Growing Up in Scotland

Birth Cohort 2

Sweep 2.5 (Web-CATI) - Age 4/pre-school- 2014/2015

Overview

Between the second and third face-to-face sweeps of data collection with Birth Cohort 2 of the Growing Up in Scotland study (GUS) a web and telephone based sweep was carried out with a parent/carer of the cohort child. This document contains details about the 'Sweep 2.5' web-CATI survey only. For details about sweep 2 and the wider GUS study, see related documentation available elsewhere on the UKDS website:

- http://doc.ukdataservice.ac.uk/doc/7432/mrdoc/pdf/7432_gus_bc2_sw2_user_guide.pdf
- http://doc.ukdataservice.ac.uk/doc/7432/mrdoc/pdf/7432_gus_bc2_sw2_data_documentation.pdf

Methodology

This sweep utilised a significantly shorter questionnaire than that used for face-to-face data collection, taking participants no longer than 15 minutes to complete either online or by telephone.

Respondents were the main carers of children who were members of 'Birth Cohort 2' (BC2). BC2 is comprised of a nationally representative sample of 6127 children living in Scotland when they were 10 months old at the time of the first interview in 2011. At the time of this data collection, children in the cohort were aged 4 years old and most were in nursery or pre-school.

All parents received an advance letter informing them about the survey. The letter highlighted that they would be asked to complete a short, online questionnaire (in contrast with previous face-to-face interviews). Parents were asked to check and update their contact details.

Once the online questionnaire was live, all respondents were sent a further letter and an email (where an email address was available) informing them that the questionnaire was ready to be completed, how to complete it (including providing a unique access code) and providing further information about issues such as confidentiality. Emails included a unique hyperlink which allowed parents to directly access the online questionnaire.

Reminders were sent one week after the online questionnaire went live, via email where this information was available and via post where not. After another week, non-respondents were contacted via email where available and via telephone where not. Where telephone contact was made, parents were either asked to complete a telephone interview or prompted to complete the online questionnaire.

Fieldwork and response

The sweep 2.5 web-CATI survey with BC2 was launched in January 2014. The fieldwork was conducted from January 2014 to February 2015. The sample was issued on a monthly basis, as in the face-to-face interviews.

The total number of cases issued was 5807. This resulted in 3237 productive interviews, representing a response rate of 56%. Participants were first invited to complete the questionnaire online and if they did not complete it were then contacted to complete a telephone interview. Of those respondents that completed an interview, 54% (1755) completed the questionnaire online via web (914) or smartphone (841) whilst the remaining 46% (1482) took part in a telephone interview.

Overall there were 2570 unproductive cases. Of these, 9% (227) were refusals. A further 57% (1467) of the unproductive cases were a result of non-contact: there were 529 cases where there was no direct contact with the named respondent, 430 cases where the telephone number had been disconnected, 402 cases were there was no answer, and a further 106 cases where the named respondent was not known at that telephone number.

A full breakdown of productive and unproductive cases is shown in the table below:

Table 1 – Fieldwork outcomes

	Total
Issued cases	**5807**
Productive	**3237**
Full telephone	1427
Full web	904
Full smart phone	826
Partial telephone	55
Partial web	10
Partial smart phone	15
Response rate (%)	**56**
Unproductive	**2570**
No direct contact with named respondent	529
Telecommunication barriers (e.g. Technical phone problems/ Number disconnected/ no answer)	844
Broken appointment	350
Refusals	227
Child no longer living in Scotland	13
Other	607

Using the data

The GUS Sweep 2.5 data consists of the following SPSS file:

GUS_BC2_ SW2.5.sav	3237 cases	Birth Cohort 2

The data file contains questionnaire variables (excluding variables used for administrative purposes). 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. Please note that variable descriptions in the variable list cannot be relied upon to capture the detail of the question wording, or the answer categories used. For the precise question wording, please refer to the questionnaire documentation. A copy of the questionnaire is provided alongside these notes. All routing is outlined in the questionnaire documentation. Note that in some cases only a partial interview was achieved.

The questionnaire focusses on topics related to pre-school and childcare arrangements, including any impacts on family life.

