ScotCen
Social Research that works for society

Survey weights and longitudinal analysis

Summary findings from a Growing Up in Scotland Working Paper

Growing Up in Scotland

Klaudia Lubian, Sarah Tipping and Paul Bradshaw
ScotCen Social Research, February 2020

At **ScotCen Social Research** we believe that social research has the power to make life better. By really understanding the complexity of people's lives and what they think about the issues that affect them, we give the public a powerful and influential role in shaping decisions and services that can make a difference to everyone. And as an independent, not for profit organisation we're able to put all our time and energy into delivering social research that works for society.

ScotCen Social Research
Scotiabank House
6 South Charlotte Street
Edinburgh, EH2 4AW
T 0131 240 0210
www.scotcen.org.uk

A Company Limited by Guarantee. Registered in England No.4392418.
A Charity registered in England and Wales (1091768) and Scotland (SC038454)
This project was carried out in compliance with ISO20252

Contents

1 Introduction and key findings

1.1 Introduction

Within a chapter you'll have a number of sections and sub-sections. For numbered sections and sub-sections use Heading 2 for second level and Heading 3 for third level numbering.

The Growing Up in Scotland study (GUS) is a longitudinal research project which tracks the lives of thousands of children and their families in Scotland from the early years, through childhood and beyond. The main aim of the study is to provide new information to support policy-making in Scotland, but it is also intended to provide a resource for practitioners, academics, the voluntary sector and parents.

To date, the study has collected information about three nationally representative cohorts of children: a child cohort and two birth cohorts. Altogether, information has been collected on around 14,000 children and families in Scotland.

This report provides a summary of the methods and findings of a review into the weights which are generated to support analysis of data on the first birth cohort (Birth Cohort 1 or BC1). Specifically, the implications of using the weights created for longitudinal analysis – analysis involving multiple sweeps of study data.

The current recommended approach for longitudinal analysis of GUS data is to apply the longitudinal weights which have been constructed for each sweep. More specifically, to apply the longitudinal weight for the latest sweep included in the analysis. Longitudinal weights are derived for households where the cohort member's main carer took part in all previous sweeps. For example, the longitudinal weight at sweep 9 has been derived only for those cases which have participated at all of sweeps 1 to 8. As a result, the sample size for analyses is considerably reduced in many situations when these weights are applied. However, very often the analysis requires using variables from just a few previous sweeps. As many cases are only missing from the study for single sweeps – but as a result are excluded from longitudinal weights following that sweep – having a means of undertaking longitudinal analysis that incorporates those cases would potentially improve the utility of the data.

The review was undertaken to find an alternative or suggest improvements to the current approach to longitudinal analysis using GUS data. The aim was to suggest potential solutions to address the issue of decreasing sample sizes for longitudinal analyses caused by increasing attrition and individual sweep non-response.

Three main solutions were considered by the review:

1. Creating additional longitudinal weights
2. Using cross-sectional weights instead of longitudinal weights
3. Not using weights at all

This report provides a summary of findings along with practical recommendations for those analysing the data.

In addition, some consideration is given to a further three issue presented by GUS data as the study continues:

1. Are separate weights for analysis of questions asked of main carers and children needed for analysis of the data from past sweeps? And will they be needed for analysis of data from future sweeps?

2. How should the boost sample be incorporated in longitudinal analyses?

3. How do we allow for comparisons of questions asked of children or main carers at future sweeps with questions asked of the cohort child's Primary 6 teacher at sweep 8?

1.2 Key findings and recommendations

- For population estimates and descriptive statistics we strongly recommend using weighted data.
- An additional longitudinal weight for analysis of cases where any of sweeps 2-6 were missing would offer a significant boost to the sample size for analysis of data from sweeps 1, 7, 8 and 9 or 1, 8 and 9, relative to using the sweep 9 longitudinal weight.
- For analysis of two 'paired' sweeps of data, or bespoke combinations of some but not all sweeps for analysis, the cross-sectional weight for the latest sweep could be applied instead of the longitudinal weight, provided that the individuals who have responded to the selected sweeps make up a substantial proportion of the cross-sectional sample from the latest sweep included in the analysis.
- The use of the cross-sectional weight for analysis of two paired sweeps or bespoke combinations of some but not all sweeps for analysis is not recommended when sweep 9 or 10 are the latest sweeps. This is because the cross-sectional weight for sweeps 9 and 10 are computed for the combined main and boost sample. A standalone, main sample only cross-sectional weight would be required for these sweeps to permit this approach.
- The sweep 9 and 10 cross-sectional weights cannot be used for separate analysis of the main (or boost) sample
- For multivariate analysis, the decision to weight or not depends on a number of factors. Analysts need to test the impact of the weights in a model to ascertain whether the model suffers if the weights are not included.
- The impact of non-response on unweighted multivariate analysis may also be controlled by incorporating variables used for compiling the weights in the multivariate model alongside other predictors. This requires a careful approach. If time does not allow for this, we recommend that non-response weights are used to address the impact of attrition on the GUS sample.
- It is possible to use the main carers' longitudinal weight for the analysis of longitudinal child data without risking the child data being significantly biased.
- In order to conduct longitudinal analysis of data from all children, including boost cases, from sweep 9 onwards, a new longitudinal weight starting at sweep 9 needs to be computed.
- For longitudinal analysis of bespoke combinations of sweep 9 with earlier years, a 'main sample' sweep 9 cross-sectional weight will be required. This would involve creating a cross-sectional weight for all main (non-boost sample) cases that responded at sweep 9.
- For analysis of data from future sweeps and questions asked of a teacher at sweep 8, a new longitudinal weight is recommended.

