



Growing Up In Scotland

Sweep 2: 2006-2007

User Guide

Authors: Paul Bradshaw, Sarah Tipping, Louise Marryat, Joan Corbett

- 1 Overview of the survey 1**
 - 1.1 Study Design 1
 - 1.2 Sample Design 2
 - 1.3 Development and Piloting..... 3
- 2 Data collection methods 3**
 - 2.1 Mode of data collection..... 3
 - 2.2 Length of Interview 4
 - 2.3 Timing of fieldwork..... 4
 - 2.4 Partner interviews..... 4
- 3 Response rates 4**
- 4 Coding and editing 6**
- 5 Weighting the data..... 6**
 - 5.1 Background 6
 - 5.2 Main interview..... 6
 - 5.2.1 Weighting method 6
 - 5.2.2 Final sweep 2 weights 8
 - 5.3 Partner weights..... 8
 - 5.3.1 Weighting method 8
 - 5.3.2 Final partner weights 9
 - 5.4 Sample efficiency 9
 - 5.5 Applying the weights..... 10
 - 5.5.1 Main interview weights 10
 - 5.5.2 Partner weights 10
- 6 Using the data 10**
 - 6.1 Variables on the files 10
 - 6.2 Variable naming convention 11
 - 6.3 Variable labels 11
 - 6.4 Derived variables..... 11
 - 6.5 Household data 13
 - 6.6 Childcare data 13
 - 6.6.1 Childcare and Pre-school arrangements..... 16
 - 6.7 Indicators and summary variables..... 18
 - 6.7.1 Socio-economic Characteristics: National Statistics Socio-economic Classification (NS-SEC)18
 - 6.7.2 Area-level variables: Scottish Government Urban/Rural Classification 18

6.7.3	Area-level variables: Scottish Index of Multiple Deprivation	19
6.7.4	Child Development: Communication and Symbolic Behaviour Scale – Infant/Toddler Checklist	19
6.7.5	Child Development: Strengths and Difficulties Questionnaire.....	20
6.7.6	Parental Health: Depression, Anxiety and Stress Scale	21
6.8	Dropped Variables.....	21
6.9	Weighting variables	22
6.10	Missing values conventions	23
7	Documentation.....	23
8	Related publications.....	23
9	Contact details	24
10	References.....	24
	Appendix A: Full non-response models.....	25

1 Overview of the survey

The data files contain data from Growing Up in Scotland (GUS) Sweep 2, the second year of a longitudinal research study aimed at tracking the lives of a cohort 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.

1.1 Study Design

The survey is 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 named sample of approximately 10,700 children was selected from the Child Benefit records to give an achieved sample of 8,000 overall.

The configuration of cohorts and sweeps for the first four 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').

Table 1.1 Proposed sample design, 2005-2011

Year	Age at interview					
	0-1	1-2	2-3	3-4	4-5	5-6
2005	BC1		CC1			
2006		BC1		CC1		
2007			BC1		CC1	
2008				BC1		CC1

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 2-3 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 to release data from the Census 2001 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 Authority¹ and then by the Scottish Index of Multiple Deprivation Score. 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.

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.

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	Child Cohort
1	01-June-2004 - 30-Jun-2004	01-June-2002 - 30-Jun-2002
2	01-Jul-2004 - 31-Jul-2004	01-Jul-2002 - 31-Jul-2002
3	01-Aug-2004 - 31-Aug-2004	01-Aug-2002 - 31-Aug-2002
4	01-Sep-2004 - 30-Sep-2004	01-Sep-2002 - 30-Sep-2002
5	01-Oct-2004 - 31-Oct-2004	01-Oct-2002 - 31-Oct-2002
6	01-Nov-2004 - 30-Nov-2004	01-Nov-2002 - 30-Nov-2002
7	01-Dec-2004 - 31-Dec-2004	01-Dec-2002 - 31-Dec-2002
8	01-Jan-2005 - 31-Jan-2005	01-Jan-2003 - 31-Jan-2003
9	01-Feb-2005 - 28-Feb-2005	01-Feb-2003 - 28-Feb-2003
10	01-Mar-2005 - 31 Mar-2005	01-Mar-2003 - 31 Mar-2003
11	01-Apr-2005 - 30-Apr-2005	01-Apr-2003 - 30-Apr-2003
12	01-May-2005 - 31-May-2005	01-May-2003 - 31-May-2003

1.3 Development and Piloting

Policy priorities and key topics of interest for the sweep 2 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 in reference to these priorities and topics. A subset of new questions was included in a small cognitive pilot in September 2005, with a full instrument initially piloted in paper form in November 2005. This instrument was revised and converted into CAPI for the second Dress Rehearsal Pilot in January 2006.

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 sweep 2, interviewers were instructed to undertake the interview with the sweep 1 respondent. Where this was not possible or appropriate, interviews were conducted with the child's main carer. In practice, most interviews were undertaken with the sweep 1 respondent and this was usually the child's mother.

2.2 Length of Interview

Overall, the average interview lasted around 79 minutes. The child cohort interview had a slightly longer average length at 82 minutes, than the birth cohort interview at 78 minutes. The median interview length for both cohorts was 70 minutes.

2.3 Timing of fieldwork

Fieldwork was undertaken over a fourteen month period commencing in April 2006. 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 2006 and remained in field until mid-July 2006.

