Growing Up in Scotland Sweep 11: 2021-23

User Guide

Paul Bradshaw and Anna Keyes, Scottish Centre for Social Research

Scottish Centre for 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)

Contents

1 Survey details

1.1 Study aims and objectives

The overarching aim of the Growing Up in Scotland study is set out in its purpose, which is:

"To generate, through robust methods, specifically Scottish data about outcomes throughout childhood and into adulthood for children growing up in Scotland across a range of key domains:

- *Cognitive, social, emotional and behavioural development*
- *Physical and mental health and wellbeing*
- *Childcare, education and employment*
- *Home, family, community and social networks*
- *Involvement in offending and risky behaviour*

Such data will encompass, in particular, topics where Scottish evidence is lacking and policy areas where Scotland differs from the rest of the UK."

1.2 Sweep 11 data collection elements

Sweep 11 **data collection** included seven main elements:

1. A brief face-to-face (f2f) 'household' Computer Assisted Personal Interview (CAPI) questionnaire which could be answered by *either* the cohort member (young person) or their main carer, where they were living together. The household ID of the person who provided this information is captured in variable MkHGrCom.
2. A face-to-face CAPI interview with the cohort member (young person).
3. A self-complete questionnaire with the cohort member. In the first instance, this was issued as a web survey (Computer Assisted Web Interview – CAWI) in advance of interviewer contact for the face-to-face interview. Where the CAWI was not completed before the interviewer visit, the web questionnaire was delivered as a self-complete instrument (Computer Assisted Self Interview – CASI) on the interviewer's laptop during the visit.
4. A self-complete questionnaire with the cohort member's main carer. This was also initially issued as a CAWI in advance of interviewer contact[1]. Where the parent/carer CAWI was not completed before the interviewer visit, an attempt was made to have it completed as a CASI during the visit.
5. Height and weight measurement of the cohort member.
6. Cognitive assessments of the cohort member.

[1] Eligibility questions were included in the main carer CAWI. Parents/carers were only asked to complete a questionnaire in cases where either the cohort member still lived with them at least some of the time or they otherwise were still in contact with them.

7. Objective physical activity and GPS measurement conducted on behalf of the SPACES2. Further details are provided in section 2.5.

When fieldwork was launched in November 2021, Scottish Government COVID restrictions preventing data collection in respondent's homes remained in place. Whilst interviewers were permitted to make contact on the doorstep, they could not enter people's homes to conduct interviews. As such, the planned f2f interview with the cohort member was delivered as a telephone interview (Computer Assisted Telephone Interview – CATI) until restrictions were lifted in May 2022 and in-home data collection could take place. The main carer self-complete could also be conducted via CATI during this initial phase of fieldwork. The content of the cohort member CAWI was deemed too sensitive to be conducted via CATI and thus remained solely as a CAWI option during this time.

Even after in-home data collection was permitted, participants could choose to complete their interview by phone if they preferred.

Where the interview was not conducted in the home, no height and weight measurements or cognitive assessments were conducted.

Devices for the objective physical activity and GPS data collection were distributed from the office and not by interviewers. As such, this element was not affected by the different approaches taken over the course of fieldwork.

A summary of the face-to-face and remote configurations of each component is provided in Table 1.1.

Table 1.1 Data collection elements at sweep 11

Respondent	Element	F2F mode	Remote mode
Cohort member (young person)	*Household questionnaire**	*CAPI in-home*	*CATI*
	Interview	CAPI in-home	CATI
	Self-completion	CASI in-home	CAWI
	Height and weight	In-home	Not included
	Cognitive assessments	In-home	Not included
Main carer	*Household questionnaire**	*CAPI in-home*	*CATI*
	Interview/self-completion	CAPI in-home	CAWI

*The household questionnaire could be completed by either the cohort member or their main carer where they lived together

1.3 Study design

GUS was initially based on two cohorts of children: the first aged approximately 10 months at the time of first interview (involving around 5217 children at the first sweep) and the second aged approximately 34 months (involving around 2800 children at the first sweep). In 2018, an additional 502 families were recruited to the study. These families took part in interviews alongside families in the original birth cohort. Further details are provided in section 1.4.2.

A second birth cohort of 6127 children aged around 10 months at the first interview was recruited in 2011.

The configuration of cohorts and sweeps for all sweeps of data collection launched to date is summarised in Table 1.2. BC1 refers to the younger of the two original cohorts ('birth cohort 1'), CC to the slightly older cohort ('child cohort') and BC2 to the later birth cohort ('birth cohort 2').