Respondent ID

As standard, the survey invitation was issued to the parent or carer who completed the previous GUS interview. For cases completed online, the data contains a variable which flags whether the respondent at sweep 2.5 is the same respondent s at the previous sweep. For telephone cases there is no equivalent flag. However, interviewers were instructed to interview the same parent or carer wherever possible.

In most cases the parent or carer invited to take part was the parent/carer who took part at sweep 2. Where the family did not take part at sweep 2, the parent/carer who took part at sweep 1 was invited to take part. Household IDs for each person in the household remain static over the course of the study. To help users match in data previously collected about the parent/carer who took part at previous sweeps, the sweep 2 respondent ID variable (DcRespID) is provided in the sweep 2.5 dataset. At sweep 1 all parent/carers who took part was assigned respondent ID=2. Therefore, no variable for sweep 1 respondent ID is provided.

Time spent in pre-school

Part-way through the fieldwork period (July 2014), the number of funded pre-school hours was increased (from 475 to 600 hours) in Scotland[1]. In response to this policy change, questions about how much time the child spent in pre-school were also amended. Before the policy change parents were asked to record how much time the child spent in pre-school to the nearest half hour. After the policy change parents were instead asked to record this to the nearest 10 minutes.

A flag (dateflag2) is provided in the dataset to identify whether data on pre-school attendance were collected before or after this change.

[1] At the time these data were collected (2014), the term 'pre-school education' was the dominant term used to describe this provision, rather than the current terminology of 'early learning and childcare'. To ensure consistency across the data and documentation, we use the term 'pre-school' here.

The dataset contains different sets of variables for how much time the child spends in pre-school:

- The original unedited variables (WdPRhrmo to WdPRhrsu). Fractions mean different things for cases who took part pre- and post the policy change.
- Edited variables for data collected pre-policy change only (DdPRhrmo1 to DdPRhrmo1) and for data collected post-policy change only. The post-policy change variables have been derived so that fractions match those used pre-policy change (i.e. .5 denotes half an hour) (DdPRhrmo2 to DdPRhrsu2)
- Variables which combine pre-and post-change data in a comparable format (i.e. .5 denotes half an hour) (DdPRhrmoA to DdPRhrsuA)
- Weekly totals for data collected pre-change (DdPreTot), post-change (DdPostTot) and combined in a comparable format (DdHrsTot).

Follow-up questions about how many of these hours were funded follow a similar pattern, i.e. where data were collected pre-change, parents were asked to answer to the nearest half hour (WdPRhrf), while for those completing post-change, the instruction was to provide details to the nearest 10 minutes (WdPRhrf2). No edited versions have been provided for these variables. However, users are referred to the syntax file for examples of how comparable variables may be derived.

See the questionnaire documentation for further details.

Weighting the data

Overview

Two weights were developed for Sweep 2.5 of BC2. These two weights were generated for analysis of information collected during the Web-CATI survey with the main carer.

The two weights were:

- A cross-sectional weight for adults that should be used for any cross-sectional analysis of data collected in the Sweep 2.5 web survey. All main carers that responded at Sweep 2.5 have a cross-sectional adult weight.
- A longitudinal weight for analysis of main carers that have responded at every previous sweep of BC2 of GUS.

The Sweep 2.5 survey follows up all main carers who responded at the Sweep 2 interview and gave NatCen permission to be re-contacted. In addition, main carers who had refused the Sweep 2 interview but had responded at Sweep 1 were contacted if they had given a ‘sweep only’ refusal at Sweep 2.

Weights for main carer interview data

The Sweep 2.5 sample of adult respondents can be split into two groups. For the purposes of describing the weighting these have been named Sample A and Sample B and are defined as follows:

- Sample A – adults who had responded at all previous sweeps
- Sample B – adults who had responded at Sweep 1 but had missed the interview in Sweep 2.

The two samples were treated separately during the weighting. This is because the Sample B respondents are likely to have different response behaviour to those in Sample A, as suggested by their much lower response rates. There were 753 individuals in Sample B, 86 (11%) of which responded at Sweep 2.5. The response rate for Sample A (4,837) was much higher at 65%. The issued and responding sample sizes are given in Table 2.

Table 2 Response rates for the two groups of main interview respondents

	Issued	Responding	Response rate
Sample A	4,837	3,152	65%
Sample B	753	86	11%
Combined (A+B)	5,590	3,238	58%

Two sets of weights were developed for the responding adults: a cross-sectional weight and a longitudinal weight. Only members of Sample A (who have responded at every sweep of GUS) received a longitudinal weight. This weight is described in more detail in Section 2.1.