To ensure that respondents in both samples were interviewed when their children were approximately the same age, each case was assigned a 'target interview date'. For the birth cohort this was identified as the date on which the child turned 22.5 months old, and for the child cohort the date the child turned 46.5 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.

2.4 Partner interviews

As well as the main interview, at sweep 2, CAPI interviews were also undertaken with the resident partner of the main respondent. Given that in the vast majority of cases the main respondent was the child's natural mother, most of the partner interviews (97%) were conducted with the child's natural father. The partner's interview was shorter than, and used a selection of questions from, the main interview. A total of 2,975 partner interviews were successfully completed in the birth cohort and 1,541 in the child cohort. These figures represent response rates of 79% and 77% respectively.

3 Response rates

Details of the number of cases issued and achieved and the response rates are detailed in Table 3.1.

Table 3.1 Number of issued and achieved cases and response rates

	Birth	Child	All Sample
All eligible children (No. of sweep 1 achieved interviews)	5217	2858	8075
Cases to field:			
All	5217	2858	8075
Achievable or 'in-scope' [*]	5158	2822	7980
Cases achieved	4512	2500	7012
Response rate			
As % of all sweep 1 cases	87%	88%	87%
As % of all 'in-scope'	88%	89%	88%

^{*}Cases which were considered out-of-scope or unachievable were mostly ineligible addresses – usually due to the family having moved away from Scotland.

4 Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted. The GUS Sweep 2 Coding Instructions provide details of the tasks that were conducted.

5 Weighting the data

5.1 Background

- The sampling frame was the child-level Child Benefit records held by the Inland Revenue. Children were selected from 120 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 should always be analysed separately.
- The Sweep 2 interview follows up all mothers who responded at the first interview and gave ScotCen permission to be re-contacted. There was no sub-sampling. Response rates were good at 87% for the birth cohort and 88% for the child cohort.
- At Sweep 2 we also carried out interviews of any resident partners of the main respondent, proxy interviews were not permitted. Response rates for the partner interviews were 79% of all couple households in the birth cohort and 77% for the child cohort.

5.2 Main interview

5.2.1 Weighting method

Unlike the sweep 1 weights, a model-based weighting technique was used at sweep 2. All cases which were issued at sweep 2 were respondents who had taken part in the sweep 1 interview. Information on the sweep 2 non-respondents taken from their sweep 1 interview could be used to model their response behaviour at sweep 2. Ineligible households (deadwood) were not included in the non-response modelling³.

Non-response behaviour was modelled using logistic regression. This is a method of analysing the relationship between an outcome variable (in this case response to the sweep 2 interview) using a set of predictor variables. The model takes account of the relationship of the predictor variables to the outcome and the relationships of the predictor variables to each other.

To speed up the modelling process a bivariate analysis was carried out prior to the modelling to identify variables that were related to response behaviour. The variables included in the shortlist are listed in table 5.1.

² Further information on the sample design and the weighting process at sweep 1 can be found in the Sweep 1 User Guide which is available from the Data Archive (SN 5760) or the 'using GUS data' section of the Growing Up in Scotland website www.growingupinScotland.org.uk

³ There were 45 individuals with ineligible outcome codes; these individuals were dropped from the

Table 5.1 Variables used in non-response model

Variable Name	Description
DaHGnp04	Whether respondent is living with spouse/partner
DaHGprim	Whether cohort child was mother's first born (amongst children in household)
MaWinc09	Household income
DaMedu03	Highest Education level of Mother
DaEthGpM	Ethnicity of Respondent
DaMsta01	Respondents employment status
DaMsta10	Household employment status
MaZhou05	Tenure
DaHGnmkd	Number of children in household
DaHGag2	Respondent's age at time of interview
DaHGmr2	Respondent's marital status
DaURind1	ONS Urban/Rural indicator - Scotland
DaADsco02	Scottish Index of Multiple Deprivation (data zone level) 2004 - Quintiles
MaCany01	Respondent regularly uses childcare
DaHacc01	Whether child had had an accident or injury for which they were taken to a doctor, health centre or hospital
MaHgen01	Selected child's general health
MaObtg01	Respondent regularly attended any baby/toddler groups in the past year
MaBFDe01	Selected child was breastfed
MaHdev01	Respondent has concerns about selected child's development, learning or behaviour
MaHpgn01	Respondent's general health
MaHcig02	Respondent currently smokes cigarettes

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

The birth and child cohorts were modelled separately, although there were similarities between the two models. The characteristics related to response behaviour at sweep 2 are given in Table 5.2 for the birth cohort and Table 5.3 for the child cohort. The full models are given in the Appendix.

Table 5.2 Characteristics associated with response behaviour in the birth cohort

Characteristics associated with response	Characteristics associated with non-response
High income (>£32,000)	Lower income or withheld information on income
Owner occupiers	Rent from a private landlord
Live as a couple	Lone parent family
From a white ethnic background	From any other ethnic background
Breastfed the baby	Did not breast feed
Area deprivation score falls in the middle quintile	Area deprivation score is in the lowest quintile
Attended baby groups with the selected child	Did not attend baby groups

Table 5.3 Characteristics associated with response behaviour in the child cohort

Characteristics associated with response	Characteristics associated with non-response
Higher income (>£32,000)	Lower income or withheld information on income
Owner occupiers	Rent from a private landlord
Live as a couple	Lone parent family
Used childcare regularly	Did not use childcare
Respondent suffer poor health	Respondent has excellent general health
Household is in remote or very remote town	Household is in large urban area

5.2.2 Final sweep 2 weights

The final sweep 2 weight is the product of the sweep 2 non-response weight and the sweep 1 interview weight. The final weights were scaled to the responding sweep 2 sample size to give a mean weight of one. This makes the weighted sample size match the unweighted sample size.