A key aim of using multiple cohorts was to allow the study to provide three types of data:

- Cross-sectional time specific data – e.g. what proportion of 17-18 year-olds were living in single parent families in 2021-23?
- Cross-sectional time series data – e.g. is there any change in the proportion of 5-year-old children living in single parent families between 2010 and 2016?
- Longitudinal cohort data – e.g. what proportion of children who were living in single parent households aged 10-11 are living in different family circumstances at the time they are aged 17-18?

Table 1.2 Ages, stages and timing of data collection sweeps by cohort

Child's age* (school stage)	Cohort, year of data collection (sweep) *achieved sample size*			
	Child cohort	Birth cohort 1		Birth Cohort 2
		Main sample	Boost sample	
10 months	-	2005/06 (1) *5217*	-	2011 (1) *6127*
1-2 years	-	2006/07 (2) *4512*	-	-
2-3 years	2005/06 (1) *2858*	2007/08 (3) *4193*	-	2013 (2) *5020*
3-4 years	2006/07 (2) *2500*	2008/09 (4) *3994*	-	2014** (2.5) *3237*
4-5 years	2007/08 (3) *2332*	2009/10 (5) *3833*	-	2015 (3) *4434*
5-6 years	2008/09 (4) *2200*	2010/11 (6) *3657*	-	-
7-8 years	-	2012/13 (7) *3453*	-	-
8-9 years (Primary 5)	-	2013/14** (7.5) *2775*	-	-
9-10 years (Primary 6)	-	2014/15 (8) *3151*	-	-
10-11 years (Primary 7)	-	2015/16** (8.5) *2099*	-	-
12-13 years (Secondary 1)	-	2017/18 (9)		-
		2917	*502*	
14-15 years (Secondary 3)	-	2019/20 (10)		-
		2669	274	
17-18 years (Secondary 6)	-	2021-23 (11)		-
		2384	265	

* Up to age 8, data collection was timed for each case so that data was captured from children when they were of the same chronological age. Typically the spread was around 3 months. From age 8 onwards, the timing aimed to capture children at the same school stage introducing a wider age range at the point of data collection.
**This data was collected via a web questionnaire or telephone interview.

1.4 Sample design

1.4.1 BC1 Main sample (the original birth cohort)

The original or 'main' BC1 sample was recruited at sweep 1.

The initial area-level sampling frame was created by aggregating Data Zones. Data Zones are small geographical output areas created for the Scottish Government. Data Zones are used 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[2] and then by the Scottish Index of Multiple Deprivation Score (SIMD). 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 were born between 1st June 2004 and 31st May 2005. 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.

1.4.2 BC1 Boost sample (refreshment sample recruited in 2018)

Whilst the overall levels of attrition seen in GUS are typical for a cohort study of its kind, the effects of attrition are spread unevenly over the sample, with some sub-groups affected more than others. Analysis of the achieved sample from Birth Cohort 1 after sweep 8 revealed that two groups in particular had become

[2] 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.

under-represented since the beginning of the study: children born to mothers aged 16-24 at time of birth and children living in the 15% most deprived areas (according to the Scottish Index of Multiple Deprivation).

To resolve this under-representation, a boost sample for BC1 was recruited to the study as part of phase 2 fieldwork for sweep 9, specifically targeting families those in the under-represented groups. Like the sample for the original birth cohort, the sample for the boost was drawn from Child Benefit records held by HMRC. For further details about the sampling approach and rationale, please see the GUS BC1 sweep 9 User Guide (https://growingupinscotland.org.uk/wp-content/uploads/2019/08/BC1-SW9-User-Guide.pdf)

At sweep 9, a total of 502 families were recruited to the boost sample. All families who consented to follow-up and who had not subsequently withdrawn from the study were issued for sweep 11 fieldwork (n=492).

1.5 Developing and piloting

Policy priorities and key topics of interest for the sweep 11 adult and young person questionnaires were initially discussed and agreed by the study's Scottish Government Project Manager and a number of internal and external stakeholder groups, including the Questionnaire Advisory Group.

Informal qualitative consultation was also undertaken with groups of young people who were the same age cohort members would be at the time of data collection. This took the form of small, online discussion groups facilitated by the research team. The aim was to find out what the topical issues were for young people. A total of three groups were convened: two with sixth year pupils attending two different schools and one with a group of young people who had left school.

Feedback from stakeholders and young people was then used by the GUS team at ScotCen to develop the questionnaires with input from the study's Questionnaire Advisory Group and policy teams across the Scottish Government.

