Scottish Centre *for* Social Research

Growing Up In Scotland
Sweep 5: 2009-2010
User Guide

Authors: Paul Bradshaw, Louise Marryat, Judith Mabelis, Mireille Ferrandon and Sarah Tipping

1 Overview of the survey

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

At sweep 5, data collection for the study included two main elements:
1. A face-to-face CAPI interview with the cohort child's main carer
2. Two cognitive assessments undertaken with the cohort child - further details are included in sections 3 and 7

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 and the second aged approximately 34 months. A second birth cohort is being recruited in 2011 with children aged around 10 months at the time of the first interview. All cohorts were named samples drawn from Child Benefit records.

The configuration of cohorts and sweeps for the first five sweeps of data collection is summarised below. BC1 refers to the younger of the two cohorts ('birth cohort') and CC1 to the slightly older cohort ('child cohort'). Note that at sweep five, data was collected from BC1 only.

Table 1.1   Sample design: sweeps 1 to 5

Sweep *Launch year*	Age at interview					
	0-1	1-2	2-3	3-4	4-5	5-6
1 *2005*	BC1		CC1			
2 *2006*		BC1		CC1		
3 *2007*			BC1		CC1	
4 *2008*				BC1		CC1
5 *2009*						BC1

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

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

1.2   Sample Design

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

The Data Zones were aggregated to give an average of 57 births per area per year (based on the average number of births in each Data Zone for the preceding 3 years). It was estimated that this number per area would provide us with the required sample size. Once the merging task was complete, the list of aggregated areas was sorted by Local Authority[1] and then by the Scottish Index of Multiple Deprivation Score (SIMD). 130 areas were then selected at random. The Department of Work and Pensions then sampled children from these 130 sample points.

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

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

---

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

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 study by cohort

Sample Number	Dates of Birth required Birth Cohort
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 5 questionnaire were initially discussed and agreed by the study's Scottish Government Project Managers and Policy Advisory Group. The questionnaire was then developed by the GUS team at ScotCen with input from colleagues at the Centre for Research on Families and Relationships (CRFR) in reference to these priorities and topics. A full instrument was initially piloted in CAPI in November 2008. This instrument was revised for the second 'Dress Rehearsal' Pilot in January 2009.

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 5, this means the same respondent as Sweep 4,

or 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 (98.8% of interviews were with the previous respondent) and this was usually the child's mother (98% of interviews were with the child's mother).

2.2 Length of Interview

Overall, the average interview lasted around 71 minutes. The median interview length was 66 minutes.

2.3 Timing of fieldwork

Fieldwork was undertaken over a fourteen month period commencing in April 2009. The sample was issued in twelve monthly waves at the beginning of each month and each month's sample was in field for a maximum period of two and a half months. For example, sample 2 was issued at the beginning of May 2009 and remained in field until mid-July 2009.

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 Child assessments

As part of the data collection for sweep 5, the birth cohort children were again asked to complete two cognitive assessments, having also done so at sweep 3 (aged approximately 34 months). The assessments – Naming Vocabulary and Picture Similarities - were taken from the Early Years battery of the 'British Ability Scales Second Edition' (BAS II). The British Ability Scales (BAS) is a battery of individually administered tests of cognitive abilities and educational achievements suitable for use with children and adolescents aged from 2 years 6 months to 17 years 11 months. The assessments are normally employed by educational psychologists in a classroom or clinical setting but have been adapted for use in a survey setting, and modified to be administered with the help of a CAPI programme pre-determining the complex set of rules for routing children through each assessment. The purpose and method of each assessment is described in table 3.1. The data is used to estimate an approximate score for each child. Further details on the data and variables associated with the cognitive assessments can be found in section 7.7.7.

3.1 BAS Naming Vocabulary

Naming Vocabulary assesses the spoken vocabulary of young children. The test items consist of a booklet of coloured pictures of objects which the child is shown one at a time and asked to name. The scale measures expressive language ability, and successful performance depends on the child's previous development of a vocabulary of nouns. Picture recognition is also crucial; however, the pictures are large and brightly coloured and are unlikely to cause problems except for children

with major visual impairments or with no experience of picture books. The items require the child to recall words from long-term memory rather than to recognise or understand the meaning of words or sentences.

Naming Vocabulary score may reflect:
- Expressive language skills
- Vocabulary knowledge of nouns
- Ability to attach verbal labels to pictures
- General knowledge
- General language development
- Retrieval of names from long-term memory
- Level of language stimulation

Low scores may reflect reluctance to speak.

3.2   BAS Picture Similarities

Picture Similarities measures the reasoning ability of young children. The test items consist of a booklet with four images on each page and a set of cards each with a single image printed on. The child is shown the row of pictures, given a corresponding card and asked to place the card under the image on the page which shares an element or concept with the image on the card. To undertake the task, the child must identify various, potentially relevant, features of the images and determine which feature the target picture on the card shares with only one of the four possible images on the page. Whilst speech is not required, good verbal-encoding may well help the child solve the problems.

Picture Similarities scores may reflect the child's:
- Non-verbal problem solving (inductive reasoning)
- Visual perception and analysis
- Ability to attach meaning to pictures
- Ability to develop and test hypotheses
- Use of verbal mediation
- General knowledge

Low scores may also reflect impulsiveness (responding without checking the response).

3.3   Further information

For more information about the development, administration, scoring and interpretation of the BAS assessments see:

Elliott, C.D., Smith, P, and McCulloch, K (1996) British Ability Scales Second Edition (BAS II): Administration and Scoring Manual. London: NFER-Nelson.
Elliott, C.D., Smith, P, and McCulloch, K (1997) British Ability Scales Second Edition (BAS II): Technical Manual. London: NFER-Nelson.

3.4 Obtaining consent for child assessments

Before undertaking the assessments with the child, parents were required to give informed consent. A bespoke information leaflet detailing the object and content of the assessments was given to parents by the interviewer. After reading the leaflet, parents were then asked to sign a consent form permitting the assessments to go ahead. Levels of consent to undertake the assessments was very high at 97% for each assessment.