Details of the weight variables are contained in section 6.9. Information on when to apply the weights is contained in section 5.5.

5.3 Partner weights

5.3.1 Weighting method

Partner interviews were carried out at sweep 2 of the survey. Partner interviews were attempted in any household with live-in partners. Partners were not interviewed at sweeps 1, 3 or 4. Whilst the response rate of the partners was good - 79% in the birth cohort and 77% in the child cohort - there could still be some bias if the partners who responded were systematically different from those that did not. A bivariate analysis suggested the partner sample was biased and a set of weights was generated to reduce the effects of this.

The methods used are the same as those used to generate the main sweep 2 non-response weights. The difference was that information from the respondents' sweep 2 interview could be used to model the response behaviour of the partners. Again, the data for the birth and child cohorts were modelled separately. However, the patterns in response behaviour were very similar and a number of variables appeared in both models. The characteristics related to response behaviour identified by modelling partner response are given in Table 5.4 for the birth cohort and Table 5.5 for the child cohort. The models are given in full in the Appendix.

Partners in both cohorts were less likely to respond if the mother was not working, *unless the partner was also not working*. However, couples in households where neither partner worked more than 16 hours per week were more likely to respond.

Table 5.4 Characteristics associated with partner response behaviour in the birth cohort

Characteristics associated with response	Characteristics associated with non-response
Mother older (40+)	Younger mother (<30)
Educated to Higher or above	No qualifications
First time mother	Other children in household
Child white ethnic background	Child other ethnic background
Mother works full-time	Mother does not work
Neither parent in work	Both parents work 16+ hours a week

Table 5.5 Characteristics associated with partner response behaviour in the child cohort

Characteristics associated with response	Characteristics associated with non-response
Household is in remote or very remote town	Household is in large urban area, or household is in accessible small town
First time mother	Other children in household
Child white ethnic background	Child other ethnic background
Mother works full-time	Mother does not work
Neither parent in work	Both parents work 16+ hours a week

5.3.2 Final partner weights

The final partner weight is the product of the partner non-response weight and the sweep 2 interview weight. The weights were scaled to the responding sample size to give a mean weight of one. This makes the weighted sample size match the unweighted sample size.

Details of the weight variables are contained in section 6.9. Information on when to apply the weights is contained in section 5.5.

5.4 Sample efficiency

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

The effect of the sample design on the precision of survey estimates is indicated by the effective sample size (neff). The effective sample size measures the size of an (unweighted) simple random sample that would have provided the same precision (standard error) as the design being implemented. If the effective sample size is close to the actual sample size then we have an efficient design with a good level of precision. The lower the effective sample size, the lower the level of precision. The efficiency of a sample is given by the ratio of the effective sample size to the actual sample size. The range of the weights, the effective sample size and sample efficiency for both sets of weights are given in Table 5.6.

Table 5.6 Effective sample size and sample efficiency

	Birth cohort		Child cohort	
	Respondent weights	Partner weights	Respondent weights	Partner weights
Minimum weight	0.72	0.66	0.67	0.63
Mean weight	1.00	1.00	1.00	1.00
Maximum weight	1.84	2.55	1.80	2.06
Effective sample size	4,293	2,809	2,389	1,478
Sample efficiency	95%	94%	96%	96%
Unweighted sample size	4,511	2,979	2,500	1,543

5.5 Applying the weights

5.5.1 Main interview weights

These weights should be used for all analyses of sweep 2 interview data, including analysis of combined sweep 1 and sweep 2 data. They should not be used for analysis of data from the partner interview.

5.5.2 Partner weights

These weights should be used for all analysis of data from the partner interview. The purpose of the weights is to make the responding partners representative of all partners in the responding households at sweep 2.

6 Using the data

The GUS Sweep 1 data consists of two SPSS files

GUS_SW2_B.sav	4512 cases	Birth cohort
---------------	------------	--------------

GUS_SW2_C.sav	2500 cases	Child cohort
---------------	------------	--------------

6.1 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.

6.2 Variable naming convention

Variables names are 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 Sweep 1 and Sweep 2 the names will 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 the table overleaf.

6.3 Variable labels

In the Sweep 2 dataset the variable labels are restricted to 40 characters; 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. The variable list shows the page numbers of the relevant questionnaire section.

6.4 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.

GUS Variable Naming Convention

Character	1		2		3	4, 5 & 6	7&8
Source	Source of data		Sweep/Wave		Key theme prefix	Sub theme stem	Question/Variable number
Type of lettering/no.	Non- sequential Capitals: D,M, P, S		Sequential lower case: a, b, c..		Non-sequential Capitals: C, P, N...	Abbreviated lower case: hea,	01 - 99
	Source code	Details	Sweep code	Details	See Theme/prefix dictionary	See Stem dictionary	
	AL	Area Level variable					
	D	Derived variable	a	Cohort 1: Sweep 1 (2005/06)			
	DP	Derived variable from partner int	b	Cohort 1: Sweep 2 (2006/07)			
	DWP	DWP variable	c	Cohort 1: Sweep 3 (2007/08)			
	M	Main carer/adult interview					
	P	Partner's interview					
	S	Scale variable					
	W	Weights and Heights					

6.5 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. Each person was identified by their person number, which they will retain through each sweep of the survey. The variable MbHGSI(n) can be used to see whether a person who was in the household at sweep 1 is in the household at sweep 2.