2 Background to the issue

2.1 Summary of the existing weighting approach

A detailed description of the methods applied to create the GUS survey weights is included in the User Guide which accompanies each sweep's dataset. User Guides can be accessed via the GUS website1 or from the documentation tab of the GUS Cohort 1 data catalogue entry on the UK Data Service website (where the datasets themselves can also be requested).2 Here we provide an overview of the weighting methodology used at sweeps 1, 2 and 9 of BC1. The weighting approach at the intermediate sweeps was similar to the approach used in sweeps 2 and 9.

Sweeps 1 and 2

Calculation of the survey weight at sweep 1 was undertaken in two stages:

- Stage 1: calculation of selection weights which accounted for the selection of one child in households where there were twins (or other multiple births) or children eligible for both the birth and child cohorts.
- Stage 2: calculation of non-response weights.

At stage 2, there was a limited amount of information available on the sampling frame (Child Benefit records) with which to build a non-response model. Therefore, the variables which were available were enhanced with the addition of area-level information using the postcodes of sampled families.

At sweep 2, a model-based weighting technique was used. 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 was used to model their response behaviour at sweep 2. Non-response behaviour was modelled using logistic regression. 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. 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 final sweep 2 weight is the product of the sweep 2 non-response weight and the sweep 1 weight.

Sweep 3 onwards

From sweep 3 onwards, both cross-sectional and longitudinal weights were developed for analysis of parent/carer data. These differentiate between those cases where a parent/carer participated at the current and every prior sweep (Sample A) and those who participated at the current sweep but missed at least one prior sweep (Sample B) on the basis that the characteristics of people who participate at all sweeps are different to those whose participation is less than 100%. For main carers in both

[1] https://growingupinscotland.org.uk/using-gus-data/data-documentation/
[2] https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=5760

samples A and B, response behaviour was modelled using logistic regression following the approach outlined above. The modelling uses individual and household characteristics collected from the previous sweeps of the study as predictor variables.

In Sweep 7, for the first time in GUS, children were asked to fill in a short self-completion questionnaire. Almost all children completed the questionnaire - 97% of those whose main carer completed an interview. Both cross-sectional and longitudinal weights are also created for analysis of child data.

Longitudinal child weights are generated for cases where the main carer had responded at every prior sweep of GUS. These weights are constructed using calibration weighting. 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, make the survey estimates match the population estimates which in this instance were calculated from Sample A, weighted by the main carer's longitudinal weight. Since the longitudinal weight for main carer interview data corrects for non-response bias at each stage of GUS, the weighted Sample A estimates are the best estimates available for children from the cohort from which Sweep 1 was sampled who remain in Scotland.

2.2 Why weighting the data is important

The GUS weights adjust the sample to correct for unequal selection probabilities and differential non-response, thereby reducing both selection bias and non-response bias. They make the sample representative of the target population of children living in Scotland. Unweighted estimates would underestimate key groups such as the lone parent families and those living in deprived areas. The experiences and outcomes of these groups would therefore be under-represented in the sample, leading to biased estimates.

The tables below show weighted and unweighted estimates of child obesity and SDQ total difficulties score by family type (lone parent/couple family). The estimates are taken from sweep 9. It can be seen that the proportion of lone parents is slightly lower in the unweighted cross-sectional sample than the weighted sample (22% of the unweighted sample and 23% of the weighted sample, Table 1), and considerably lower in the unweighted longitudinal sampled compared to the weighted longitudinal sample (18% of the unweighted sample and 23% of the weighted sample, Table 2). This is because response rates were lower for lone parents, leading to them being under-represented in the sample. The weights address this by 'weighting up' lone parents.

Table 1 shows the impact of the weights on two key survey outcomes in the cross-sectional sample: child obesity and SDQ total difficulties score. The children of lone parents are more likely to be outside the healthy weight range and to have higher SDQ total difficulties scores. Thus unweighted estimates – which under-represent lone parents – under-estimate the proportion of children outside the healthy weight range and under-estimate the mean SDQ total difficulties score. Table 2 shows the corresponding figures for the longitudinal sweep 9 sample. The same pattern is seen but the impact of the weights on the SDQ total difficulties score for the longitudinal sample is much stronger than for the cross-sectional sample; the longitudinal sample is more affected by the attrition bias (and it does not include the boost sample added at sweep 9).