Table 3.1 Details of cognitive assessments used at sweep 5

Assessment name	Assesses	Method	Max no of items
BAS – Picture Similarities	Non-verbal reasoning	Child is shown a row of 4 pictures and is given a card with a 5th picture. The child places the card under the picture which shares an element or concept with the card.	33
BAS-Naming Vocabulary	A verbal task which concerns knowledge of names	Child is shown a picture and asked to say its name.	36

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
Cases to field at sweep 5:	
All issued to field*	4196
Eligible i.e. achievable or 'in-scope'**	4177
Cases achieved at sweep 5	3833
Response rate	
As % of all eligible cases at sweep 5	92%
As % of all sweep 1 cases	73%

* The number of cases issued to the field at sweep 5 is higher than the number of
Interviews achieved at sweep 4 because some of the sweeps 1 to 3 respondents missed
at sweep 4 came back at sweep 5.
** Cases which were considered out-of-scope or unachievable were mostly ineligible
addresses – usually due to the family having moved away from Scotland.

5 Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted. The GUS Sweep 5 Coding Instructions, deposited along with this User Guide, provide details of the tasks that were conducted.

6 Weighting the data

6.1 Overview

Two sets of weights have been developed for the birth cohort at Sweep 5:

1. A cross-sectional weight that should be used for any cross-sectional analysis of Sweep 5 data only. All sample members that responded at Sweep 5 have a cross-sectional weight.

2. A longitudinal weight for analysis of more than one wave of data. Sample members that have responded at every wave of GUS thus far have a longitudinal weight.

6.2 Background

- The sampling frame was the child-level Child Benefit records held by the Inland Revenue. Children were selected from 130 sample points in Scotland. The sample points consist of aggregations of Data Zones[2].

- There are two cohorts of children: the birth cohort and child cohort. Children in the birth cohort were aged approximately 10 months at the time of first interview whereas children in the child cohort were aged around 34 months. Weights for the birth and child cohorts have been generated separately, since these two groups are always analysed separately. The child cohort was not followed up at Sweep 5.

- The Sweep 5 interview follows up all main carers who responded at the previous interview and gave ScotCen permission to be re-contacted. In addition, some of the respondents of Sweep 1 to 3 who had asked not to take part for a year but were willing to be contacted the following year were also included at Sweep 5.

- At Sweep 5 we used proxy interviews to gather information on the main respondent's resident partner.

6.3 The sweep 5 sample

The Sweep 5 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 – Sweep 5 respondents who responded at all waves
- Sample B – Sweep 5 respondents who responded at Wave 1 but had missed an intervening wave.

The two samples will be treated separately during the weighting. This is because the Sample B respondents are likely to have different response behaviour to respondents in Sample A, as

---

[2] Further information on the sample design and the weighting process at sweeps 1, 2, 3 and 4 can be found in the User Guides for those sweeps which are available from the Data Archive or the 'using GUS data' section of the Growing Up in Scotland website www.growingupinscotland.org.uk

suggested by their much lower response rates. The birth cohort contained 362 families in Sample B, 212 (59%) of which responded at Sweep 5. The response rates for Sample B were much lower than the response rates for Sample A (95%). The issued and responding sample sizes by sample type are given in Table 6.1.

Table 6.1 Response rates for different samples

	Issued	Responding	Response rate
Baby cohort			
Sample A	3815	3621	95%
Sample B	362	212	59%
Combined (A+B)	4177	3833	92%

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 wave of data. Only members of Sample A (who have responded at every wave of GUS) will have a longitudinal weight. This weight is described in more detail in section 6.4.

The cross-sectional weight will be used for any cross-sectional analysis of Sweep 5 data. All Sweep 5 respondents will have a cross-sectional weight (Sample A + B). These are described in more detail in section 6.5.

6.4   Longitudinal weights

Longitudinal weights were only generated for respondents in Sample A. A model-based weighting technique was used to develop the Sweep 5 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 4. 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 5 interview) and a set of predictor variables. The predictor variables were a set of socio-demographic respondent and household characteristics collected from the previous sweeps.

The model generated a predicted probability for each respondent. This is the probability the respondent would take part in the interview, given the characteristics of the respondent and the household. Respondents with characteristics associated with non-response (such as being a private tenant) are under-represented in the sample and will receive a low predicted probability. 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 5 are given in Table 6.2. The full models are given in Table A1 in the Appendix.

---

[3] There were 19 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 behaviour

Characteristics associated with response	Characteristics associated with non-response
Birth cohort	
Owner occupiers	Rent from a private landlord; rent from Housing Association
At least one parent/carer in full-time employment	No parent/carer working; at least one parent/carer in part-time employment
Mother aged 25 or over	Younger mother aged under 20
Does not live in the 20% most deprived Data Zones in Scotland	Lives in the 20% most deprived Data Zones in Scotland
Interviewer made contact on the first call	Interviewer needed more than one call to make contact with the household

6.4.1 Final Sweep 5 longitudinal weights

The final Sweep 5 weight is the product of the Sweep 5 non-response weight and the Sweep 4 interview weight. The final weight was scaled to the responding Sweep 5 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 4 weights, showing the reduction in bias caused by the Sweep 5 weights.

6.5 Cross-sectional weights

Cross-sectional weights were generated for all respondents at Sweep 5 (the combined A and B samples) and should be used for any cross-sectional analysis of Sweep 5 data.

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

The pre-calibration weights were the Sweep 5 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 for the birth cohort.

