Scottish Centre *for*
Social Research

Growing Up In Scotland

Sweep 1 - 2005

User Guide

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

1 Overview of the survey

The data files contain data from Growing Up in Scotland (GUS) Sweep 1, the first 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 Executive Education Department, 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 planned configuration of cohorts and sweeps is summarised below: BC1 refers to the younger of the two cohorts ('birth cohort') and CC1 to the slightly older cohort ('child cohort'). At present, the intention is to launch a new birth cohort (BC2) in 2009.

Year	Age at interview					
	0-1	1-2	2-3	3-4	4-5	5-6
2005	BC1		TC1			
2006		BC1		TC1		
2007			BC1		TC1	
2008				BC1		TC1
2009	(BC2)				BC1	
2010		(BC2)				BC1
2011			(BC2)			

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 Executive. 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 Authority[1] 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

A pilot of the computer-assisted personal interview (CAPI) instrument, including the self-completion questionnaire and associated documents was conducted in February 2005. All pilot respondents were drawn from child benefit records. The pilots tested the fieldwork procedure and interview content. Some new questions underwent cognitive testing, and some innovative measures were introduced.

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 this wave, 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.

2.2   Length of Interview

Overall, the average interview lasted around 65 minutes. The child cohort interview had a slightly longer average length at 66 minutes, than the birth cohort interview at 64 minutes.

2.3   Timing of fieldwork

Fieldwork was undertaken over a fourteen month period commencing in April 2005. 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 2005 and remained in field until mid-July 2005.

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 10.5 months old, and for the child cohort the date the child turned 34.5 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 two weeks.

3   Response rates

Details of the number of eligible cases identified by the DWP, 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	8218	4712	12930
Cases removed	966	655	1621
Cases to field:			
All	7252	4057	11309
Achievable or 'in-scope'*	6583	3605	10143
Cases achieved	5217	2858	8075
Response rate			
As % of all eligible children	63%	61%	62%
As % of all 'in-scope'	80%	79%	80%

*Cases which were considered out-of-scope or unachievable were mostly ineligible or incorrect addresses.

4   Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted. The GUS Sweep 1 Code Book and Edit Instructions provide details of the tasks that were conducted.

4.1   Coding

The coding of responses was mostly dealt with by the CAPI questionnaire through the use of fully closed questions. There were, however, a number of questions (e.g. **MaTotb01**) where a code frame was not used in order to capture all responses to a particular question verbatim. The responses to these 'open' questions were coded into separate variables after the interview was conducted**.**

The coded responses to open questions are in the core dataset, but as mentioned above the original text responses have been removed to reduce the potential to identify individuals.

Other questions in the CAPI questionnaire had code frames, which included an 'other' option (e.g. **MaCwya01**). In these 'other-specify' questions, interviewers could use this option if the respondent's answer did not fit any of the codes or if they were not confident of coding into the prescribed codes. In these cases, the interviewer recorded the full 'other' answer at a follow up question (e.g. 'other' responses relating to **MaCwya01** were recorded in **MaCwya04**). For all multi-coded variables that were coded, the original and coded answers have been merged into a single set of variables.

4.2   Editing

As with the coding, most of the editing for the GUS Wave 1 questionnaire was carried out by the interviewers in the field. However, there were a couple of additional checks that related to inconsistencies in the data that were carried out after the interview.

5 Weighting the data

The final weights were generated in a number of stages. The first stage generated selection weights to correct the differential selection probabilities for some children. The second stage modelled non-response bias and generated a non-response weight. This weight corrects for the effects of non-response. The final stage adjusted the composite selection and non-response weight to create a set of calibration weights. These weights make the (weighted) sample match the population in terms of the variables used to calibrate, in this case, age sex and month of issue. This corrects for the effects of the exclusions made by the DWP.