Table 1 GUS Sweep 9 cross-sectional sample, comparing weighted and unweighted estimates of child obesity and SDQ total difficulties score by family type

Family type	Proportion of the sample		Key outcomes	
	Unweighted	Weighted by cross-sectional main carer weight	% of children outside healthy ISD range (unweighted)	SDQ total difficulties score (unweighted)
Lone Parent	22%	23%	46%	10.14
Couple	78%	77%	34%	7.09
Base (unweighted)	3418			
Overall estimates			**Key outcomes**	
			% of children outside healthy ISD range	SDQ total difficulties score
Unweighted estimate			36%	7.74
Weighted estimate (weighted by cross-sectional main carer weight)			38%	8.02

Table 2 GUS Sweep 9 longitudinal sample, comparing weighted and unweighted estimates of child obesity and SDQ total difficulties score by family type

Family type	Proportion of the sample		Key outcomes	
	Unweighted	Weighted by longitudinal main carer weight	% of children outside healthy ISD range (unweighted)	SDQ total difficulties score (unweighted)
Lone Parent	18%	23%	47%	9.30
Couple	82%	77%	32%	6.71
Base (unweighted)	2528			
Overall estimates			**Key outcomes**	
			% of children outside healthy ISD range	SDQ total difficulties score
Unweighted estimate			36%	7.16
Weighted estimate (weighted by longitudinal main carer weight)			38%	8.02

The impact of the weights therefore depends on the relationship between the key outcomes and non-response behaviour; if the key outcome is related to a characteristic that is affected by non-response bias and therefore under-represented in the sample, then it will be affected by the weights. If there was no relationship between the key outcome and response behaviour then we would expect weighted and unweighted estimates to be the same, although this is very rare in social sciences.

For population estimates and descriptive statistics we strongly recommend using weighted data. There is a general consensus amongst survey practitioners as to the need for weights for such estimates.

2.3 Attrition rates and impact on analysis

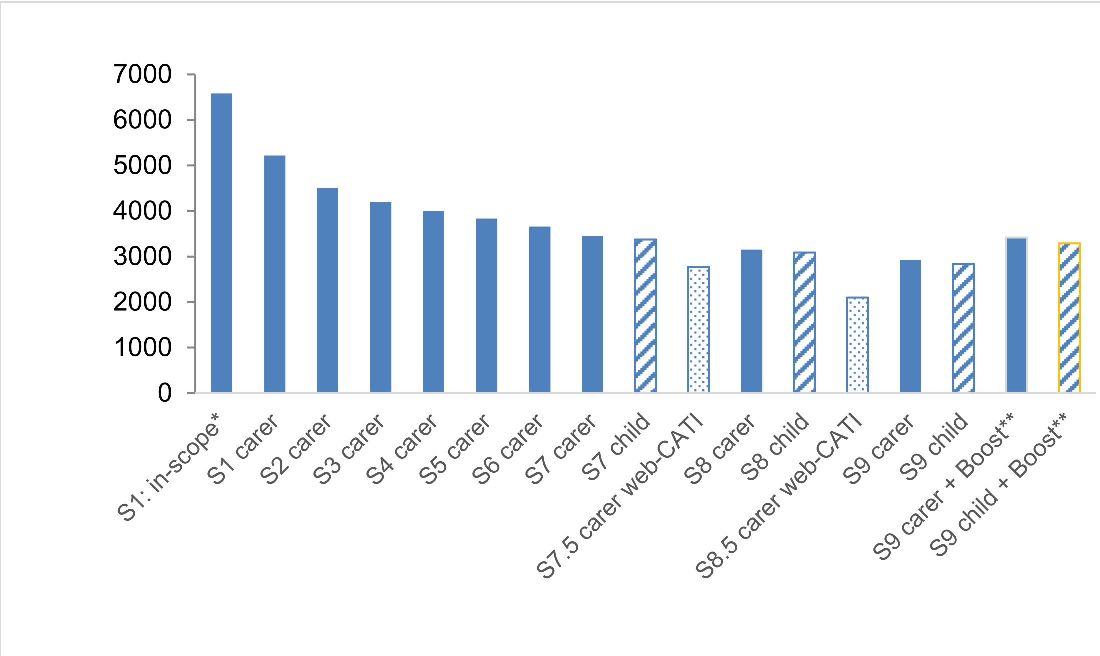
The GUS sample was drawn from Child Benefit records held by the DWP[3]. Across the geographic clusters sampled, at total of 8218 eligible children were identified. A

[3] Further details on the sample design can be found in the sweep 1 User Guide.

number of cases were removed by DWP and others were deemed to be ineligible or out of scope, leaving 6583 ‘in-scope’ cases for sweep 1 fieldwork. Of these, an interview was achieved with 5217/80%.

Figure 1 presents sample sizes for all BC1 sweeps (both main carer and child interviews) compared with the number of *achievable/’in-scope’* cases from sweep 1. The biggest drop in response rates among main carers occurred at sweep 1 (79% of achievable cases responded), followed by quite a significant drop at sweep 2, by 10 percentage points compared to sweep 1 (69% of achievable cases from sweep 1 responded). At sweep 9, 44% of achievable cases from sweep 1 participated. Sweep 9 was the first BC1 Sweep to include a boost sample. The addition of the boost sample resulted in a total sample size of 3,418 main carers responding at sweep 9.