6.6 Sample efficiency

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

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

Table 6.3 Range of weights and sample efficiency

	Minimum	Maximum	Mean	N	Neff	Efficiency
Birth cohort						
Longitudinal weight	0.64	2.88	1.00	3621	3221	88.9%
Cross-sectional weight	0.65	2.81	1.00	3833	3460	90.3%

6.7 Applying the weights

The cross-sectional weights should be used for any cross-sectional analysis, i.e. any analysis of Sweep 5 data only. All sample members that responded at Sweep 5 have a cross-sectional weight.

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

6.8 Weighting variables

The final interview sweep 5 weights are described in Table 6.4.

Table 6.4 Description of weight variables in the data file

Variable name	Label
DeWTbrth	De Birth cohort Sw5 weight (cross sectional sample)
DeWTbth2	De Birth cohort Sw5 weight - longitudinal

7 Using the data

The GUS Sweep 5 data consists of the following SPSS file

GUS_SW5_B.sav	3833 cases	Birth cohort

7.1 Variables on the files

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

For variables with answers following a scale, such as 'Strongly agree' to 'Strongly disagree' for instance, it must be noted that the order of the answer categories may not follow systematically an ascending or descending scale throughout the list of variables. Also the answers may equally refer to positive or negative statements as in the Strength and Difficulties questions MeSDQ01 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

Character No:				
1	**2**	**3**	**4, 5 & 6**	**7&8**
Source of data	Sweep/Wave	Key theme prefix	Sub theme stem	Question/Variable number
Non- sequential Capitals: D,M, P, S	Sequential lower case: a, b, c..	Non-sequential Capitals: C, P, N…	Abbreviated lower case: e.g. hea,	01 - 99

Source code	Details	Sweep code	Details			
AL	Area Level variable					
D	Derived variable	a	Sweep 1 (2005/06)			
DP	Derived variable from partner int	b	Sweep 2 (2006/07)			
DWP	DWP variable	c	Sweep 3 (2007/08)			
M	Main					

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

7.3 Variable labels

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

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

DeHGnmad	De - Number of adults (16 or over) in household
DeHGnmkd	De - Number of children in household
DeHGnmsb	De - Number of siblings in household
DeHGnp01	De - Number of natural parents in household
DeHGrsp01	De - Whether respondent is natural mother
DeHGrsp02	De - Whether respondent is natural father
DeHGnp02	De - Natural mother in household
DeHGnp03	De - Natural father in household
DeHGnp04	De - Respondent living with spouse/partner
DeMothID	De – Mother's ID (= Person number in household)
DeFathID	De - Father's ID
DeRespID	De – Respondent's ID

DePartID	De – Respondent's partner's ID
DeRPAge	De – Respondent's partner's age (banded)
DeRPsex	De - 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 5, 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 5 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/../DdCtya01	Sw1 / ... / Sw4 1st childcare provider type
MaCtma01/DbCtma01/../DdCtma01	Sw1 / ... / Sw4 1st childcare provider - no of hours per week
MaCdya01/DbCdya01/../DdCdya01	Sw1 / ... / Sw4 1st childcare provider - no of days per week
MaCtyb01/DbCtyb01/../DdCtyb01	Sw1 / ... / Sw4 2nd childcare provider type
MaCtmb01/DbCtmb01/../DdCtmb01	Sw1 / ... / Sw4 2nd childcare provider - no of hours per week
MaCdyb01/DbCdyb01/../DdCdyb01	Sw1 / ... / Sw4 2nd childcare provider - no of days per week
MaCtyc01/DbCtyc01/../DdCtyc01	Sw1 / ... / Sw4 3rd childcare provider type
MaCtmc01/DbCtmc01/../DdCtmc01	Sw1 / ... / Sw4 3rd childcare provider - no of hours per week
MaCdyc01/DbCdyc01/../DdCdyc01	Sw1 / ... / Sw4 3rd childcare provider - no of days per week
MaCtyd01/DbCtyd01/../DdCtyd01	Sw1 / ... / Sw4 4th childcare provider type
MaCtmd01/DbCtmd01/../DdCtmd01	Sw1 / ... / Sw4 4th childcare provider - no of hours per week
MaCdyd01/DbCdyd01/../DdCdyd01	Sw1 / ... / Sw4 4th childcare provider - no of days per week
MaCtye01/DbCtye01/../DdCtye01	Sw1 / ... / Sw4 5th childcare provider type
MaCtme01/DbCtme01/../DdCtme01	Sw1 / ... / Sw4 5th childcare provider - no of hours per week
MaCdye01/DbCdye01/../DdCdye01	Sw1 / ... / Sw4 5th childcare provider - no of days per week
MeCsta01	Me Whether still using 1st provider from last sweep
MeCcta01	Me Previous 1st ccare provider - revised hrs at Sw5
MeCcda01	Me Previous 1st ccare provider - revised days at Sw5
MeCrsa01	Me - Why not using prev provider 1 at Sw5

MeCstb01	Me Whether still using 2nd provider from last sweep
MeCctb01	Me Previous 2nd ccare provider - revised hrs at Sw5
MeCcdb01	Me Previous 2nd ccare provider - revised days at Sw5
MeCrsb01	Me - Why not using prev provider 2 at Sw5
MeCstc01	Me Whether still using 3rd provider from last sweep
MeCctc01	Me Previous 3rd ccare provider - revised hrs at Sw5
MeCcdc01	Me Previous 3rd ccare provider - revised days at Sw5
MeCrsc01	Me - Why not using prev provider 3 at Sw5
MeCstd01	Me Whether still using 4th provider from last sweep
MeCctd01	Me Previous 4th ccare provider - revised hrs at Sw5
MeCcdd01	Me Previous 4th ccare provider - revised days at Sw5
MeCrsd01	Me - Why not using prev provider 4 at Sw5
MeCste01	Me Whether still using 5th provider from last sweep
MeCcte01	Me Previous 5th ccare provider - revised hrs at Sw5
MeCcde01	Me Previous 5th ccare provider - revised days at Sw5
MeCrse01	Me - Why not using prev provider 5 at Sw5
DeCstp01	De Whether any of the previous ccare arrgmts stopped
DeCstp02	De No of previous sweep providers stopped
DeCnpv01	De No of ccare provs from last sweep still being used
DeCapv01	De 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 5. 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 5

