

Chapter 12: Appendix A: Technical Notes

12.1 Sampling

12.1.1 Sample 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 12,930 children was selected from the Child Benefit records to give an achieved sample of approximately 8,000 overall.

The area-level sampling frame was created by aggregating Data Zones. Data Zones are small geographical output areas created for the Scottish Executive. They were used to release data from the Census 2001 and 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 enough issued cases to achieve a sample of 8,000 (this is reliant on the birth rate remaining roughly constant). Once the merging task was complete, the list of aggregated areas was sorted by Local Authority ²⁰ and then by the Scottish Index of Multiple Deprivation score. One hundred and thirty areas were then selected at random. The Department of Work and Pensions (DWP) 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 12.1). 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 within the same household were in different age cohorts the baby had a higher chance of being selected, this was to ensure the ratio of babies to children remained constant.

After selecting the eligible children, the DWP made a number of exclusions before transferring the sample details to ScotCen. These exclusions included cases they considered 'sensitive' and children that had been sampled for research by the DWP in the last 3 years.

Table 12.1 Eligible child dates of birth for inclusion in the Growing Up in Scotland study by sample type

Sample Number	Dates of Birth required - Baby sample	Dates of Birth required - Toddler sample
Pilot 1	01-Jan-2004 - 31-Jan-2004	01-Jan-2002 - 31-Jan-2002
Pilot 2	01-Mar-2004 - 31 Mar-2004	01-Mar-2002 - 31 Mar-2002
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

12.1.2 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 11.2.

Table 12.2 Number of issued and achieved cases and response rates

	Babies	Toddlers	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.

12.2 Data collection

12.2.1 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, inputted their responses directly into the questionnaire programme.

Interviews were conducted with the child's main carer. At this sweep, 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.

12.2.2 Length of interview

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

12.2.3 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 baby cohort this was identified as the date on which the child turned 10.5 months old, and for the toddler 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.

12.3 Analysis

12.3.1 Weighting

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.

The selection weights are equivalent to the inverse of the selection probabilities. For samples 1 to 11 all babies and three-fifths of toddlers were included in the sample. The toddler selection weight is the inverse of the selection probabilities and is equal to 5/3 for toddlers and 1 for babies. For sample 12, where smaller numbers were initially drawn by the DWP, this weight is equal to 1.

12.3.2 Estimating the precision of estimates

Each percentage quoted in this report has an associated margin of error, due to the fact that it is based on only a *sample*, rather than *all* children. This margin can be estimated for each proportion, p (where p is the percentage divided by 100) by:

$$\pm 2x \sqrt{\frac{p(1-p)}{n}}$$

where n is the unweighted sample size (base). This margin corresponds to 95% confidence. In other words there is a 95% chance that the true value across *all* children in the subgroup (as opposed to just those in the sample) falls within this margin.

For example, in Table 5.2, the proportion of children in the baby sample who are looked after by their grandparents during the day on a daily or almost daily basis is estimated as 20% and the unweighted base is 5179. The margin of error around this estimate can be calculated as:

$$\pm 2x \sqrt{\frac{(0.2 \times 0.8)}{5179}}$$

which comes to 0.01. In other words, there is a 95% chance that the true value is within the range 0.20 ± 0.01 , i.e. between 0.19 and 0.21 or between 19% and 21%. In general, the larger the base, the more accurate the estimate is likely to be. ²¹

12.3.3 Presentation of results

Percentages

The percentages shown in the report text have all been rounded to the nearest whole number. Those shown in tables and graphs have been rounded to one decimal place. Consequently, the percentages in one column of a table will not necessarily add to exactly 100.

A dash (-) indicates no respondents at all. All figures are *column* percentages, except where otherwise indicated.

Bases

Each table shows the weighted and unweighted bases corresponding to each percentage. The data were weighted to compensate for differential non-response across the subgroups. The weighted bases can be used to (approximately) combine two different columns in a table. The unweighted bases can be used to calculate the precision of estimates.

12.3.4 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 study generally used the five category system in which respondents are classified as managerial and professional, intermediate, small employers and own account workers, lower supervisory and technical, and semi-routine and routine occupations. Unless otherwise stated, the analysis employs a household level measure of NS-SEC.

NS-SEC was introduced in 2001 and replaced Registrar General's Social Class (which had been used in the 1995 and 1998 surveys) as the main measure of socio-economic status.

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

12.3.5 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 in 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 12.3 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>

12.3.6 Scottish Index of Multiple Deprivation (SIMD)

The Scottish Index of Multiple Deprivation (SIMD) identifies small area concentrations of multiple deprivation across all of Scotland in a fair way. It allows effective targeting of policies and funding where the aim is to wholly or partly tackle or take account of area concentrations of multiple deprivation.

The first Index (SIMD 2004) was published in June 2004 and was based on 31 indications in the six individual domains of Current Income, Employment, Housing, Health, Education, Skills and Training and Geographic Access to Services and Telecommunications.

The SIMD 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.

For the purposes of this report, the full index has been separated into quintiles and each case has been assigned a quintile based on the residential postcode. Quintiles are percentiles which divide a distribution into fifths, i.e., the 20th, 40th, 60th, and 80th percentiles. For example, 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.