Figure 1 Number of cases in cross-sectional samples in all BC1 sweeps compared with number of 'in-scope' cases at sweep 1

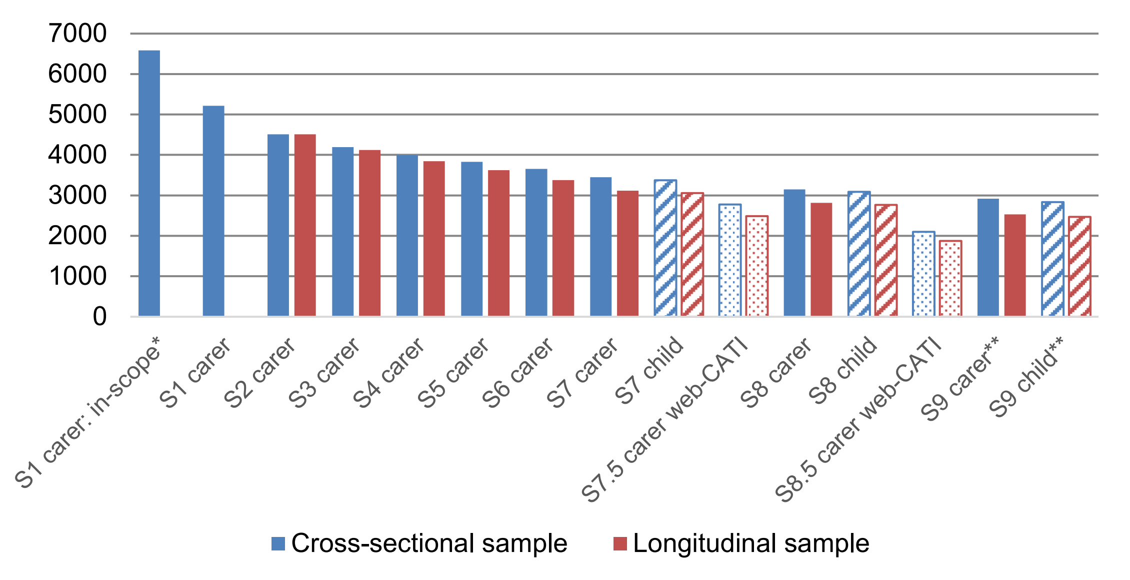


* 7252 BC1 children were first issued to field. Cases which were considered out-of-scope or unachievable were mostly ineligible or incorrect addresses.
** Sweep 9 was the first BC1 Sweep to include a Boost Sample. Only cross-sectional weights were computed for the main carers and children from the Boost Sample.

Figure 2 compares the cross-sectional sample sizes (presented in Figure 1) with longitudinal sample sizes – i.e. the number of cases that have responded at all preceding sweeps including the one of interest.

The longitudinal sample size for sweep 2 is the same as the cross-sectional sample size, but from sweep 3 onwards we can observe a difference between the two. This difference is due to respondents that had missed one or more intervening sweeps. For example, although 3822 main carers responded at sweep 5, only 3621 are included in the longitudinal sample - 212 missed one or more of sweeps 2, 3 or 4. The difference increases from 73 at sweep 3 to 389 cases at sweep 9 for main carer interviews and from 317 at sweep 7 to 365 at sweep 9 for child interviews.

Figure 2 Number of cases in cross-sectional and longitudinal samples in all BC1 sweeps compared with number of 'in-scope' cases at sweep 1



*7252 BC1 children were initially issued to field. Cases which were considered out-of-scope or unachievable were mostly ineligible or incorrect addresses.
**Sweep 9 was the first BC1 Sweep to include a Boost Sample. Longitudinal weights could not be computed for cases from the Boost Sample, therefore the cross-sectional sample sizes presented for Sweep 9 are coming from the Main Sample.

Table 3 presents how the longitudinal samples reduced over time. Almost 40% of eligible cases issued to the first sweep are lost by sweep 3. By sweep 6 almost 50% are lost. Despite the fact that the cross-sectional and longitudinal samples at sweep 9, (taking only main sample cases into account) make up similar proportions of the overall sweep in scope cases (i.e. 43% v 38%,Table 3), the accumulative attrition affects the sample in a different way to intermittent attrition. This is because the types of people who drop out entirely are likely to be different to the types of people who drop out just for a wave or two, which is why the profile of the sweep 9 longitudinal sample may be different from the cross-sectional sample.

Currently, it is recommended that for any longitudinal analyses the longitudinal weight prepared for the most recent sweep included in the analysis is applied. This means that only cases which have taken part in every sweep of GUS up to and including the latest sweep which is of interest will be included the analysis. Depending on the analysis being undertaken and the sweeps being used, this can result in the ‘loss’ of potentially valid cases by applying the longitudinal weight. For example, analysis which predominantly uses data from sweeps 5 and sweep 8 would apply the sweep 8 longitudinal weight resulting in 2726 children being included in the analysis. This means that even though a higher number of cases replied to both sweeps 5 and 8, the base is limited to the 2726 case which have taken part in every sweep of GUS up to sweep 8.