Variable name	Description
MeCany02	Me If no ccare at last sweep whether using ccare at Sw5
MeCany03	Me If ccare at last sweep - any new prov at Sw5
MeCtya01	Me New provider 1 - type
MeCtma01	Me 1st new ccare provider - hours per week
MeCdya01	Me 1st new ccare provider - number of days per week
MeCaga01	Me Age (months) started new provider 1
MeCwya01 – MeCwya18	Me Reasons for using 1st new provider
MeCtyb01	Me New provider 2 - type
MeCtmb01	Me 2nd new ccare provider - hours per week
MeCdyb01	Me 2nd new ccare provider - number of days per week
MeCagb01	Me Age (months) started new provider 2
MeCwyb01 – MeCwyb18	Me Reasons for using 2nd new provider
MeCtyc01	Me New provider 3 - type
MeCtmc01	Me 3rd new ccare provider - hours per week
MeCdyc01	Me 3rd new ccare provider - number of days per week
MeCagc01	Me Age (months) started new provider 3
MeCwyc01 – MeCwyc18	Me Reasons for using 3rd new provider
MeCtyd01	Me New provider 4 - type
MeCtmd01	Me 4th new ccare provider - hours per week

MeCdyd01	Me 4th new ccare provider - number of days per week
MeCagd01	Me Age (months) started new provider 4
MeCwyd01 – MeCwyd18	Me Reasons for using 4th new provider
DeCnnp01	De No of new childcare arrangements at Sweep 5

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 5 interview - the provider type, number of hours and days of the arrangement, and whether or not it is a new arrangement at sweep 5. 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 5

Variable name	Description
DeCtya01	De - Childcare prov A: provider type
DeCnwa	De - Provider A: new or existing
DeCtma01	De Provider A: No of hours per week
DeCdya01	De Provider A: No of days per week
DeCtyb01	De - Childcare prov B: provider type
DeCnwb	De - Provider B: new or existing
DeCtmb01	De Provider B: No of hours per week
DeCdyb01	De Provider B: No of days per week
DeCtyc01	De - Childcare prov C: provider type
DeCnwc	De - Provider C: new or existing
DeCtmc01	De Provider C: No of hours per week
DeCdyc01	De Provider C: No of days per week
DeCtyd01	De - Childcare prov D: provider type
DeCnwd	De - Provider D: new or existing
DeCtmd01	De Provider D: No of hours per week
DeCdyd01	De Provider D: No of days per week
DeCtye01	De - Childcare prov E: provider type
DeCnwe	De - Provider E: new or existing
DeCtme01	De Provider E: No of hours per week
DeCdye01	De Provider E: No of days per week
DeCany01	De Whether resp uses regular CCare at Sw5 (not including the excluded pre-school cases – see 7.6.1)
DeCtot01	De 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 5 questionnaire which accompanies this user guide.

7.6.1 Childcare and Pre-school arrangements

Children in the birth cohort at Sweep 4 and 5, were aged between 3 and 5 years old. At this age, children in Scotland are eligible for funded pre-school places in private and education authority run nursery classes, nursery schools, and playgroups. It became clear on analysis of data from sweep 2 that a number of parents whose children were attending pre-school had not provided those pre-school details in the childcare section. At subsequent sweeps, pre-school data was collected separately and excluded from the module on childcare. As such, to obtain an accurate measure of the proportion of parents using childcare, the number of providers being used, the mix of provision and the total number of hours, it is necessary to create a number of derived variables which incorporate information from both modules.

The individual details of the pre-school place as childcare at sweep 5 have been derived for the birth cohort using the information from Sweep 5 variables *MePRyn01* 'Me Child currently attends pre-school' and *MePRty01* 'Me - Type of pre-school place'.

The final set of derived variables detailing the *current* childcare arrangements for all cases at sweep 5, including the pre-school omissions, are listed in Table 7.6.

Table 7.6 Childcare variables incorporating information on pre-school arrangements

Variable name	Description
DeCany02	Whether or not using childcare (including pre-school arrangements)
DeCtot02	Number of childcare providers being used at sw5 (including pre-school arrgts)
DeCPrSpv	Sw5 Pre-School provider type for those who did not provide pre-school details in childcare section
DeCPrSHr	No of hours looked after per week by the missed Pre-School childcare provider
DeCPrSDy	No of days looked after per week by the missed Pre-School childcare provider
DeCtmi01	No of hrs child looked after by someone else (average week)
DeCtmi02	No of hrs child looked after by someone else in an average week - BANDED
DeCday01	Highest number of days per week in any one childcare arrangement
DeCtyp01	Does respondent use grandparents for childcare?
DeCtyp02	Does respondent use another relative for childcare?
DeCtyp03	Does respondent use private creche/nursery for childcare?
DeCtyp04	Does respondent use a childminder for childcare?
DeCtyp05	Does respondent use a local authority playgroup for childcare?
DeCtyp06	Does respondent use a local authority nursery for childcare?
DeCtyp07	Does respondent use a private playgroup for childcare?
DeCtyp08	Does respondent use a community/voluntary playgroup for childcare?
DeCtyp09	Does respondent use an ex-spouse or partner for childcare?
DeCtyp10	Does respondent use the childs older sibling for childcare?
DeCtyp11	Does respondent use a friend or neighbour for childcare?
DeCtyp12	Does respondent use a daily visiting nanny for childcare?
DeCtyp13	Does respondent use a live-in nanny for childcare?
DeCtyp14	Does respondent use a babysitter for childcare?
DeCtyp15	Does respondent use a workplace creche or nursery for childcare?
DeCtyp16	Does respondent use a family centre for childcare?
DeCtyp17	Does respondent use a nursery class attached to a primary school for childcare?

