximum	Mean	N	Neff	Efficiency
Baby cohort						
Longitudinal weight	0.62	3.00	1.00	3657	3262	89.2%
Cross-sectional weight	0.63	3.06	1.00	3375	2949	87.4%

6.3 Applying the weights

The cross-sectional weights should be used for any cross-sectional analysis, i.e. any analysis of Sweep 6 data only. All sample members that responded at Sweep 6 have a cross-sectional weight.

The longitudinal weight should be used for any analyses of more than one sweep of data that incorporates data from Sweep 6. Sample members that have responded at every sweep of GUS have a longitudinal weight.

Table 6.4 Description of weight variables in the data file

Variable name	Label
DfWTbrth	Df Birth cohort Sweep 6 weight
DfWTbth2	Df Birth cohort Sweep 6 weight - longitudinal

7 Using the data

The GUS Sweep 6 data consists of the following SPSS file:

GUS_SW6_B.sav	3657 cases	Birth cohort

7.1 Variables on the data file

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

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 MfSDQ01 to 25. The phrasing of the question and the list of answers provided on the showcards - if any - shape the variables. The user must therefore take these variations into account when creating derived variables.

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

7.2 Variable naming convention

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

Table 7.1 GUS variable naming conventions

<table>
<tr><th colspan="4">Character No:</th></tr>
<tr><th colspan="2">1</th><th colspan="2">2</th><th>3</th><th>4, 5 & 6</th><th>7&8</th></tr>
<tr><td colspan="2">Source of data</td><td colspan="2">Sweep/Sweep</td><td>Key theme prefix</td><td>Sub theme stem</td><td>Question/ Variable number</td></tr>
<tr><td colspan="2">Non- sequential Capitals: D,M, P, S</td><td colspan="2">Sequential lower case: a, b, c..</td><td rowspan="8">Non-sequential Capitals: C, P, N…</td><td rowspan="8">Abbreviated lower case: e.g. hea,</td><td rowspan="8">01 - 99</td></tr>
<tr><td>Source code</td><td>Details</td><td>Sweep code</td><td>Details</td></tr>
<tr><td>AL</td><td>Area Level variable</td><td>a</td><td>Sweep 1 (2005/06)</td></tr>
<tr><td>D</td><td>Derived variable</td><td>b</td><td>Sweep 2 (2006/07)</td></tr>
<tr><td>DP</td><td>Derived variable from partner int</td><td>c</td><td>Sweep 3 (2007/08)</td></tr>
<tr><td>DWP</td><td>DWP variable</td><td>d</td><td>Sweep 4 (2008/09)</td></tr>
<tr><td>M</td><td>Main carer/adult interview</td><td>e</td><td>Sweep 5 (2009/10)</td></tr>
<tr><td>P</td><td>Partner's interview</td><td>f</td><td>Sweep 6 (2010/11)</td></tr>
</table>

7.3 Variable labels

In the Sweep 6 dataset the variable labels have been shortened to 40 characters as far as possible; the first 2 show the source and year of the data (as in the variable name). Although the labels give an indication of the topic of the question it is essential to refer to the questionnaire to see the full text of the question and the routing applied to that variable. The variable list shows the page numbers of the relevant questionnaire section.

7.4 Derived variables

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

7.5 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 MfHGSl(n) can be used to see whether a person who was in the household at sweep 1, 2, 3 4 or 5 is still in the household at sweep 6.