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 6.1. 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 and partner age, marital status and relationship to other people in the household.

Table 6.1 Key household derived variables

DbHGnmad	Db - Number of adults in household
DbHGnmkd	Db - Number of children in household
DbHGnmsb	Db - Number of siblings in household
DbHGnp01	Db - Number of natural parents in household
DbHGrsp01	Db - Whether respondent is natural mother
DbHGrsp02	Db - Whether respondent is natural father
DbHGnp02	Db - Natural mother in household
DbHGnp03	Db - Natural father in household
DbHGrsp04	Db - Respondent living with spouse/partner
DbMothID	Db - Person number of mother
DbFathID	Db - Person number of father
DbRespID	Db - Person number of respondent
DbPartID	Db - Person number of partner
DbRPAge	Db - Respondent partners age
DbRPsex	Db - Respondent partners sex

6.6 Childcare data

The childcare section of the CAPI questionnaire utilises feed-forward data. This technique allows information collected at the previous sweep 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 2, for those cases where childcare had been used at sweep 1, details of sweep 1 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 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 2 but had not been in place at sweep 1. 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 sweep 1 childcare arrangements, 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 6.2.

Table 6.2 Childcare variables for exploring continuity of provision

Variable name	Description
MaCtya01	Sw1 1st childcare provider type
MaCtma01	Sw1 1st childcare provider - no of hours per week
MaCdya01	Sw1 1st childcare provider - no of days per week
MaCtyb01	Sw1 2nd childcare provider type
MaCtmb01	Sw1 2nd childcare provider - no of hours per week
MaCdyb01	Sw1 2nd childcare provider - no of days per week
MaCtyc01	Sw1 3rd childcare provider type
MaCtmc01	Sw1 3rd childcare provider - no of hours per week
MaCdyc01	Sw1 3rd childcare provider - no of days per week
MaCtyd01	Sw1 4th childcare provider type
MaCtmd01	Sw1 4th childcare provider - no of hours per week
MaCdyd01	Sw1 4th childcare provider - no of days per week
MaCtye01	Sw1 5th childcare provider type
MaCtme01	Sw1 5th childcare provider - no of hours per week
MaCdye01	Sw1 5th childcare provider - no of days per week
MbCsta01	Mb Whether still using 1st provider from sweep 1?
MbCcta01	Sw1 1st ccare provider - revised hrs at sw2
MbCcda01	Sw1 1st ccare provider - revised days at sw2
MbCrsa01	Main reason no longer using provider 1 from sw1 at sw2
MbCstb01	Mb Whether still using 2nd provider from sweep 1?
MbCctb01	Sw1 2nd ccare provider - revised hrs at sw2
MbCcdb01	Sw1 2nd ccare provider - revised days at sw2
MbCrse01	Main reason no longer using provider 2 from sw1 at sw2
MbCstc01	Mb Whether still using 3rd provider from sweep 1?
MbCctc01	Sw1 3rd ccare provider - revised hrs at sw2
MbCcde01	Sw1 3rd ccare provider - revised days at sw2
MbCrsc01	Main reason no longer using provider 3 from sw1 at sw2
MbCstd01	Mb Whether still using 4th provider from sweep 1?
MbCctd01	Sw1 4th ccare provider - revised hrs at sw2
MbCcdd01	Sw1 4th ccare provider - revised days at sw2
MbCrse01	Main reason no longer using provider 4 from sw1 at sw2
MbCste01	Mb Whether still using 5th provider from sweep 1?
MbCcte01	Sw1 5th ccare provider - revised hrs at sw2
MbCcde01	Sw1 5th ccare provider - revised days at sw2
MbCrse01	Main reason no longer using provider 5 from sw1 at sw2
DbCstp01	Db Has respondent stopped using any of the childcare arrangements they were using at sweep 1?
DbCstp02	Db How many of the childcare arrangements they were using at sweep 1

	has the respondent stopped using?
DbCnpv01	Db Number of childcare provs from sweep 1 still being used
DbCapv01	Is respondent still using a childcare provider that had been used at sweep 1?

The second section is concerned with the details of **new arrangements** which were in place at sweep 2. 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 6.3.

Table 6.3 Variables for exploring new childcare arrangements at sweep 2

Variable name	Description
MbCany02	Mb Whether using any childcare (those who had no childcare at sw1)
MbCany03	Mb Whether using any new childcare arrangements (those who were using childcare at sw1)
Mbctya01	Mb 1st new provider type
Mbctma01	Mb 1st new ccare provider - hrs per week
Mbcdya01	Mb 1st new ccare provider - no of days per week
Mbcaga01	Mb Age (months) child started with 1st new provider
MbCwya01 – MbCwya18	Mb Reasons for using 1 st new provider
Mbctyb01	Mb 2nd new provider type
Mbctmb01	Mb 2nd new ccare provider - hrs per week
Mbcdyb01	Mb 2nd new ccare provider - no of days per week
Mbcagb01	Mb Age (months) child started with 2nd new provider
MbCwyb01 – MbCwyb18	Mb Reasons for using 2 nd new provider
mbctyc01	Mb 3rd new provider type
mbctmc01	Mb 3rd new ccare provider - hrs per week
mbcdyc01	Mb 3rd new ccare provider - no of days per week
mbcagc01	Mb Age (months) child started with 3rd new provider
MbCwyc01 – MbCwyc18	Mb Reasons for using 3 rd new provider
mbctyd01	Mb 4th new provider type
mbctmd01	Mb 4th new ccare provider - hrs per week
mbcdyd01	Mb 4th new ccare provider - no of days per week
mbcagd01	Mb Age (months) child started with 4th new provider
MbCwyd01 – MbCwyd18	Mb Reasons for using 4 th new provider
DbCnnp01	Db No of new childcare arrangements in place at sweep 2