DeCtyp18	Does respondent use an agency carer?
DeCtyp19	Does respondent use another type of childcare provider for childcare?
DeCtyp20	Does respondent currently use OTHER INFORMAL childcare?
DeCtyp21	Does respondent currently use NURSERY OR CRECHE for childcare?
DeCtyp22	Does respondent currently use PLAYGROUP for childcare?
DeCtyp33	Does respondent currently use CHILDREN S CLUB for childcare?
DeCtyp23	Does respondent currently use OTHER PROVIDERS for childcare?
DeCtyp30	Does respondent currently use informal childcare?
DeCtyp31	Does respondent currently use formal childcare?
DeCtyp32	Current use of formal and informal childcare

7.7   Indicators and summary variables

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

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

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

Further information on NS-SEC is available from the National Statistics website at:
http://www.ons.gov.uk/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 (DeEqvinc) and quintiles of the scale (DeEqv5) 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 5 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 'ALeHBdBc' 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 Cognitive Assessments: British Ability Scales – Naming Vocabulary and Picture Similarities

Score variables in the dataset

The dataset provides the following scores for each assessment:

Raw score – the number of correct responses amongst the items administered. Note that because children take different sets of items, their raw scores cannot be compared directly.

Ability Score – Derived from the raw score and the item set administered (using tables provided on the BAS Naming Vocabulary and Picture Similarities Score Sheets) this is an estimate of child's level on the ability being measured. It reflects the raw score and the difficulty of the items administered. The ability score is not a normative score. The numbers used are arbitrary and simply provide a common scale of performance level, regardless of the items a child was given. Note that DePSAbSc/DeNVAbSc variables must be used for comparisons with Birth Cohort 2 (see below).

Normative scores – Derived from standard BAS tables and defined with reference to the standardisation sample used in developing the assessment. Both T-scores (with mean=50 and standard deviation=10) and their equivalent Percentiles are provided. Note that DePSTSc/DeNVTSc variables must be used for comparisons with Birth Cohort 2 (see below).

The sweep 5 GUS report *Changes in child cognitive ability in the pre-school years* (Bradshaw, 2011) used within-GUS z-scores of the sweep 5 ability score rather than the BAS-derived T-scores. Z-scores are easily derived in SPSS using the 'Descriptives' command but have also been included in the data file.

Variable names and labels are detailed in table 7.10.

Table 7.10   BAS Summary Score Variables

Variable name	Label
DePicRaw	De Picture Similarities Raw Score
DePicSAS	De Picture Similarities Ability Score
ZDePicSAS	De Picture Similarities Z-Score
DePicSTS	De Picture Similarities T-Score
DePicSPt	De Picture Similarities Percentile Equivalent

DeNamRaw	De Naming Vocabulary Raw Score
DeNamVAS	De Naming Vocabulary Ability Score
ZDeNamVAS	De Naming Vocabulary Z-Score
DeNamVTS	De Naming Vocabulary T-Score
DeNamVPt	De Naming Vocabulary Percentile Equivalent
DePSAbSc	De Picture Similarities Ability Score (for comparing with BC2)
DePSTSc	De Picture Similarities T-Score (for comparing with BC2)
DeNVAbSc	De Naming Vocabulary Ability Score (for comparing with BC2)
DeNVTSc	De Naming Vocabulary T-Score (for comparing with BC2)

Comparison with BC2

Birth Cohort 2 children carried out Naming Vocabulary and Picture Similarities exercises when they were the same age (58 months). However, different editions of the assessments were used: For BC1, the 2nd edition assessment was used (BAS-II), whereas for BC2 the 3rd edition was used (BAS3). Whilst the assessments are almost identical, there are a small number of differences – for example in the individual items, the order of the items and the stopping points – which would introduce caveats when making a straightforward comparison of ability scores.

To allow for comparison between the cohorts, the assessment authors provided the GUS team with a calibration formula to be applied to the original BC1 scores. Once applied, the revised scores can be used in comparisons between the cohorts. No adjustment is required to the BC2 scores.

Note that any comparisons of cognitive ability scores across the two cohorts <u>MUST</u> use the adjusted BC1 scores (DePSAbSc, DePSTSc, DeNVAbSc, DeNVTSc).

Note also that because of the adjustments, it is not possible to convert differences in average cognitive ability scores to developmental age in months when using the adjusted scores.

General influences on test scores

It is important to note that the child's performance may have been affected by influences extraneous to those that the assessment is intended to measure. The conditions listed below can lead either to a higher or lower score than would normally be obtained.

- Non-standard administration of the scale
- Administration disrupted by noise or other interruptions
- Difficulty in establishing rapport with the child
- Child has difficulty in concentrating on the tasks or is easily distracted
- Child is excessively anxious to the extent that concentration/flexibility of thought seem impaired
- Child is reluctant to respond and/or refuses to persevere on more difficult items
- Child has permanent/temporary sensory impairment (particularly vision/hearing) or motor impairment
- Child is on medication of a type that could affect performance
- Child is over-tired or ill.

In anticipation of these issues, the specification of CAPI program and the training of interviewers were designed to ensure standard administration of the assessment. The training of interviewers

was also designed to ensure that risks were minimised. To allow for the consideration of such issues when analyzing the data, interviewers were asked to record details of any interruptions, distractions, behaviours or health circumstances in CAPI. Table 7.11 contains details of the relevant variables where this information is recorded.