Table 3 Number of cases in cross-sectional and longitudinal samples in all BC1 sweeps compared with number of 'in-scope' cases at sweep 1

Sweep	cross-sectional sample size	% of S1 'in-scope'	longitudinal sample size	% of S1 'in-scope'	difference
S1 carer: in-scope*	6583	100.0%			
S1 carer	5217	79.2%			
S2 carer	4512	68.5%	4512	68.5%	0
S3 carer	4193	63.7%	4120	62.6%	73
S4 carer	3994	60.7%	3844	58.4%	150
S5 carer	3833	58.2%	3621	55.0%	212
S6 carer	3657	55.6%	3375	51.3%	282
S7 carer	3453	52.5%	3119	47.4%	334
S7 child	3374	51.3%	3057	46.4%	317
S7.5 carer web-CATI	2775	42.2%	2487	37.8%	288
S8 carer	3149	47.8%	2815	42.8%	334
S8 child	3088	46.9%	2765	42.0%	323
S8.5 carer web-CATI	2099	31.9%	1874	28.5%	225
S9 carer**	2917	44.3%	2528	38.4%	389
S9 child**	2834	43.1%	2469	37.5%	365

*7252 BC1 children were initially issued to field. Cases which were considered out-of-scope or unachievable were mostly ineligible or incorrect addresses.
**Sweep 9 was the first BC1 Sweep to include a Boost Sample. Longitudinal weights could not be computed for cases from the Boost Sample, therefore the cross-sectional sample sizes presented for Sweep 9 are coming from the Main Sample.

3 Potential alternative approaches to increase sample size in longitudinal analyses

The smaller sample size created by use of the longitudinal weight limits the ability to conduct more in-depth analyses (users need to combine smaller categories of some variables) or to include all variables that are necessary.

In this section we consider three potential alternative approaches to longitudinal analyses which may improve the sample available:

- Construction of a second longitudinal weight
- Longitudinal analysis using cross-sectional weights
- Longitudinal analysis on unweighted data

3.1 The merit of a second longitudinal weight

One of the options for dealing with the reduced sample sizes created by applying the longitudinal weight is to consider a second longitudinal weight starting from one of the later sweeps rather than sweep 1. In sections that follow, we aim to answer the following questions:

- How many cases are lost due to the use of longitudinal weights when analysis does not require data from all preceding sweeps?
- Is there a merit in computing a second longitudinal weight for such analyses, and if so, which cross-sectional weight would be best to serve as a starting weight for the second longitudinal weight?

To help answer these questions, in the sections that follow we present results of simulation of sample sizes available for longitudinal analysis if certain sweeps are allowed to be missed. The analysis was conducted only for main carer data.

3.1.1 Simulation of sample size for longitudinal analyses with first few sweeps missing

In this section we present the simulation of sample sizes for longitudinal analysis that relaxes the requirement of participation in all preceding sweeps. In other words, we check how many more cases would be available for analysis of responses of cases that participated at sweeps 1, 4 and 5 (rather than 1, 2, 3, 4 and 5 – as currently assumed).

Table 4 presents sample sizes that would be available for longitudinal analysis at each sweep if we allow respondents to miss the following individual or combinations of sweeps: sweep 2, sweeps 2 and 3, sweeps 2-4, sweeps 2-5, sweeps 2-6 and sweeps 2-7. The last scenario assumes analysis would include cases that responded at sweeps 1, 8 and 9.

Table 5 presents percentage increase in sample sizes available for longitudinal analysis (from Table 4) when compared with sample sizes available at each sweep with the use of current longitudinal weights.

Table 4 Simulation of longitudinal sample sizes if one or more consecutive sweeps can be missed

Sweep	responded at all sweeps	could miss S2	could miss S2-S3	could miss S2-S4	could miss S2-S5	could miss S2-S6	could miss S2-S7
3	4120						
4	3844	3895					
5	3621	3665	3739				
6	3375	3411	3470	3536			
7	3119	3150	3194	3243	3323		
8	2815	2843	2883	2922	2979	3070	
9	2528	2554	2584	2608	2649	2704	2760

Table 5 Percentage increase in sample sizes available for longitudinal analysis if one or more consecutive sweeps can be missed

Sweep	responded at all sweeps	could miss S2	could miss S2-S3	could miss S2-S4	could miss S2-S5	could miss S2-S6	could miss S2-S7
3	4120						
4	3844	1.3%					
5	3621	1.2%	3.3%				
6	3375	1.1%	2.8%	4.8%			
7	3119	1.0%	2.4%	4.0%	6.5%		
8	2815	1.0%	2.4%	3.8%	5.8%	9.1%	
9	2528	1.0%	2.2%	3.2%	4.8%	7.0%	9.2%

The increases in sample sizes for longitudinal analysis range from 1% to 9.2% relative to sample sizes available at each sweep with the use of current longitudinal weights. The biggest gain could be achieved from computing a new weight for analysis of responses of main carers who responded at sweeps 1, 8 and 9 where analysis could be done with 2760 cases instead of the 2528 available if the sweep 9 longitudinal weight is applied (an increase of 9.2%).

For the analysis of past sweeps, there seems to be a merit in computing a new weight starting with cross-sectional weight from sweep 7, i.e. for analysis that does not require questions from sweeps 2-6. The difference of 9% (255 cases) in sample size available for analysis of responses from sweeps 1, 7 and 8 can positively affect the ability to conduct more in-depth analysis. Adding questions from sweep 9 would reduce the gain to 7% (176 cases), which is still a considerable boost to sample size.