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

DfHGnmad	Df - Number of adults (16 or over) in household
DfHGnmkd	Df - Number of children in household
DfHGnmsb	Df - Number of siblings in household
DfHGnp01	Df - Number of natural parents in household
DfHGrsp01	Df - Whether respondent is natural mother
DfHGrsp02	Df - Whether respondent is natural father
DfHGnp02	Df - Natural mother in household
DfHGnp03	Df - Natural father in household
DfHGnp04	Df - Respondent living with spouse/partner
DfMothID	Df – Mother's ID (= Person number in household)
DfFathID	Df - Father's ID
DfRespID	Df – Respondent's ID
DfPartID	Df – Respondent's partner's ID
DfRPAge	Df – Respondent's partner's age (banded)
DfRPsex	Df - Respondent partners sex

7.6 Childcare data

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

At Sweep 6, for those cases where childcare had been used at the previous sweep, details of the previous sweep arrangements – including the provider name, provider type, the number of hours they looked after the child per week and the number of days over which those hours were spread – were fed-forward. The respondent could confirm whether all details were still correct, change the number of hours or days, or indicate that the arrangement was no longer being used. All respondents could also provide details of new arrangements which were in place at sweep 6 but had not been in place at the previous sweep. The multiple sets of information collected create a particularly complex data structure.

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

Table 7.3 Childcare variables for exploring continuity of provision

Variable name	Description
MaCtya01/DbCtya01/../DfCtya01	Sw1 / ... / Sw6 1st childcare provider type
MaCtma01/DbCtma01/../DfCtma01	Sw1 / ... / Sw6 1st childcare provider - no of hours per week
MaCdya01/DbCdya01/../DfCdya01	Sw1 / ... / Sw6 1st childcare provider - no of days per week
MaCtyb01/DbCtyb01/../DfCtyb01	Sw1 / ... / Sw6 2nd childcare provider type
MaCtmb01/DbCtmb01/../DfCtmb01	Sw1 / ... / Sw6 2nd childcare provider - no of hours per week
MaCdyb01/DbCdyb01/../DfCdyb01	Sw1 / ... / Sw6 2nd childcare provider - no of days per week
MaCtyc01/DbCtyc01/../DfCtyc01	Sw1 / ... / Sw6 3rd childcare provider type

MaCtmc01/DbCtmc01/../DfCtmc01	Sw1 / ... / Sw6 3rd childcare provider - no of hours per week
MaCdyc01/DbCdyc01/../DfCdyc01	Sw1 / ... / Sw6 3rd childcare provider - no of days per week
MaCtyd01/DbCtyd01/../DfCtyd01	Sw1 / ... / Sw6 4th childcare provider type
MaCtmd01/DbCtmd01/../DfCtmd01	Sw1 / ... / Sw6 4th childcare provider - no of hours per week
MaCdyd01/DbCdyd01/../DfCdyd01	Sw1 / ... / Sw6 4th childcare provider - no of days per week
MaCtye01/DbCtye01/../DfCtye01	Sw1 / ... / Sw6 5th childcare provider type
MaCtme01/DbCtme01/../DfCtme01	Sw1 / ... / Sw6 5th childcare provider - no of hours per week
MaCdye01/DbCdye01/../DfCdye01	Sw1 / ... / Sw6 5th childcare provider - no of days per week
MfCsta01	Mf Whether still using 1st provider from last sweep
MfCcta01	Mf Previous 1st ccare provider - revised hrs at sweep 6
MfCcda01	Mf Previous 1st ccare provider - revised days at Sw6
MfCrsa01	Mf - Why not using prev provider 1 at Sw6
MfCstb01	Mf Whether still using 2nd provider from last sweep
MfCctb01	Mf Previous 2nd ccare provider - revised hrs at Sw6
MfCcdb01	Mf Previous 2nd ccare provider - revised days at Sw6
MfCrsb01	Mf - Why not using prev provider 2 at Sw6
MfCstc01	Mf Whether still using 3rd provider from last sweep
MfCctc01	Mf Previous 3rd ccare provider - revised hrs at Sw6
MfCcdc01	Mf Previous 3rd ccare provider - revised days at Sw6
MfCrsc01	Mf - Why not using prev provider 3 at Sw6
MfCstd01	Mf Whether still using 4th provider from last sweep
MfCctd01	Mf Previous 4th ccare provider - revised hrs at Sw6
MfCcdd01	Mf Previous 4th ccare provider - revised days at Sw6
MfCrsd01	Mf - Why not using prev provider 4 at Sw6
MfCste01	Mf Whether still using 5th provider from last sweep
MfCcte01	Mf Previous 5th ccare provider - revised hrs at Sw6
MfCcde01	Mf Previous 5th ccare provider - revised days at Sw6
MfCrse01	Mf - Why not using prev provider 5 at Sw6
DfCstp01	Df Whether any of the previous ccare arrgmts stopped
DfCstp02	Df No of previous sweep providers stopped
DfCnpv01	Df No of ccare provs from last sweep still being used
DfCapv01	Df Whether resp still uses a previous ccare provider