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 2 interview - the provider type, number of hours and days of the arrangement, and whether or not it is a new arrangement at sweep 2. 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 6.4.

Table 6.4 Variables for exploring new current childcare arrangements at sweep 2

Variable name	Description
DbCtya01	Sw2 Childcare provider A - provider type
DbCnwa	Is Provider A a new or existing arrangement?
DbCtma01	Db Provider A: No of hours child looked after per week
DbCdya01	Db Provider A: No of days child looked after per week
DbCtyb01	Sw2 Childcare provider B - provider type
DbCnwb	Is Provider B a new or existing arrangement?
DbCtmb01	Db Provider B: No of hours child looked after per week
DbCdyb01	Db Provider B: No of days child looked after per week
DbCtyc01	Sw2 Childcare provider C - provider type
DbCnwc	Is Provider C a new or existing arrangement?
DbCtmc01	Db Provider C: No of hours child looked after per week
DbCdyc01	Db Provider C: No of days child looked after per week
DbCtyd01	Sw2 Childcare provider D - provider type
DbCnwd	Is Provider D a new or existing arrangement?
DbCtmd01	Db Provider D: No of hours child looked after per week
DbCdyd01	Db Provider D: No of days child looked after per week
DbCtye01	Sw2 Childcare provider E - provider type
DbCnwe	Is Provider E a new or existing arrangement?
DbCtme01	Db Provider E: No of hours child looked after per week
DbCdye01	Db Provider E: No of days child looked after per week
DbCany01	Db Does respondent currently get help with childcare for ^childname on a regular basis? (not including the excluded pre-school cases – see 6.6.1)
DbCtot01	Db Number of childcare providers being used at sweep 2 (not including the excluded pre-school cases – see 6.6.1)

Although not listed in Table 6.4, 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 2 questionnaire which accompanies this user guide.

6.6.1 Childcare and Pre-school arrangements

At the time of the sweep 2 interview, children in the child cohort were aged just under 4 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. As such, a module on the transition to and early experiences of pre-school was included in the questionnaire for parents in the child cohort. The pre-school module collected only broad details about the actual provision; questions in the childcare section, which encompassed pre-school, would collect more precise information on the provider type, the number of hours and the number of days. However, it became clear on analysis that a number of parents whose children were attending pre-school had not provided those pre-school details in the childcare section. The exclusion of these pre-school arrangements from the childcare data meant that data on the proportion of parents using childcare, the number of providers being used, the mix of provision and the total number of hours, was inaccurate in that it missed the pre-school arrangement.

To resolve this, a number of derived variables have been created which incorporate information from the pre-school module and create more a more accurate picture of **current** childcare use amongst parents. These variables are listed in Table 6.5.

Table 6.5 Childcare variables including a correction for the excluded pre-school cases

Variable name	Description
DbCany02	Whether or not using childcare (including those who had excluded pre-school arrangements)
DbCtot02	Number of childcare providers being used at sw2 (including previously excluded pre-school arrgts)
DbCtyf01	Sw2 Childcare provider E – derived provider type for those who did not provide pre-school details in childcare section
DbCtmf01	No of hours looked after per week by provider F (excluded pre-school provider)
DbCdyf01	No of days looked after per week by provider F (excluded pre-school provider)
DbCtmi01	Db Total number of hours child is currently looked after by someone else in an average week
DbCtmi02	Db Total number of hours child is currently looked after by someone else in an average week - BANDED
DbCday01	Db Highest number of days child is in any one childcare arrangement
DbCtyp01	Db Does respondent use grandparents for childcare?
DbCtyp02	Db Does respondent use another relative for childcare?
DbCtyp03	Db Does respondent use private creche/nursery for childcare?
DbCtyp04	Db Does respondent use a childminder for childcare?
DbCtyp05	Db Does respondent use a local authority playgroup for childcare?
DbCtyp06	Db Does respondent use a local authority nursery for childcare?
DbCtyp07	Db Does respondent use a private playgroup for childcare?
DbCtyp08	Db Does respondent use a community/voluntary playgroup for childcare?
DbCtyp09	Db Does respondent use an ex-spouse or partner for childcare?
DbCtyp10	Db Does respondent use the childs older sibling for childcare?
DbCtyp11	Db Does respondent use a friend or neighbour for childcare?
DbCtyp12	Db Does respondent use a daily visiting nanny for childcare?
DbCtyp13	Db Does respondent use a live-in nanny for childcare?
DbCtyp14	Db Does respondent use a babysitter for childcare?
DbCtyp15	Db Does respondent use a workplace creche or nursery for childcare?
DbCtyp16	Db Does respondent use a family centre for childcare?
DbCtyp17	Db Does respondent use a nursery class attached to a primary school for childcare?
DbCtyp18	Db Does respondent use a childcarer (provided via childcare agency) for childcare?
DbCtyp19	Db Does respondent use another type of childcare provider for childcare?
DbCtyp20	Db Does respondent currently use OTHER INFORMAL childcare?
DbCtyp21	Db Does respondent currently use NURSERY OR CRECHE for childcare?
DbCtyp22	Db Does respondent currently use PLAYGROUP for childcare?
DbCtyp23	Db Does respondent currently use OTHER PROVIDERS for childcare?
DbCtyp30	Db Does respondent currently use informal childcare?
DbCtyp31	Db Does respondent currently use formal childcare?
DbCtyp32	Db Current use of formal and informal childcare