Table 7.11   Variables recording difficulties experienced during assessments

Variable name	Label
CAePrb01	CAe - No difficulties experienced during assessments
CAePrb02	CAe - Difficulties experienced because: Assessment was interrupted
CAePrb03	CAe - Difficulties experienced because: Child was ill
CAePrb04	CAe - Difficulties experienced because: Child was tired
CAePrb05	CAe - Difficulties experienced because: Parent interfered
CAePrb06	CAe - Difficulties experienced because: Something else

7.7.8   Child Development: Strengths and Difficulties Questionnaire

The Strengths and Difficulties Questionnaire (SDQ) is a brief behavioural screening questionnaire designed for use with 3-16 year olds (Goodman, 1997). The scale includes 25 questions which are used to measure five aspects of the child's development – emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and pro-social behaviour. A score is calculated for each aspect, as well as an overall 'difficulties' score which is generated by summing the scores from all the scales except pro-social. For all scales, except pro-social where the reverse is true, a higher score indicates greater evidence of difficulties. The dataset includes the constituent items, and the derived variables including the various composite scores and total score. Details of these variables are included in Table 7.12 with syntax illustrated in the derived variables documentation.

Table 7.12   Derived variables associated with the Strengths and Difficulties Questionnaire

Variable name	Description
DeDsdem1	De SDQ: Emotional symptoms score
DeDsdco1	De SDQ: Conduct problems score
DeDsdhy1	De SDQ: Hyper-activity or inattention score
DeDsdpr1	De SDQ: Peer problems score
DeDsdps1	De SDQ: Pro-social score
DeDsdto1	De SDQ: Total difficulties score

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

7.7.9   Parental Health: Medical Outcomes Study 12-Item Short Form (SF-12)

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

Table 7.13 Constituent and derived variables associated with the SF-12

Variable name	Description
MeHpgn01	Me - How is resp health in general
MeHlmt01	Me - Resp health limits moderate activities
MeHlmt02	Me - Resp health limits climbing stairs
MeHlmt03	Me - Resp health limited accomplishments past 4 wks
MeHlmt04	Me - Resp health limited reg activities past 4 wks
MeHlmt05	Me - Resp mental health limited accomplishments past 4 wks
MeHlmt06	Me - Resp mental health limited quality of accomplishments past 4 wks
MeHlmt07	Me - Resp physical pain limited normal work past 4 wks
MeHpgn02	Me - Time resp felt calm in past 4 wks
MeHpgn03	Me - Time resp felt energetic in past 4 wks
MeHpgn04	Me - Time resp felt down in past 4 wks
MeHpgn05	Me - Time resp health interfered socially in past 4 wks
DeSF12ph	De - Physical PCS - 12 Scale
DeSF12mn	De - Mental MCS - 12 Scale

7.7.10 Parenting: Home chaos

The sweep 5 questionnaire included a subset of four questions from the 15-item Confusion, Hubbub, and Order Scale (CHAOS), an instrument specifically designed to be administered to parents for assessing turmoil in the child's home (Matheny et al, 1995). CHAOS is used to assess a child's home life and the GUS items ask parents how strongly they agree/disagree with questions about disorganisation, noise, having a calm atmosphere, and having a regular routine at home.

US research has shown household chaos to be associated with behaviour problems, inattention and cognitive development problems in children (Deater-Deckard et al, 2009; Dumas et al, 2006).

Table 7.14 Constituent variables associated with home chaos

Variable name	Description
MePcha01	*It's really disorganised in our home*
MePcha02	*You can't hear yourself think in our home*
MePcha03	*The atmosphere in our home is calm*
MePcha04	*First thing in the day, we have a regular routine at home*

7.7.11 Parenting: Pianta Child-Parent Relationship Scale

The Pianta scale (Pianta, 1992) is used to measure the mother-child relationship at year 5. The scale is constructed using the responses on the extent to which the respondent feels a series of statements apply to her relationship with her child (such as *'I share an affectionate, warm relationship with [my child]'*).

The full scale has 30 items and looks at 3 dimensions of the relationship – warmth, conflict and dependency. The 15 items included in the sweep 5 GUS questionnaire are a subset of the full scale that were also used in the Millennium Cohort Study (MCS2; 2004/05) and which relate to warmth and conflict. Measures can be constructed for these two dimensions, with a high score corresponding to a high degree of warmth or conflict. Each measure uses seven items, shown below.

Table 7.15 Constituent and derived variables associated with the Pianta Child-Parent Relationship Scale

Variable name	Description
Warmth	
MePpia01	*I share an affectionate, warm relationship with [Child's name]*
MePpia03	*[Child's name] will seek comfort from me*
MePpia05	*[Child's name] values his/her relationship with me*
MePpia06	*When I praise [Child's name], he/she beams with pride*
MePpia07	*[Child's name] spontaneously shares information about [him/herself]*
MePpia09	*It is easy to be in tune with what [Child's name] is feeling*
MePpia15	*[Child's name] openly shares his/her feelings and experiences with me*
Conflict	
MePpia02	*[Child's name] and I always seem to be struggling with each other*
MePpia08	*[Child's name] easily becomes angry at me*
MePpia10	*[Child's name] remains angry or is resistant after being disciplined*
MePpia11	*Dealing with [Child's name] drains my energy*
MePpia12	*When [Child's name] wakes up in a bad mood, I know we're in for a long and difficult day*
MePpia13	*[Child's name]'s feelings towards me can be unpredictable or can change suddenly*
MePpia14	*[Child's name] is sneaky or manipulative with me*

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 as described in section 7.7.6.

7.9 Missing values conventions

-1   Not applicable: Used to signify that a particular variable did not apply to a given respondent, usually because of internal routing.

-8   Don't know, Can't say.

-9   No answer/ Refused

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

8 Documentation

The documentation has been organised into the following sections:

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

9 References

Bradshaw, P. (2011) ***Growing Up in Scotland: Changes in child cognitive ability in the pre-school years,*** Edinburgh: Scottish Government

Deater-Deckard, K., Mullineaux,P.Y., Beekman, C., Petrill, S.A., Schatschneider,S. and Thompson, L.A. (2009) "Conduct problems, IQ, and household chaos:
a longitudinal multi-informant study". *Journal of Child Psychology and Psychiatry* 50:10, pp 1301–1308

Dumas, J.E., Nissley, J. , Nordstrom, A., Phillips Smith, E., Prinz, R.J. and Levine, D.W. (2006) "Home Chaos: Sociodemographic, Parenting, Interactional, and Child
Correlates." *Journal of Clinical Child & Adolescent Psychology*, 34: 1, pp. 93 - 104

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

Matheny, A. P., Jr., Wachs, T.d., Ludwig, J.L., and Phillips, K. (1995). "Bringing Order Out of Chaos: Psychometric Characteristics of the Confusion, Hubbub, and Order Scale." *Journal of Applied Developmental Psychology*, 16, pp. 429-444.