The second section is concerned with the details of **new arrangements** which were in place at Sweep 6. These variables include details of the provider type, the number of hours and days per week they look after the child, the child's age when the arrangement commenced and the reasons given for using the provision. Details of the variables are listed in Table 7.4.

Table 7.4 Variables for exploring new childcare arrangements at sweep 6

Variable name	Description
MfCany02	Mf If no ccare at last sweep whether using ccare at Sw6
MfCany03	Mf If ccare at last sweep - any new prov at Sw6
MfCtya01	Mf New provider 1 - type
MfCtma01	Mf 1st new ccare provider - hours per week
MfCdya01	Mf 1st new ccare provider - number of days per week
MfCaga01	Mf Age (months) started new provider 1

MfCwya01 – MfCwya18	Mf Reasons for using 1st new provider
MfCtyb01	Mf New provider 2 - type
MfCtmb01	Mf 2nd new ccare provider - hours per week
MfCdyb01	Mf 2nd new ccare provider - number of days per week
MfCagb01	Mf Age (months) started new provider 2
MfCwyb01 – MfCwyb18	Mf Reasons for using 2nd new provider
MfCtyc01	Mf New provider 3 - type
MfCtmc01	Mf 3rd new ccare provider - hours per week
MfCdyc01	Mf 3rd new ccare provider - number of days per week
MfCagc01	Mf Age (months) started new provider 3
MfCwyc01 – MfCwyc18	Mf Reasons for using 3rd new provider
MfCtyd01	Mf New provider 4 - type
MfCtmd01	Mf 4th new ccare provider - hours per week
MfCdyd01	Mf 4th new ccare provider - number of days per week
MfCagd01	Mf Age (months) started new provider 4
MfCwyd01 – MfCwyd18	Mf Reasons for using 4th new provider
DfCnnp01	Df No of new childcare arrangements at Sweep 6

Information from the first two sections was used to derive a set of variables forming the third section – **current arrangements.** These derived variables indicate - for all childcare arrangements in place at the time of the Sweep 6 interview - the provider type, number of hours and days of the arrangement, and whether or not it is a new arrangement at Sweep 6. A range of summary variables indicating, for example, use of any childcare, total number of providers, total hours looked after by all providers and use of different provision are also included. These variables are detailed in Table 7.5.

Table 7.5 Variables for exploring current childcare arrangements at sweep 6

Variable name	Description
DfCtya01	Df - Childcare prov A: provider type
DfCnwa	Df - Provider A: new or existing
DfCtma01	Df Provider A: No of hours per week
DfCdya01	Df Provider A: No of days per week
DfCtyb01	Df - Childcare prov B: provider type
DfCnwb	Df - Provider B: new or existing
DfCtmb01	Df Provider B: No of hours per week
DfCdyb01	Df Provider B: No of days per week
DfCtyc01	Df - Childcare prov C: provider type
DfCnwc	Df - Provider C: new or existing
DfCtmc01	Df Provider C: No of hours per week
DfCdyc01	Df Provider C: No of days per week
DfCtyd01	Df - Childcare prov D: provider type
DfCnwd	Df - Provider D: new or existing
DfCtmd01	Df Provider D: No of hours per week
DfCdyd01	Df Provider D: No of days per week
DfCtye01	Df - Childcare prov E: provider type
DfCnwe	Df - Provider E: new or existing
DfCtme01	Df Provider E: No of hours per week

DfCdye01	Df Provider E: No of days per week
DfCany01	Df Whether resp uses regular CCare at Sw5 (not including the excluded pre-school cases – see 7.6.1)
DfCtot01	Df Number of ccare providers at Sw5 (not including the excluded pre-school cases – see 7.6.1)

Although not listed in Table 7.5, this section also covers variables associated with cost, availability, choice and preferences. Details of these questions and the corresponding variables are available in the Sweep 6 questionnaire documentation which accompanies this user guide.