6.7 Indicators and summary variables

6.7.1 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.

Further information on NS-SEC is available from the National Statistics website at: http://www.statistics.gov.uk/methods_quality/ns_sec/cat_subcat_class.asp

6.7.2 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 6.2 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 Executive (2004) *Scottish Executive Urban Rural Classification 2003 – 2004*. This document is available online at <http://www.scotland.gov.uk/Publications/2004/06/19498/38784>

6.7.3 Area-level variables: Scottish Index of Multiple Deprivation

The Scottish Index of Multiple Deprivation (SIMD) 2006 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 2006 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 1 datasets is based on the 2006 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 had not been published at the time the 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>

6.7.4 Child Development: Communication and Symbolic Behaviour Scale – Infant/Toddler Checklist

Within the self-completion section of the interview, respondents had to complete questions which assessed their child's communication, emotional development, understanding and interaction with peers. Questions for parents in the birth cohort form the Infant/Toddler checklist of the Communication and Symbolic Behaviour (CSBS) (Wetherby and Prizant, 2001).

The 24 questions are grouped into categories called clusters. The items in each cluster can be totalled to yield seven individual cluster scores. Results from the clusters are then used to produce three composite scores each assessing different aspects of the child's development – social communication, expressive speech/language and symbolic functioning⁴. A total score can also be calculated by summing the three composite scores. Those children who score below a certain level on the scale are considered to be 'of concern' in relation to their development. As well as containing the constituent items, the dataset also includes a set of derived variables including the various composite scores and total score. Details of these variables are included in Table 6.2. Corresponding syntax is detailed in the derived variable documentation which accompanies this User Guide.

Table 6.3 Derived variables associated with the CSBS Infant/Toddler Checklist

Variable name	Description
DbDcsc01	CSBS Cluster 1: Emotion and eye gaze'
DbDcsc02	CSBS Cluster 2 Score: Communication'
DbDcsc03	CSBS Cluster 3: Gestures'
DbDcsc04	CSBS Cluster 4: Sounds'
DbDcsc05	CSBS Cluster 5: Words'
DbDcsc06	CSBS Cluster 6: Understanding'
DbDcsc07	CSBS Cluster 7: Object Use'.
DbDcsc11	CSBS Social Composite Score (0-26
DbDcsc12	CSBS Speech Composite Score (0-14
DbDcsc13	CSBS Symbolic Composite Score (0-17).'
DbDcsc20	CSBS Total Score (0-57)'. '
DbDcsc21	CSBS: Whether child is in concern or no concern range for social composite
DbDcsc22	CSBS: Whether child is in concern or no concern range for speech composite
DbDcsc23	CSBS: Whether child is in concern or no concern range for symbolic composite
DbDcsc30	CSBS: Whether child is in concern or no concern range for overall measure

Further details on the CSBS can be found at:

<http://www.brookespublishing.com/store/books/wetherby-csbsd/index.htm#checklist>

6.7.5 Child Development: Strengths and Difficulties Questionnaire

Parents in the child cohort completed the Strengths and Difficulties Questionnaire (SDQ). The SDQ is a brief behavioural screening questionnaire designed for use with 3-16 year olds. 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 6.4 with syntax illustrated in the derived variables documentation.

Table 6.4 Derived variables associated with the Strengths and Difficulties Questionnaire

Variable name	Description
DbDsdem1	SDQ: Emotional symptoms score
DbDsdco1	SDQ: Conduct problems score
DbDsdhy1	SDQ: Hyper-activity or inattention score
DbDsdpr1	SDQ: Peer problems score
DbDsdps1	SDQ: Pro-social score
DbDsdto1	SDQ: Total difficulties score

Further details on the SDQ can be found at:

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

6.7.6 Parental Health: Depression, Anxiety and Stress Scale

Six items from the Depression, Anxiety and Stress (DASS) scale (Lovibond & Lovibond, 1995) were included in the self-completion section of the interview. DASS is available in a 42-item, or 21-item scale in full. We took 6 items: 3 measuring stress, and 3 measuring depression. These items can be combined to create a stress scale and depression scale. Standardized versions of the scales (z-scores) can be combined to produce a single scale measuring evidence of negative emotional symptoms in the respondent. The constituent items and the derived scale variables are detailed in Table 6.4 below. Syntax for compiling the derived variables is detailed in the derived variables documentation.