The sample sizes presented in tables 7 and 8 are absolute/actual sample sizes, not effective sample sizes that take into account the effect of the sample design on the precision of survey estimates and extent of bias removed with weighting. It is challenging to estimate effective sample sizes for the simulated samples, as they can be computed using actual weights only. However, if a new weight was created for longitudinal analysis of data from later sweeps, e.g. sweeps 7, 8 and 9, it is likely that its efficiency (i.e. effective sample size divided by actual sample size) would be higher than the currently used longitudinal weight. This is because the cross-sectional weight from sweep 7 would be used as the starting point, while the existing longitudinal weight is computed by removing the bias that stays after weighting the model by the longitudinal weight from the previous sweep.

3.1.2 Simulation of sample size for longitudinal analyses with non-consecutive sweeps missing

Recently, a growing number of longitudinal analyses conducted with GUS data starts from later sweeps and utilise responses from non-consecutive sweeps, e.g. combinations of responses from sweeps 5 and 7 or sweeps 5 and 9. Longitudinal sample sizes for a selection of such combinations have been calculated and presented in Table 6.

The calculations show by how many (+ in percentage terms) is the absolute sample size for longitudinal analysis increased when compared to the sample size available using the current longitudinal weight. For example, there are 3373 main carers that responded to sweeps 1, 3 and 7, whilst longitudinal weight computed for sweep 7 is available for 3119 cases. If analysis is conducted with 3373 cases this would mean an increase of 8.1% compared to currently available longitudinal sample.

The figures in the table suggest that if an additional longitudinal weight was to be computed to enable including a significant additional number of cases for longitudinal analysis of both consecutive sweeps (scenarios discussed in section 2.2.4) and non-consecutive sweeps (section 2.2.5), it would be **a weight for analysis of respondents that could have missed any of sweeps 2-6**. This would offer a significant boost to the sample size for analysis of data from sweeps 1, 7, 8 and 9 or 1, 8 and 9 (see table 7), while limiting loss of cases to only 138 if analysis required data from sweeps 7 and 9 only (2842 – 2704) compared to 314 cases (2842 –2528) that would be discarded if the current longitudinal weight for sweep 9 was used.

Table 6 Percentage increase in sample sizes and absolute sample size available for longitudinal analysis if data from selected sweeps is used

Sweep	*N* responded at all sweeps	S1, S3, S5	S1, S3, S7	S1, S3, S9	S1, S5, S9	S1, S7, S9
5	3621	3747 (3.5%)				
6	3375					
7	3119		3373 (8.1%)			
8	2815					
9	2528			2860 (13.1%)	2833 (12.1%)	2842 (12.4%)

3.2 Longitudinal analysis using cross-sectional weights

Researchers using GUS often require simple pairs of sweeps, or bespoke combinations of sweeps during analysis. The current longitudinal weight is only calculated for individuals who have responded to every wave of GUS. Thus when analysing paired or combinations of sweeps, applying the longitudinal weight means large numbers of interviews are discarded. In this section we investigate whether it would be appropriate to apply cross-sectional weights for the analysis of pairs or bespoke combinations of sweeps.

At each sweep, all responding cases in the sample are given a cross-sectional weight, calculated using the population totals estimated from the survey weighted by the longitudinal weights. Using the cross-sectional weights for cross-wave analysis will maximise the sample size available for the analysis of different combinations of sweeps. The larger sample should have particular benefits for sub-group analysis. The cross-sectional weight for the latest wave would be applied. Thus for analysis of sweep 5 and sweep 8, the sweep 8 cross-sectional weight would be applied and the analysis sub-sample would be all sweep 8 cases that also responded in sweep 5.

However, the cross-sectional weights are not designed for cross-wave analysis. They have been generated to ensure the cross-sectional sample (i.e. all individuals who respond to a given sweep) is generalizable to the target population. Individuals who have responded to two or more sweeps make up a sub-set of this sample. For the example above, there are 3149 responding cases at sweep 8 with a cross-sectional weight, 3048 of these cases (97%) also completed an interview in sweep 5. In this instance the paired sweeps make up a substantial proportion of the cross-sectional sample - only 3% of the cross-sectional sample is missing. **Applying the cross-sectional weights to such a large sub-set of the cross-sectional sample means the weighted sub-sample is still likely to be representative of the target population.**

The loss of cases will be greater if more than two sweeps are included, as the sub-set of the sample that is being used will be a smaller proportion of the overall sample. However, the proportion is still generally high for combinations of three sweeps. For example, applying sweep 8 cross-sectional weights to longitudinal analysis of cases that responded at sweeps 8, 3 and 5 reduces the available sample size to 95% of the cases responding to sweep 8.

Estimates from paired sweeps of data that are run using cross-sectional weights will have a larger sample size, and therefore smaller confidence intervals. However, estimates for the same paired sweeps that are run using longitudinal weights would be expected to contain less bias. There is therefore a trade-off to be made. This can be demonstrated by comparing estimates of child obesity and SDQ total difficulties scores from sweep 7, estimated using the sweep 7 cross-sectional weights, to the same estimate for sweep 9 respondents using both the sweep 9 cross-sectional and sweep 9 longitudinal weights. The results are presented in Table 7. Estimates of sweep 7 obesity and SDQ total difficulties score that are generated for sweep 9 respondents using the longitudinal weights are closer to the sweep 7 figures. The same estimates using the cross-sectional weights are still very close (within the upper and lower confidence limits for the longitudinal estimates) but have smaller confidence intervals and a larger sample size.