7.7 Indicators and summary variables

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

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

The Growing Up in Scotland dataset includes the five category system in which respondents and their partner, where applicable, are classified as managerial and professional, intermediate, small employers and own account workers, lower supervisory and technical, and semi-routine and routine occupations. A sixth category 'never worked' is also coded on this variable. The decision on whether or not this category should be included as a separate category, incorporated with category 5 'Semi-routine or routine' or set to 'missing' is dependent on the particular analysis to which it is being applied.

Further information on NS-SEC is available from the National Statistics website at: http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html.

7.7.2 Socio-economic characteristics: Equivalised household annual income

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

Table 7.7 Income equivalence scales for household members

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

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

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

GUS collects a banded version of total net household income from all sources in the main CAPI interview. This income data is adjusted, using the above equivalence scale, according to the characteristics of the household, to produce an equivalised annual household income value. Variables with the full equivalised income scale (DfEqvinc) and quintiles of the scale (DfEqv5) are available in the datasets.

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

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

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

Table 7.8 Scottish Government Urban Rural Classification

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

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

7.7.4 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 6 datasets is based on the 2009 version of SIMD. It should be noted that the analyses in the GUS Sweep 1 report are based on the 2004 version of SIMD as the 2006 version – which was used for the GUS Sweep 2 to 4 reports - had not been published at the time the Sweep 1 report was being written.

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

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

7.7.5 Area-level variables: Carstairs Index

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

7.7.6 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 'ALfHBdBc' 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
Lanarkshire	Identified
Lothian	Identified
Orkney	*Aggregated*
Shetland	*Aggregated*
Tayside	Identified
Western Isles	*Aggregated*

7.7.7 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[4]. 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
DfDsdem1	Df SDQ: Emotional symptoms score
DfDsdco1	Df SDQ: Conduct problems score
DfDsdhy1	Df SDQ: Hyper-activity or inattention score
DfDsdpr1	Df SDQ: Peer problems score
DfDsdps1	Df SDQ: Pro-social score
DfDsdto1	Df SDQ: Total difficulties score

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

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

Body Mass Index (BMI), i.e. weight divided by height squared, is a score that adjusts your weight for your height. Taken as a number in isolation, the BMI it does not actually represent anything medically. It is only meaningful in the context of a distribution of values for a population. Individuals are 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.

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

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 (Prentice, 1998) and Barlow and Dietz (Barlow & Dietz, 1998).

The main child overweight and obesity variables have been produced using the International Obesity Taskforce cut-offs. These cut-offs are based on BMI reference data from six different countries around the world (over 190,000 subjects in total aged 0 to 25 from UK, Brazil, Hong Kong, the Netherlands, Singapore, and the United States). In summary, the BMI percentile curves that pass through the values of 25 and 30 kg/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 7.11 Child Derived BMI variables

Variable name	Description
DfBMI	Dd BMI (reliable child weight measurements only)
DfUKbmi	Dd UK BMI national classification standards
DfINTbmi	Dd International BMI cut-offs
DfINTbmi2	Dd BMI status (ovrwt inc. obese) - international cut-offs
DfINTbmi3	Dd BMI status (non-obese vs obese) - international cut-offs
DfISDbmi	Dd ISD BMI 5 group classification
DfISDcbmi	Dd ISD BMI 5 group classification (excl. far outliers)
DfISDHWt	Dd Study child weight within/outwith ISD healthy range
DfISDcHWt	Dd Study child weight within/outwith ISD healthy range (excl. far outliers)
DfISDovW	Dd Study child overweight, including obese (ISD)
DfISDcOvW	Dd Study child overweight, including obese (ISD excl. far outliers)