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

10 Related publications

Further information about GUS Sweep 5 is available in:
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

Other publications which include the use of GUS data include:

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/

11 Contact details

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

GUS Project Manager	Paul Bradshaw	0131 221 2564	paul.bradshaw@scotcen.org.uk
GUS Senior Researcher	Judith Mabelis	0131 221 2560	Judith.Mabelis@scotcen.org.uk
GUS Data Manager	Mireille Ferrandon	0131 221 2570	mireille.ferrandon@scotcen.org.uk

Appendix A: Full non-response models

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

	B	S.E.	Wald	df	Sig.	Exp(B)
Tenure			6.6	2	0.037	
Owner occupiers					baseline	
Rents HA/council	-0.45	0.196	5.2	1	0.023	0.64
Rents private	-0.52	0.246	4.5	1	0.034	0.59
Mother's age at birth (grouped)			15.9	4	0.003	
<20					baseline	
20-24	-0.09	0.229	0.2	1	0.682	0.91
25-29	0.28	0.246	1.3	1	0.259	1.32
30-34	0.68	0.267	6.4	1	0.011	1.97
35+	0.65	0.296	4.9	1	0.027	1.92
Household employment			12.0	2	0.002	
At least one parent/carer in full-time employment					baseline	
At least one parent/carer in part-time employment	-0.59	0.183	10.3	1	0.001	0.56
Household employment	-0.03	0.211	0.0	1	0.877	0.97
Scottish Index of Multiple Deprivation 2009-quintiles			10.6	4	0.031	
0.63 - 7.75 (least deprived)					baseline	
7.76 - 13.76	0.49	0.199	5.9	1	0.015	1.62
13.77 - 21.02	0.62	0.236	6.9	1	0.009	1.86
21.03 - 33.72	0.18	0.222	0.7	1	0.406	1.20
33.73 -90.05 (most deprived)	0.46	0.273	2.8	1	0.092	1.58
Total number of calls			8.5	4	0.074	
1					baseline	
2	-0.16	0.194	0.7	1	0.399	0.85
3	-0.01	0.234	0.0	1	0.957	0.99
4	-0.44	0.254	3.0	1	0.082	0.64
5+	-0.53	0.216	5.9	1	0.015	0.59
Constant	2.67	0.311	73.7	1	0.000	14.45

Notes:
1. The response is 1 = sample A response to wave 4, 0 = sample A non-response.
2. Model is weighted by wave 2 baby weight
3. The model R^2 = 0.03 (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

	Birth cohort		
	Sweep 4 weighted by Sweep 4 weight	Sweep 5 weighted by Sweep 4 weight	Sweep 5 weighted by Sweep 5 weight
	%	%	%
Tenure			
Owner occupier	63.0	64.5	63.1
Rents HA/council	28.1	26.9	28.1
Rents private	8.9	8.6	8.8
Family status			
Lone parent	18.7	17.7	18.5
Couple parent	81.3	82.3	81.5
Mother's age at birth			
<20	7.6	7.1	7.5
20-24	17.2	16.3	17.1
25-29	23.5	23.5	23.5
30-34	31.3	32.1	31.4
35+	20.5	21.0	20.5
Highest education level of respondent			
Degree or equivalent	27.8	28.4	27.7
Vocational qualification below degree	39.0	39.1	39.2
Higher Grade or equivalent	7.4	7.4	7.4
Standard Grade or equivalent	17.2	16.5	16.8
No Qualifications	8.6	8.6	8.8
Household income			
<£10,000	10.3	9.8	10.3
£10,000-£19,999	20.0	19.6	20.2
£20,000-£31,999	22.9	23.0	23.0
£32,000+	41.5	42.4	41.4
Missing	5.2	5.1	5.2
Respondent NSSEC - 5 Category			
Managerial and professional occupations	34.4	35.2	34.5
Intermediate occupations	19.8	19.9	19.9
Small employers and own account workers	5.7	5.8	5.7
Lower supervisory and technical occupations	5.8	5.7	5.8
Semi-routine and routine occupations	30.3	29.5	30.1
Missing/never worked	3.9	3.8	4.0
Ethnicity of respondent			
White	96.2	96.4	96.4
Other ethnic background	3.8	3.6	3.6

Table A2 Distribution of sample A (continued)

	Birth cohort		
	Sweep 4 weighted by Sweep 4 weight	**Sweep 5 weighted by Sweep 4 weight**	**Sweep 5 weighted by Sweep 5 weight**
	%	%	%
Household employment			
At least one parent/carer in full-time employment	72.6	73.9	72.7
At least one parent/carer in part-time employment	13.6	12.8	13.5
No parent/carer working	13.8	13.3	13.8
Mother's employment status			
Childs mother working - full-time	15.7	16.0	15.7
Childs mother working - part-time	49.6	49.7	49.6
Childs mother not working	34.6	34.3	34.7
Number of children in the household			
1	24.4	24.1	24.3
2	49.5	49.9	49.6
3	18.8	18.8	18.7
4+	7.3	7.3	7.3
Urban/rural indicator (Scotland)			
1 Large urban area (125,000+)	37.1	36.2	36.4
2 Other urban area (10,000-125,000)	32.5	33.3	33.5
3 Accessible small town (3,000-10,000)	10.3	9.9	9.9
4 Remote small town (3,000-10,000)	1.0	1.0	1.0
5 Very remote small town (3,000-10,000)	1.8	1.7	1.7
6 Accessible rural (<3,000)	11.9	12.2	12.0
7 Remote rural (<3,000)	3.0	3.0	2.9
8 Very remote rural (<3,000)	2.4	2.7	2.6
Use regular childcare			
Yes	63.5	63.7	63.5
No	36.5	36.3	36.5
Scottish Index of Multiple Deprivation 2009 - quintiles			
0.63 - 7.75 (least deprived)	18.8	19.3	18.8
7.76 - 13.76	20.0	20.2	20.1
13.77 - 21.02	18.8	19.1	18.7
21.03 - 33.72	19.4	19.5	19.5
33.73 -90.05 (most deprived)	23.0	21.8	22.9
SIMD09 Flag lowest 15%			
Less deprived 85% Data Zones	82.6	83.5	82.7
Most deprived 15% Data Zones	17.4	16.5	17.3
Base (unweighted)	3844	3621	3621