Table 7 Comparing estimates of child obesity and SDQ total difficulties score from sweep 7 using weights from sweep 7 and 9

	Sw7 estimate using Sw7 weights	Sw7 estimate using Sw9 cross-sectional weights	Sw7 estimate using Sw9 longitudinal weights
Study child weight outside ISD healthy range	26.5%	26.0%	26.3%

95% CI lower	24.9%	24.3%	24.3%
95% CI upper	28.2%	27.8%	28.3%
Base (unweighted)	*3362*	*2782*	*2481*
SDQ total difficulties score	7.86	7.71	7.98
95% CI lower	7.62	7.44	7.67
95% CI upper	8.10	7.97	8.30
Base (unweighted)	*3424*	*2825*	*2516*

This suggests that the cross-sectional weights are likely to be appropriate for analysis of up to three sweeps as long as the sub-sample makes up a substantial proportion of the cross-sectional sample. There is no recommended cut off point, the decision as to whether the cross-sectional weights are appropriate should be made on the basis of similar analysis as presented in Table 12. **However, such an option is not easily available for analysis of sweep 9 data as the cross-sectional weight for sweep 9 was computed for the combined main and boost sample.** If one of the sweeps analysed includes boost sample (e.g. when data from sweeps 9 and 7 is analysed), a new cross-sectional weight would need to be computed for the main sample of the most recent wave (excluding boost cases).

3.3 Longitudinal analyses without using weights

Whilst for population estimates and descriptive statistics the consensus generally is that using weighted data is preferable, there is some disagreement over the use of weights for multivariate analysis, such as regression modelling, particularly for more complex models, such as longitudinal models and multi-level models[4]. This is partially due to issues around estimation procedures for complex models that do not always run on weighted data, for example, when estimating maximum likelihood for complex regression models (although this is less the case as survey packages continue to improve and computational power increases).

The decision to weight or not depends on a number of factors: the nature of the outcome variable, the complexity of the model, and the variance structure of the data. It should be noted that these factors should routinely be considered when modelling causality, and not only considered in the context of weighting. Analysts need to assess whether the weights are causing or exacerbating any issues in the data.

There are ways of testing the impact of the weights in a model, and therefore ascertaining whether the model suffers if the weights are not included. This can be done by including the weighting variable as a covariate alongside the substantive variables in an unweighted model. The model should also include an interaction term between the weighting variable and each substantive variable. If the weighting variable and interactions are not significant it can be concluded that the weights are having little impact on the relationship between substantive variables and outcomes, meaning an unweighted model will correctly summarise the relationship between substantive variables and outcomes. If the weighting terms are proven to be significant, then the implication is that the relationship between the substantive variables and the outcome is affected by the weights. This means it is not appropriate to run a model on

[4] For example: https://www.nber.org/papers/w18859
http://www.stat.columbia.edu/~gelman/research/published/STS226.pdf, and
Kott, 2007

unweighted data without making an attempt to correct for non-response and selection bias.

An important aspect to consider is that weights will impact on different outcomes in different ways, depending on the relationship between the outcome and response behaviour. This means it may be acceptable to exclude weights for one analysis but not another. Thus, the above test would need to be carried out for each outcome, and a separate decision made each time. There is no one-size-fits-all approach.

An alternative to running a model on weighted data would be to run models on unweighted data but include any variables used to construct the weights as covariates in the model. The model would need to include the weighting variables as both main effects and as interaction effects with any substantive variables A model specified in this way should provide unbiased and consistent parameter estimates. It should also have smaller standard errors than the equivalent model based on weighted data. This approach works because it controls for the three-way relationship between the outcome variable (which is only known for respondents), the substantive variable (which is only known for respondents), and characteristics that are known to be affected by non-response bias (because auxiliary information exists that is available for both respondents and non-respondents). This final set of variables – the characteristics known to be affected by non-response – can be thought of as confounders; they are masking the 'true' relationship between substantive variables and outcomes, we therefore want to reduce their impact by controlling for them in the model. By including an interaction between the substantive variable and the characteristic with known response bias, we are mitigating for the way in which the relationship between the outcome and substantive variable may be affected by non-response.

The success of this approach depends on the model being correctly specified, with all weighting factors and interactions accounted for so that they may be tested. This may not be straightforward for GUS as it does not use a consistent set of variables to generate weights at each wave. Instead, a subset of variables is selected at each wave from a wider pool of potential covariates. The variables most strongly related to non-response are included and the remaining variables removed, leading to a parsimonious model that best reflects the nature of non-response at that particular wave. In addition, longitudinal weights are built up over time, with each weight incorporating the weight from the previous wave. This means a larger number of weighting variables would be required in the model should this approach be taken on GUS.

This approach is therefore feasible when there is a large sample size to allow the interaction effects to be estimated robustly and may therefore be less robust for sub-group analysis where there are fewer cases available.

Finally, it should be noted that statistical packages are constantly improving, and the computational power of computers is continually increasing. This means analysts are better able to handle weighting during complex analysis than even relatively recently. This point is pertinent as these improvements allow weights to be incorporated into analyses where they were previously viewed as inappropriate, or where the computing power required was too great.