5.1 Stage 1 - Selection weights for twins and multiple cohort households

There were 225 households where more than one child was eligible (154 pairs of twins and 71 cross-cohort multiple births). In households where the eligible children were in the same cohort, each child had a 1 in n chance of being selected (where n is the number of eligible children in the household). In households where the children were in different cohorts older children had a smaller chance of being selected, this was to reflect the smaller chance they had of being included in the sample overall. Children in the child cohort had a selection probability of 3/8, for those in the birth cohort this was 5/8. The multi-birth weight (wt_sel) is the inverse of these selection probabilities.

There were no selection weights generated to correct for sampling fewer older children in relation to younger children. **This is because data from the birth cohort and that from the child cohort should always be analysed separately. They can not be analysed together because together they do not represent a coherent or real population and any results would be misleading.**

5.2 Stage 2 - Non-response weights

There was a limited amount of information available on the sampling frame with which to build a non-response model. The aim was to model the non-response behaviour using logistic regression. Once a model had been generated it was used to get predicted probabilities for each case. These probabilities are the probability the parent will respond to the survey, given their characteristics. The non-response weights (wt_NR) are then generated as the inverse of these saved probabilities. The modelling was run on data weighted by the selection weight.

A number of the variables used in the modelling were taken from the child benefit records, these were; age of claimant, sex of claimant, number of children in the household and method of benefit payment. The other variables were at the area-level[2], these were, Scottish index of multiple deprivation (quintiles), population density measured by the number of persons in private households per hectare and ONS urban rural indicator. These variables were matched into the data by postcode.

The two cohorts were modelled separately. The model suggests responders in the birth cohort were more likely to be older parents, have fewer than 4 children in the household and live in rural and less deprived areas. Responders in the child cohort were also more likely to be older parents, live in rural and less deprived areas and to receive payment by account.

---

2 Area level information has been used because there was not much data available from the sampling frame with which to generate non-response weights. At the second wave there will be more information available to model non-response with. The high response rate (62%) means bias from non-response at this stage is likely to be minimal.

The final logistic regression non-response models are shown in Tables 5.1 and 5.2. All the variables, with the exception of number of children per household, were included in the final model.

Table 5.1   Non-response model for birth cohort

Variables in the Equation	B	S.E.	Wald	df	Sig.	Exp(B)
Age of claimant			48.0	5	0.000	
20 or less	-0.7	0.16	19.0	1	0.000	0.51
21-24	-0.5	0.14	9.6	1	0.002	0.64
25-29	-0.4	0.14	10.4	1	0.001	0.64
30-34	-0.2	0.14	1.4	1	0.244	0.85
35-39	-0.1	0.14	0.5	1	0.485	0.91
40+	(baseline)					
Area deprivation index			19.8	4	0.001	
Least deprived (0.5393 - 7.7347)	0.4	0.09	17.5	1	0.000	1.45
(7.7354 - 13.5231)	0.2	0.09	6.4	1	0.011	1.24
(13.5303 - 21.0301)	0.2	0.08	5.3	1	0.021	1.21
(21.0421 - 33.5214)	0.1	0.08	0.7	1	0.399	1.07
(33.5277 - 87.5665) most deprived	(baseline)					
Population density			41.0	4	0.000	
Least dense (0.0012 - 0.1729)	0.6	0.14	21.0	1	0.000	1.87
(0.1739 - 1.3274)	0.5	0.10	29.7	1	0.000	1.72
(1.3625 - 8.9643)	0.3	0.08	11.6	1	0.001	1.29
(9.0377 - 25.6086)	0.2	0.07	6.2	1	0.013	1.19
(25.8113 - 128.49) most dense	(baseline)					
Number of children in the household			9.7	3	0.021	
1	0.3	0.12	5.4	1	0.020	1.31
2	0.4	0.12	9.2	1	0.002	1.43
3	0.3	0.13	5.7	1	0.017	1.36
4 or more	(baseline)					
Constant	0.5	0.18	8.6	1	0.003	1.72