Table 6.4 Constituent and derived variables associated with the Depression, Anxiety and Stress scale

Variable name	Description
MbHdas01	I found myself getting upset by quite trivial things (stress)
MbHdas02	I found it difficult to relax (stress)
MbHdas03	I felt that I had nothing to look forward to (depression)
MbHdas04	I felt sad and depressed (depression)
MbHdas05	I found that I was very irritable (stress)
MbHdas06	I was unable to become enthusiastic about anything (depression)
DbHdas01	DASS Raw Stress Score
DbHdas02	DASS Raw Depression Score (0-9)
ZDbHdas01	Standardised DASS Stress Score
ZDbHdas02	Standardised DASS Depression Score
DbHdas03	Composite DASS score

Further information on DASS is available at:

<http://www2.psy.unsw.edu.au/groups/dass/>

6.8 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 recorded in multiple variables instead).

The following types of variables (specified below) 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
 - Specific country of birth
 - 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 and Scottish index of multiple deprivation summary variable.

6.9 Weighting variables

The final main interview sweep 2 weights are **DbWTbrth** (birth cohort) and **DbWTchld** (Child cohort). The partner weights are **DbWtBrnP** (bith cohort) and **DbWtchIP** (child cohort).

Separate weights are provided for each cohort because analysis should always treat each cohort as a distinct population. However, key analysis using this data may involve comparison between the cohorts. It is usually more convenient to undertake this analysis by combining the two cohort datasets into a single dataset and then ensuring that subsequent analysis is either filtered to select a single cohort, or that output is nested by cohort type ('SampType'). On merging the datasets it is possible to create a combined weight variable in order that nested analysis uses just a single weight variable. The value of the combined weight is equal to the value of the corresponding cohort weight variable for that child. Syntax to create the combined main interview weight is included below:

Compute DbWTbrch = DbWTbrth.

If (SampType = 2) DbWTbrch = DbWTchld.

6.10 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.

7 Documentation

The documentation has been organised into the following sections

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

8 Related publications

Further information about GUS Sweep 2 is available in:

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

Other publications which include the use of GUS data include:

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

These reports and other related information are available on the GUS website:
<http://www.growingupinScotland.org.uk/>

9 Contact details

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

GUS Project Manager Paul Bradshaw 0131 228 2167 p.bradshaw@scotcen.org.uk
GUS Data Manager Joan Corbett 0131 221 2560 j.corbett@scotcen.org.uk

10 References

Goodman, R. (1997) "The Strengths and Difficulties Questionnaire: A Research Note", *Journal of Child Psychology and Psychiatry*, 38, 581-586.

Lovibond, S.H. & Lovibond, P.F. (1995). *Manual for the Depression Anxiety Stress Scales*. (2nd. Ed.) Sydney: Psychology Foundation

Wetherby, A.M. and Prizant, B.M. (2001), *Communication and Symbolic Behaviour Scales - Infant/Toddler Checklist*, Baltimore: Paul H. Brookes Publishing Co.

Appendix A: Full non-response models

Table A1 Full model for non-response to sweep 2 interview (birth cohort)

Variables in the Equation	B	S.E.	Wald	df	Sig.	Exp(B)
Household income			13.7	4	0.01	
Less than £9,999					(baseline)	
£10,000-£19,999	0.14	0.13	1.1	1	0.30	1.15
£20,000-£31,999	0.15	0.16	0.9	1	0.35	1.16
£32,000 or more	0.25	0.18	2.0	1	0.16	1.28
Missing	-0.29	0.16	3.3	1	0.07	0.75
Tenure			21.1	2	0.00	
Owner occupier	0.59	0.14	18.8	1	0.00	1.80
Rents - HA/council	0.52	0.14	14.5	1	0.00	1.68
Rents - private					(baseline)	
Family type			11.2	1	0.00	
Lone parent family	-0.40	0.12	11.2	1	0.00	0.67
Couple family					(baseline)	
Ethnicity of respondent			16.4	1	0.00	
White ethnic background	0.74	0.18	16.4	1	0.00	2.09
Other ethnic background					(baseline)	
Selected child was breastfed			10.0	1	0.00	
Yes	0.29	0.09	10.0	1	0.00	1.34
No					(baseline)	
Scottish Index of Multiple Deprivation			11.4	4	0.02	
Least deprived (0.5393 - 7.7347)	0.21	0.16	1.7	1	0.19	1.23
(7.7354 - 13.5231)	0.24	0.14	3.0	1	0.09	1.28
(13.5303 - 21.0301)	0.43	0.14	9.9	1	0.00	1.53
(21.0421 - 33.5214)	0.05	0.12	0.2	1	0.69	1.05
(33.5277 - 87.5665) most deprived					(baseline)	
Respondent regularly attended baby groups in the past year			5.6	1	0.02	
Yes	0.23	0.10	5.6	1	0.02	1.26
No					(baseline)	
Constant	-0.07	0.65	0.0	1	0.91	0.93

Notes: Outcome is 1 = respondent gave a sweep 2 interview, 0 = no sweep 2 interview,
 Base is all households eligible for sweep 2 in the birth cohort (n = 5,185),
 R squared (Cox and Snells) 0.052.