We therefore recommend that non-response weights are used to address the impact of attrition on the GUS sample. This recommendation is in line with the advice for other major longitudinal surveys.

4 Considerations for analysis of future sweeps

As more data is collected and the analysis scenarios become potentially more complex, a number of additional issues are raised regarding the application of weights. We have identified three principle issues likely to arise and provided suggestions for how they may be handled:

- Are separate weights for analysis of questions asked of main carers and children needed for analysis of the data from past sweeps? And will they be needed for analysis of data from future sweeps?
- How should the boost sample be incorporated in longitudinal analyses?
- How do we allow for comparisons of questions asked of children or main carers at future sweeps with questions asked of the cohort child's Primary 6 teacher at sweep 8?

4.1.1 Are separate parent and child weights needed?

Longitudinal weights for the child interviews were generated using calibration weighting which made the survey estimates match the population estimates from the main carer longitudinal sample, weighted by the main carer's longitudinal weight. Since the longitudinal weight for main carer interview data corrects for non-response bias at each stage of GUS, the weighted estimates were assumed to be the best estimates available for children from the cohort from which Sweep 1 was sampled who remain in Scotland.

Over sweeps 7 to 9, 98% of children whose main carers responded to the CAPI survey participated in child interviews. Therefore, the bias corrected by the additional calibration weighting for child interviews has been always very small. At sweep 9 for example, where longitudinal data was available for 2469 children and 2528 main carers, the maximum absolute difference corrected for by the child weight was 0.5 percentage points, which is considered very small. This means that if the main carers' longitudinal weight was used for analysis of the children who have responded to all child surveys (sweeps 7, 8 and 9), the weighted sample would be off by a maximum of 0.5 percentage points from its 'population totals'. It is therefore possible to use the main carers' longitudinal weight for the analysis of longitudinal child data without risking the child data being significantly biased. This is certainly the case for sweeps 7, 8 and 9, hence if there is a need to compare questions asked of children at future sweeps with questions asked of main carers and children at earlier sweeps, the main carer weight can be used to combine parents and children data.

Thinking about different scenarios of longitudinal analyses, we can suggest the use of the following weights:

1. In order to compare questions asked of children at future sweeps with questions asked of main carers at earlier sweeps:
    - if there is a significant difference in the profile of households of responding children and responding main carers, a child longitudinal weight would need to be computed and used for analysis,
    - if the difference in the profiles of households of responding children and responding main carers turns out to be small, longitudinal weight for main carers can be used instead,

- If the analysis will compare data from a future sweep with sweep 9 data (the first sweep to include the boost sample), the cross-sectional weights from the most recent sweep may be appropriate (although analysis similar to that presented in Table 12 should be conducted);

2. In order to compare questions asked of children at future sweeps with questions asked of main carers and children at earlier sweeps (combined main carers and children data): see point 1.

3. In order to compare questions asked of main carers at future sweeps with questions asked of main carers at earlier sweeps, the longitudinal weight for main carers can be used. If the analysis will compare data from a future sweep with sweep 9 data, the cross-sectional weights from the most recent sweep may be appropriate (although analysis similar to that presented in Table 12 should be conducted).

4.1.2 Incorporating the boost sample in longitudinal analysis

In order to conduct longitudinal analysis of data from all children, including boost cases, a new longitudinal weight starting at sweep 9 needs to be computed. Children in the boost sample received only the cross-sectional weights as sweep 9 was the first wave to which they were invited and thus they could not have completed any of the previous Sweeps. The cross-sectional weight for sweep 9 was computed for all cases responding at sweep, and it can be used to weight the non-response model at sweep 10, in exactly the same way as the sweep 1 cross-sectional weight was used to compute sweep 2 longitudinal weights. Such a weight would be available for longitudinal analysis of data from sweep 9 onwards.

For analysis of non-consecutive future sweeps (e.g. sweep 11 and 9), cross-sectional weights from the most recent sweep can be considered. Similar considerations as presented in chapter 4 apply.

For longitudinal analysis of paired waves with earlier years, a ‘main sample’ sweep 9 cross-sectional weight will be required. This would involve creating a cross-sectional weight for all main (non-boost) sample cases that responded at sweep 9. This weight would be used for all paired longitudinal comparisons between sweep 9 and earlier sweeps, for example, analysis using sweep 5 and sweep 9. The current sweep 9 cross-sectional weight would not be appropriate for paired waves of analysis since the weight is designed to incorporate the boost sample cases, and these cases do not appear in earlier sweeps.

4.1.3 Incorporating Primary 6 teachers’ data

92% of children who responded to sweep 8 had a data collected from their Primary 6 teacher. Therefore, additional cross-sectional weights were computed to adjust for non-response to the Teachers’ survey. Only a subsample of children with the teacher’s weight will respond at future sweeps, hence a model-based weighting technique can be used to develop the longitudinal weight for analysis of data from future sweeps and questions asked of a teacher at sweep 8. Cases issued to a future sweep of interest with a valid teachers’ weight should be included in a model (weighted by the teachers’ weight) that predicts the response at a future sweep. The inverse of the probability of response combined with the teachers’ weight can be then used as a weight for longitudinal analysis.