Table 7.10 Adult Derived BMI variables

Variable name	Description
DfMumBMI	Df Nat mother's valid BMI inc. estimated wt>130kg
DfMbmig5	Df Nat Mother's valid BMI (grouped)

7.8 Dropped Variables

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

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

1. Those containing text
2. Those which contained a personal identifier (e.g. name/address)
3. Those considered to be disclosive, such as:

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

There are no geographical variables in the archived dataset beyond area urban-rural classification, the Scottish index of multiple deprivation summary variable, and a derived variable identifying some of the Scottish Health Board areas.

7.9 Missing values conventions

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

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

8 Documentation

The documentation has been organised into the following sections:

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

9 Related publications

A series of Scottish Government reports have already been published which utilised the sweep 6 data. The reports can be found on the Scottish Government website with links also available from the Growing Up in Scotland website.

Bradshaw, P., Hall, J., Hill, T., Mabelis, J. and Philo, D. (2012) *Growing Up in Scotland: Early experiences of Primary School,* Edinburgh: Scottish Government
http://www.scotland.gov.uk/Publications/2012/05/7940/0

Parkes, A., Sweeting, H. and Wight, D. (2012) *Growing Up in Scotland: Overweight, obesity and activity,* Edinburgh: Scottish Government
http://www.scotland.gov.uk/Publications/2012/05/5385/0

Jamieson, L., Warner, P. and Bradshaw, P. (2012) *Growing Up in Scotland: The involvement of grandparents in children's lives,* Edinburgh: Scottish Government
http://www.scotland.gov.uk/Publications/2012/05/4455/0

Other publications which contain findings based on analysis of GUS data include:

Bradshaw, P. (2011) *Growing Up in Scotland: Changes in child cognitive ability in the pre-school years,* Edinburgh: Scottish Government
http://www.scotland.gov.uk/Publications/2011/05/31085122/0

Chanfreau, J., Barnes, M., Tomaszewski, W., Philo, D., Hall, J. and Tipping, S. (2011) *Growing Up in Scotland: Change in early childhood and the impact of significant events,* Edinburgh: Scottish Government
http://www.scotland.gov.uk/Publications/2011/05/25092325/0

Mabelis, J. and Marryat, L. (2011) *Growing Up in Scotland: Parental service use and informal networks in the early years*, Edinburgh: Scottish Government
http://www.scotland.gov.uk/Publications/2011/05/25092504/0

Parkes, A. and Wight D. (2011) *Growing Up in Scotland: Parenting and children's health*, Edinburgh: Scottish Government
http://www.scotland.gov.uk/Publications/2011/05/25092122/0

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

10 Contact details

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

GUS Project Manager	Paul Bradshaw	0131 221 2564	paul.bradshaw@scotcen.org.uk
GUS Senior Researcher	Judith Mabelis	0131 221 2560	judith.mabelis@scotcen.org.uk
GUS Data Manager	Joan Corbett	0131 221 2570	joan.corbett@scotcen.org.uk