Table A2 Full model for non-response to sweep 2 interview (child cohort)

Variables in the Equation	B	S.E.	Wald	df	Sig.	Exp(B)
Tenure			16.2	2	0.00	
Owner occupier	0.74	0.20	14.0	1	0.00	2.11
Rents - HA/council	0.24	0.19	1.6	1	0.20	1.27
Rents - private					(baseline)	
Household income			18.0	4	0.00	
Less than £9,999					(baseline)	
£10,000-£19,999	-0.18	0.18	1.1	1	0.31	0.83
£20,000-£31,999	-0.05	0.23	0.0	1	0.84	0.96
£32,000 or more	0.25	0.24	1.1	1	0.29	1.29
Missing	-0.59	0.22	7.4	1	0.01	0.56
Respondent regularly uses childcare			13.4	1	0.00	
Yes	0.47	0.13	13.4	1	0.00	1.60
No					(baseline)	
Family type			8.4	1	0.00	
Lone parent family	-0.46	0.16	8.4	1	0.00	0.63
Couple family					(baseline)	
Respondent's general health			10.5	3	0.02	
Excellent	-0.53	0.20	6.9	1	0.01	0.59
Very good	-0.05	0.18	0.1	1	0.76	0.95
Good	-0.29	0.18	2.6	1	0.11	0.75
Fair/poor					(baseline)	
ONS Urban/Rural indicator - Scotland			13.8	5	0.02	
Large urban area (125,000+)	-0.91	0.38	5.8	1	0.02	0.40
Other urban area (10,000-125,000)	-0.60	0.38	2.5	1	0.12	0.55
Accessible small town (3,000-10,000)	-0.54	0.41	1.8	1	0.19	0.58
Remote and very remote small town (3,000-10,000)	0.07	0.58	0.0	1	0.90	1.07
Accessible rural (<3,000)	-0.62	0.40	2.4	1	0.12	0.54
Remote and very remote rural (<3,000)					(baseline)	
Constant	1.61	0.44	13.4	1	0.00	5.02

Notes: Outcome is 1 = respondent gave a sweep 2 interview, 0 = no sweep 2 interview,

Base is all households eligible for sweep 2 in the child cohort (n = 2,845),

R squared (Cox and Snells) = 0.047

Table A3 Full model for non-response to the partner interview (birth cohort)

	B	S.E.	Wald	df	Sig.	Exp(B)
Highest Education level of Respondent -						
Banded			23.4	2	0.000	
Higher or above	0.57	0.15	13.6	1	0.000	1.77
Standard grade or other	0.15	0.17	0.8	1	0.366	1.17
No qualifications					(baseline)	
Mothers employment status						
Child's mother working - full-time	0.96	0.18	27.2	1	0.000	2.60
Child's mother working - part-time	0.40	0.13	8.9	1	0.003	1.49
Child's mother not working					(baseline)	
Household employment and family type						
Couple family both mother and partner working >16 hours	-1.23	0.24	27.2	1	0.000	0.29
Couple family either mother or partner working >16 hours	-0.86	0.20	18.4	1	0.000	0.42
Couple family both unemployed or <16 hours					(baseline)	
Ethnicity of Child						
White	0.50	0.17	8.9	1	0.003	1.64
Other ethnic background					(baseline)	
Age of mother at birth of sample child (banded)						
Under 20	-0.40	0.32	1.6	1	0.209	0.67
20 to 29	-0.40	0.24	2.9	1	0.088	0.67
30 to 39	-0.08	0.23	0.1	1	0.720	0.92
40 or older					(baseline)	
Whether first-time (Primiparous) mother						
Primiparous	0.35	0.09	15.7	1	0.000	1.42
Other children					(baseline)	
Constant	1.10	0.34	10.4	1	0.001	3.00

Notes: Outcome is 1 = partner gave an interview, 0 = no partner interview, Base is all households with an eligible partner in the birth cohort (n = 3,764), R squared (Cox and Snells) 0.032.

Table A4 Full model for non-response to the partner interview (child cohort)

	B	S.E.	Wald	df	Sig.	Exp(B)
ONS Urban/Rural indicator - Scotland			12.3	5	0.031	
Large urban area (125,000+)	-0.31	0.29	1.2	1	0.283	0.73
Other urban area (10,000-125,000)	-0.55	0.29	3.7	1	0.053	0.57
Accessible small town (3,000-10,000)	-0.34	0.32	1.1	1	0.290	0.71
Remote and very remote small town (3,000-10,000)	-1.06	0.39	7.2	1	0.007	0.35
Accessible rural (<3,000)	-0.61	0.30	4.1	1	0.042	0.54
Remote and very remote rural (<3,000)					(baseline)	
Mothers employment status			16.8	2	0.000	
Child's mother working - full-time	0.91	0.23	16.1	1	0.000	2.49
Child's mother working - part-time	0.62	0.18	11.9	1	0.001	1.86
Child's mother not working					(baseline)	
Household employment and family type			13.0	2	0.001	
Couple family both mother and partner working >16 hours	-1.03	0.33	10.0	1	0.002	0.36
Couple family either mother or partner working >16 hours	-0.51	0.29	3.2	1	0.076	0.60
Couple family both unemployed or <16 hours					(baseline)	
Ethnicity of Child			3.8	1	0.052	
White	0.47	0.24	3.8	1	0.052	1.60
Other ethnic background					(baseline)	
Whether first-time (Primiparous) mother			5.1	1	0.024	
Primiparous	0.25	0.11	5.1	1	0.024	1.29
Other children					(baseline)	
Constant	1.41	0.44	10.1	1	0.001	4.11

Notes: Outcome is 1 = partner gave an interview, 0 = no partner interview,

Base is all households with an eligible partner in the child cohort (n = 1,998),

R squared (Cox and Snells) = 0.021