Notes:
1. The response is 1 = responded to sweep 1 of GUS, 0 = did not respond
2. Only variables that are significant at the 0.05 level are included in the model.
3. The model R2 is 0.039 (Cox and Snell).
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 5.2 - Non-response model for child cohort

Variables in the Equation	B	S.E.	Wald	df	Sig.	Exp(B)
Age of claimant			69.7	5	0.000	
20 or less	-0.9	0.17	29.2	1	0.000	0.39
21-24	-0.9	0.13	51.0	1	0.000	0.39
25-29	-0.8	0.12	39.8	1	0.000	0.47
30-34	-0.5	0.12	21.2	1	0.000	0.59
35-39	-0.4	0.12	11.9	1	0.001	0.67
40+	(baseline)					
Payment type			4.9	1	0.026	
Account	0.4	0.16	4.9	1	0.026	1.43
Other	(baseline)					
Area deprivation index			8.2	4	0.050	
Least deprived (0.5393 - 7.7347)	0.1	0.09	1.9	1	0.166	1.13
(7.7354 - 13.5231)	0.1	0.09	1.9	1	0.171	1.13
(13.5303 - 21.0301)	0.1	0.08	0.6	1	0.455	1.07
(21.0421 - 33.5214)	-0.1	0.08	1.5	1	0.220	0.90
(33.5277 - 87.5665) most deprived	(baseline)					
Population density			65.3	4	0.000	
Least dense (0.0012 - 0.1729)	0.9	0.15	38.4	1	0.000	2.47
(0.1739 - 1.3274)	0.7	0.10	39.7	1	0.000	1.92
(1.3625 - 8.9643)	0.3	0.08	19.2	1	0.000	1.42
(9.0377 - 25.6086)	0.2	0.07	6.4	1	0.012	1.20
(25.8113 - 128.49) most dense	(baseline)					
Number of children in the household			7.3	3	0.063	
1	-0.3	0.13	6.0	1	0.014	0.73
2	-0.2	0.13	2.9	1	0.086	0.80
3	-0.3	0.14	4.2	1	0.041	0.76
4 or more	(baseline)					
Constant	1.1	0.23	22.9	1	0.000	3.07

Notes:

1. The response is 1 = responded to sweep 1 of GUS, 0 = did not respond

2. Only variables that are significant at the 0.05 level are included in the model.

3. The model R2 is 0.038 (Cox and Snell).

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.

6 Using the data

The GUS Sweep 1 data consists of two SPSS files

GUSSW1B.sav	5217 records	Birth cohort

GUSW1C.sav	2859 records	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 "**List of Variables**" document in the data section of the documentation. This document is the best place to look at in order to plan your analysis.

Once you have decided which variables to include in your analysis, you can look up details of the question wording using the interview section documentation (all variables on the data file are given by name in the copy of the interview schedules provided), or use the "**Derived Variables Specification**" document in the data section of the documentation for derived variables.

6.2 Indicators and summary variables

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

Scottish Executive Urban/Rural Classification

The Scottish Executive Urban Rural Classification was first released in 2000 and is consistent with the Executive'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 6.1   Scottish Executive Urban Rural Classification

Scottish Executive Urban Rural Classification	
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

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

Medical Outcomes Study 12-Item Short Form (SF-12)

In the GUS Sweep 1 report, health-related quality of life was measured by the Medical Outcomes Study 12-Item Short Form (SF-12). This was also used in the Scottish Health Survey in 2003, 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.

6.3 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 replace 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.

6.4 Weighting variables

The final wave 1 weights are called **DaWTbrth** (birth cohort) and **DaWTchld** (Child cohort).

6.5　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 (contains the CAPI documentation - questionnaires and showcards, interviewer, and coding instructions)
- Data documentation (contains the list of variables, list of derived variables and SPSS syntax for derived variables)

8　Related publications

Further information about GUS Sweep 1 is available in:

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

Or 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