Table A3 Birth cohort - weighted distribution of key variables for samples A and B

	Sample A Weighted by Sweep 5 weight	Combined Sweep 5 sample (A+B)	
		Weighted by pre-calibration weight[1]	Calibrated to sample A
Family type			
Lone parent	19.3	20.0	19.3
Couple family	80.7	80.0	80.7
Household income			
<£10,000	10.2	10.6	10.2
£10,000-£19,999	19.7	19.9	19.7
£20,000-£31,999	22.3	22.2	22.3
£32,000+	41.8	41.1	41.8
Missing	6.0	6.2	6.0
Respondent's age			
<25	7.8	8.1	7.8
25-29	17.0	17.3	17.0
30-34	24.1	24.1	24.1
35-39	30.5	30.0	30.5
40+	20.7	20.6	20.7
Respondent employment status			
Employed	64.6	64.2	64.6
Not employed	35.4	35.8	35.4
Scottish Index of Multiple Deprivation 2009 - quintiles			
0.63 - 7.75 (least deprived)	18.8	18.5	18.8
7.76 - 13.76	20.1	19.8	20.1
13.77 - 21.02	18.7	18.6	18.7
21.03 - 33.72	19.5	19.7	19.5
33.73 -90.05 (most deprived)	22.9	23.6	22.9
Base (unweighted)	3621	3833	3833

[1]This is the Sweep 5 weight for Sample A and the weight from the last completed wave for Sample B

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

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

- In the Pre-School section, dates have been entered incorrectly as 8200 and 5082 in CAPI for two cases at variable MePRwn01 (in the dataset as Derived Variable DePRwnY 'Year child started pre-school'). In addition an entry in the future was entered for case IDnumber 1001535 (Aug-10, so after fieldwork for Sweep 5 ended).

- In the Pre-School and Childcare sections, sometimes the respondent answered that they paid for pre-school or childcare, but gave the actual amount as £0. In the cost derived variables DeCcst01 to DeCcst04 these cases have been put together with those who answered 'free' or 'someone else pays for it' at the original questions.

- In the Childcare section, cases IDnumber 1000836 and 1005725 show a total number of hours of childcare above 168 hours (= 24 hrs x 7) per week at the Derived Variable DeCtmi01 'Number of hrs child looked after by someone else in an average week'.

- Case IDnumber 1001994 has an extreme value of £9,125 at MeCpay11 'Amount for other children's childcare per week', which impacts on the derived variable for the cost of all childcare DeCcst03.

- 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 5 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, as happened with case IDnumber 1005455: 4 illnesses recorded, hence additional variable MeHlsd01, but no corresponding set of answers regarding whether it limits activities or the type of treatment/advice received.

- In the Activities section, the two cases IDnumber 1005963 and 1002416 showed some inconsistencies in the answers questions about watching TV: answered 0 at MeAtv01 'Days child watched TV last week' and No at MeAtv24 'Does child ever watch TV (incl DVDs)' but said 'every day' or 'occasionally' at MeAtv35 'Me - Child watches TV channels with adverts'.

- Although the Self-complete section was asked to all respondents, 26 respondents chose not to complete it and these cases show as missing values ('Not Applicable') in the dataset.

- 14 respondents who said they currently smoked at variable MeHcig02 have answered '0' at MeHcig03 'How many cigarettes per day', presumably because less than one per day on average.

- In the Employment section, case IDnumber 1006568 gave a date in the future (Dec-10 at DeYendM 'Month partner's job ended' and DeYendY 'Year partner's job ended').

- The Tenure derived variables for previous sweeps DaZten02 to DdZten02 included in the Sweep 5 dataset have been calculated at the current sweep (Sweep 5), based on the original raw variables imported from the previous sweeps into the Sweep 5 dataset. The syntax used to derive these variables from Sweep 2 to Sweep 4 in the current sweep replaces all missing values, including those missing because the sweep was skipped, with the last available valid value from a previous sweep (see full details in the 'Derived Variables syntax' document). For example 94 cases in Sweep 5 skipped Sweep 4 (coded 2='No' at variable Sweep4), and therefore show as 'system missing' at the original raw variable MdZhou05 imported from Sweep 4. However derived variable DdZten01 (and its recoded version DdZten02) has been built first on MdZhou05, then all cases showing as missing at Sweep 5 have been filled gradually with Sweep 3, then Sweep 2 then Sweep 1 values, so there are no system missings anymore in the derived variable. To distinguish the 94 cases which were skipped at Sweep 4, DdZten01 and DdZten02 would have to be recoded as -1 'Not Applicable' using the Sweep4 var:

  DO IF (Sweep4=2).
  COMPUTE DdZten01=-1.
  COMPUTE DdZten02=-1.
  END IF.
  Exe.

  The same remark applies for the Sweep 2 & Sweep 3 derived variables DbZten01/ DbZten02 and DcZten01/DcZten02.

- Four cases had a partial interview (code 210 at variable MeOutcom), 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.