Appendix A: Full non-response models

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

	B	S.E.	Wald	df	Sig.	Exp(B)
Age of mother (grouped)			24.465	4	0	
<20					(baseline)	
20-24	-0.367	0.246	2.231	1	0.135	0.693
25-29	0.462	0.271	2.911	1	0.088	1.588
30-34	0.445	0.268	2.753	1	0.097	1.56
35+	0.399	0.286	1.946	1	0.163	1.49
Income			14.145	4	0.007	
<£10,000					(baseline)	
£10,000-£19,999	-0.065	0.244	0.071	1	0.790	0.937
£20,000-£31,999	-0.06	0.294	0.042	1	0.838	0.942
£32,000+	-0.016	0.302	0.003	1	0.957	0.984
Missing	-0.863	0.304	8.077	1	0.004	0.422
Household employment			7.807	2	0.02	
At least one parent/carer in full-time employment					(baseline)	
At least one parent/carer in part-time employment	-0.136	0.224	0.368	1	0.544	0.873
Household employment	-0.63	0.232	7.355	1	0.007	0.533
Urban/rural indicator (Scotland)			12.495	4	0.014	
1 Large urban area (125,000+)					(baseline)	
2 Other urban area (10,000-125,000)	-0.264	0.154	2.937	1	0.087	0.768
3,4,5 All small town (3,000-10,000)	0.275	0.242	1.292	1	0.256	1.316
6 Accessible rural (<3,000)	0.442	0.277	2.536	1	0.111	1.556
7, 8 very remote and Remote rural (<3,000)	0.508	0.395	1.653	1	0.199	1.662
Constant	2.692	0.344	61.148	1	0	14.755

Notes:
1. The response is 1 = sample A response to Sweep 6, 0 = sample A non-response.
2. Model is weighted by sweep 2 baby weight
3. The model R^2 = 0.023 (Cox and Snells).
4. B is the estimate coefficient with standard error S.E.
5. The Wald-test measures the impact of the categorical variable on the model with the appropriate number of degrees of freedom df. If the test is significant (sig < 0.05) then the categorical variable is considered to be 'significantly associated' with the response variable and therefore included in the model.
6. The Wald test for each level of the categorical variable is also shown. This tests the difference between that level and the baseline category.

Table A2 Distribution of sample A

	Sweep 5 weighted by sweep 5 weight	Sweep 6 weighted by sweep 5 weight	Sweep 6 weighted by sweep 6 weight
	%	%	%
Tenure			
Owner occupier	63.9	64.8	63.8
Rents HA/council	27.5	26.9	27.7
Rents private	8.6	8.3	8.5
Family status			
Lone parent	19.4	18.5	19.2
Couple parent	80.6	81.5	80.8
Mother's age	7.4	7.1	7.4
<20	17.1	16.0	17.0
20-24	23.5	23.8	23.5
25-29	31.4	32.0	31.5
30-34	20.6	21.0	20.7
35+	63.9	64.8	63.8
Highest education level of respondent			
Degree or equivalent	28.0	28.4	27.9
Vocational qualification below degree	39.7	40.1	40.1
Higher Grade or equivalent	7.1	7.2	7.2
Standard Grade or equivalent	16.3	15.9	16.1
No Qualifications	8.9	8.4	8.7
Household income			
<£10,000	10.2	9.8	10.2
£10,000-£19,999	19.7	19.5	19.8
£20,000-£31,999	22.3	22.4	22.3
£32,000+	41.7	42.7	41.8
Missing	6.0	5.7	6.0
Respondent NSSEC - 5 Category			
Managerial and professional occupations	33.6	34.1	33.6
Intermediate occupations	19.4	19.4	19.3
Small employers and own account workers	6.4	6.5	6.4
Lower supervisory and technical occupations	5.4	5.3	5.4
Semi-routine and routine occupations	31.7	31.2	31.8
Missing/never worked	3.5	3.4	3.6
Ethnicity of respondent			
White	96.3	96.5	96.5
Other ethnic background	3.7	3.5	3.5

(continued…)