All Sweep 2.5 respondents will have a cross-sectional weight (Sample A + B). These are described in more detail below.

Longitudinal weights for main carer interview data

Longitudinal weights were only generated for respondents in Sample A. A model-based weighting technique was used to develop the Sweep 2.5 longitudinal weights, where response behaviour is modelled using data from previous Sweeps. This is the same method used to generate weights for adults who completed the main interview at Sweep 2. Ineligible households (deadwood) were not included in the non-response modelling.

Response behaviour was modelled using logistic regression. This models the relationship between an outcome variable (in this case response to Sweep 2.5) and a set of predictor variables. The predictor variables were a set of socio-demographic individual and household characteristics collected from the previous sweeps of the study (mainly from Sweep 2).

The model generated a predicted probability of response for each individual. A set of non-response weights were generated equal to the inverse of these predicted probabilities; hence respondents who had a lower than average predicted probability received a higher than average weight, increasing their representation in the sample.

Variables found to predict response at Sweep 2.5 are shown in Table 3. All of them were entered in the non-response model which was used to calculate the non-response weights.

Table 3	Variables used in adult non-response weighting (longitudinal sample)
Number of adults 16+ in household	
Number of children in household	
Highest education level of respondent	
Respondent's religion	
Respondent's age	
Child's general health	
Type of accommodation	
Last known tenure	
Urban or rural classification of the postcode	
SIMD 2012 quintile	
Mother's age at cohort child's birth	

The final Sweep 2.5 weight was calculated as the product of the non-response weight and the Sweep 2 interview weight. The final weights were scaled to the responding Sweep 2.5 sample size, so that the weighted sample size matches the unweighted sample size.

Cross-sectional weights for main carer interview data

Cross-sectional weights were generated for all respondents at Sweep 2.5 (the combined A and B samples) and should be used for any cross-sectional analysis of Sweep 2.5 data.

Calibration weighting was applied to the combined sample to create the cross-sectional weights. This method adjusts a set of starting weights using an iterative procedure so that they match pre-defined population totals. The resulting weights, when applied to the combined data, will make the survey estimates match the population estimates which in this instance were calculated from Sample A, weighted by the longitudinal weight. Since the longitudinal weight corrects for sampling error and non-response bias at each stage of GUS, the weighted Sample A estimates are the best population estimates available.

The choice of the variables used in the calibration was dictated by the differences remaining after the Sweep 2.5 longitudinal weights were applied to Sample A and the cross-sectional weight from the last completed sweep for Sample B. The variables used in the weighting are listed in Table 4.

Table 4	Variables used in calibration of the adult cross-sectional sample
Mother's age at cohort child's birth	
Last known tenure	
SIMD 2012 quintile	

The calibration adjusts for any differences due to differential non-response between Sample A and Sample B.

Sample efficiency of main carer interview data

Weighting affects the statistical efficiency of a sample: the more variable the weights the larger the variance of the (weighted) survey estimates. More variable weights will result in larger standard errors and wider confidence intervals, so there is less certainty over where the "true" population values lie.

The precision of weighted survey estimates is indicated by the effective sample size (neff) which measures the size of an (unweighted) simple random sample that would provide the same precision (standard error) as the weighted sample. The efficiency of the weights 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.

Table 5 Range of adult weights and sample efficiency

	Min	Max	Mean	N	Neff	Efficiency
Main carer longitudinal weight	0.37	5.23	1	3,152	2,191	70%
Main carer cross-sectional weight	0.38	5.36	1	3,238	2,267	70%

3. Applying the weights

For each sample, the cross-sectional weights should be used for any cross-sectional analysis, i.e. any analysis of Sweep 2.5 data only. All sample members that responded at Sweep 2.5 have a cross-sectional weight.

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

Table 6 Description of weight variables in the data file

Variable name	Label
WdWTbrth	Wd sweep 2.5 crossectional weight
WdWTbth2	Wd sweep 2.5 longitudinal weight

Contact details

Jackie Palmer, Data Manager: jackie.palmer@scotcen.org.uk

Paul Bradshaw, Project Director: paul.bradshaw@scotcen.org.uk