Table A2 Distribution of sample A *(continued)*

	Sweep 5 weighted by sweep 5 weight	Sweep 6 weighted by sweep 5 weight	Sweep 6 weighted by sweep 6 weight
	%	%	%
Household employment			
At least one parent/carer in full-time employment	70.9	72.0	71.0
At least one parent/carer in part-time employment	14.0	13.9	13.9
No parent/carer working	15.1	14.2	15.0
Mother's employment status			
Childs mother working - full-time	15.5	15.6	15.3
Childs mother working - part-time	49.9	50.6	50.1
Childs mother not working	34.6	33.9	34.6
Number of children in the household			
1	21.1	20.6	20.8
2	50.7	51.0	50.8
3	20.5	20.6	20.6
4+	7.7	7.7	7.8
Urban/rural indicator (Scotland)			
Large urban area (125,000+)	36.1	36.1	36.1
Other urban area (10,000-125,000)	33.4	32.6	33.3
All small town (3,000-10,000)	12.7	13.0	12.8
Accessible rural (<3,000)	12.1	12.5	12.2
Remote & Very remote rural (<3,000)	5.6	5.8	5.7
Use regular childcare			
Yes	74.7	75.3	75.1
No	25.3	24.7	24.9
SIMD06 quintiles			
0.95 - 7.75 (least deprived)	19.3	19.5	19.2
7.75 - 13.56	20.4	20.7	20.4
13.56 - 21.04	18.9	19.1	19.0
21.05 - 33.70	19.6	19.0	19.2
33.71 -89.09 (most deprived)	21.9	21.7	22.2
Base (unweighted)	*3595*	*3375*	*3375*

Table A3 Weighted distribution of key variables for samples A and B

	Sample A	Combined Sweep 6 sample (A+B)	
	Weighted by sweep 6 weight	Weighted by pre-calibration weight[1]	Calibrated to sample A
Family type			
Lone parent	19.3	20.1	19.3
Couple family	80.7	79.9	80.7
Household income			
<£10,000	9.1	9.6	9.1
£10,000-£19,999	20.8	21.0	20.8
£20,000-£31,999	20.6	20.8	20.6
£32,000+	43.5	42.6	43.5
Missing	5.9	5.9	5.9
Respondent's age			
<25	5.2	5.4	5.2
25-29	16.0	16.3	16.0
30-34	22.0	22.3	22.0
35-39	31.0	30.3	31.0
40+	25.8	25.6	25.8
Respondent employment status			
Employed	66.2	65.8	66.2
Not employed	33.8	34.2	33.8
SIMD06 quintiles			
0.95 - 7.75 (least deprived)	19.2	18.7	19.2
7.75 - 13.56	20.4	20.2	20.4
13.56 - 21.04	19.0	18.8	19.0
21.05 - 33.70	19.2	19.1	19.2
33.71 -89.09 (most deprived)	22.2	23.2	22.2
Base (unweighted)	*3657*	*3375*	*3375*

[1]This is the SWEEP 6 weight for Sample A and the weight from the last completed sweep for Sample B

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

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

- In the Health and Development section, the CAPI programme used individual variables for each type of long standing illness, the set of variable being repeated three times since up to three illnesses could be recorded. The total number of illnesses mentioned over the combined 3 sets of CAPI variables was never higher than 3. However there could be more than one illness mentioned in a particular set. In the archived dataset, the individual variables for each set have been recoded into one variable for the 1st long standing illness, one for the 2nd and one for the 3rd. When 2 illnesses had been mentioned in the 1st set of CAPI answers, the 1st answer was kept as the 1st illness and the 2nd answer became the 2nd illness. If there was an answer in the 2nd set of illnesses, it was recoded as being the 3rd illnesses, etc. At Sweep 6 this recoding means that
    - for the 2nd illness there will be a higher number of illnessses mentioned in MeHlsb01 than answers at the next variables regarding this 2nd illness, MeHlsb02 to MeHntb07, and
    - there can be more than 3 illnesses mentioned in total, hence additional variable MeHlsd01, but no corresponding set of answers regarding whether it limits activities or the type of treatment/advice received.

- Although the Self-complete section was asked to all respondents, some respondents chose not to complete it and these cases show as missing values ('Not Applicable') in the dataset.

- Seven cases had a partial interview (code 210 at variable MfOutcom), so some information may be missing towards the end of the questionnaire (for religion and ethnicity in particular); those cases show either as -1 'Not Applicable' or as -3 'information not available' in the dataset.