Cognitive testing of selected new items in the young person questionnaire was carried out in May 2021. Cognitive interviews were undertaken with 17 young people of a similar age to cohort members and with a range of characteristics representing key expected differences in the sample at this sweep. In particular, the cognitive pilot sample included some young people who were still at school and some who had left school and were in further education, employment or training.

A full dress rehearsal pilot, including near to final versions of all cohort member and main carer CAWI, CASI and CAPI questionnaires, was conducted in July/August 2021. The existing GUS pilot sample was used and a total of 84 cases were issued for the pilot. Overall, 54 (64% response rate) young person CAPI telephone interviews were conducted, 35 (42%) young person CAWI questionnaires were completed and 40 (48%) main carer CAWI questionnaires were completed.

1.6 Sweep 11 fieldwork timing

In sweeps 1-7, fieldwork was conducted over a 14-month period with cases issued to field according to the child's age and interviews taking place around a specified date calculated according to the child's birthday (the 'target interview date'). Ahead of sweep 8 there was interest in interviewing families according to the child's school year. Therefore, from sweep 8, fieldwork moved from an 'ages' to a 'stages' approach. This means that the age gap between the young people at the time of interview is larger at sweeps 8-11 than at previous sweeps.

Because of how children were initially sampled, cohort members in BC1 span two different school years. Following the approach taken in sweeps 8-10, the fieldwork for sweep 11 was intended to be conducted over two phases coinciding with when the cohort would be in their first or second term of their sixth (final) year of secondary school or an equivalent stage if they had left school:

- **Phase 1** fieldwork was scheduled to take place between September 2021 and February 2022.
- **Phase 2** fieldwork was scheduled to take place between September 2022 and February 2023.

With Scottish Government COVID restrictions on conducting data collection in-home still in place in September 2021, a decision was taken to delay fieldwork launch until November in the hope that restrictions may be eased. This was not the case and fieldwork launched in November utilising a 'knock-to-nudge' and telephone interview approach with height and weight measurements and cognitive assessments initially dropped from the fieldwork.

The significant limits the COVID pandemic placed on the conduct of face-to-face survey data collection had a major impact on the strength of survey interviewer panels in fieldwork agencies across the UK and beyond. Many interviewers left, meaning panels were vastly lower in number than pre-pandemic. As a result, fully resourcing the survey to deliver within the planned timescales proved to be incredibly challenging and ultimately unachievable. Fieldwork at both phases was extended to ensure all cases were adequately covered and data was collected from as many cases as possible.

This combination of issues meant that the actual timing of fieldwork was as follows:

- **Phase 1** fieldwork took place between November 2021 and November 2022.
- **Phase 2** fieldwork took place between September 2022 and June 2023.

Please see the **project instructions** for further details.

1.7 Incentives and thank you gifts

To boost response and, in particular, encourage participation amongst under-represented sub-groups of the sample and those at higher risk of not taking part, targeted incentives were used at sweep 11. The criteria for incentive eligibility was similar to that used at previous sweeps:

- Case is in the boost sample

- Case is in the main sample and did not participate at sweep 10 (selected unproductive outcomes)
- Case was unproductive at sweep 11 first issue and was selected for re-issue fieldwork.

Incentives were conditional on participation. Cohort members were given a £20 Love2Shop voucher and main carers were given a £5 voucher.

All cohort members who participated at sweep 11 were sent a thank you letter along with a branded gift (a multi-cable adaptor).

1.8 Response

Table 1.3 shows historical response/attrition for BC1. At sweep 11, a total of 2381 interviews were achieved with the original sample through either face-to-face, web and/or telephone data collection with either a cohort member or their parent. This represents 46% of cases achieved at sweep 1.

Table 1.3 BC1 historical response

	Cases achieved	% of sweep 1 cases
Sweep 1	5217	-
Sweep 2	4512	86%
Sweep 3	4193	80%
Sweep 4	3994	77%
Sweep 5	3833	73%
Sweep 6	3657	70%
Sweep 7	3456	66%
Sweep 8	3150	60%
Sweep 9 (Main sample only)	2917	56%
Sweep 10 (Main sample only)	2669	51%
Sweep 11 (Main sample only)	2384	46%

Details of the number of cases issued and achieved at sweep 11 are presented in Table 1.4. Following a top-level overall response rate for the entire sweep, separate rates are provided for the different data collection elements.

Note that all Boost sample cases were issued as part of phase 2 fieldwork. Note also that neither cognitive assessments nor height and weight measurements were collected as part of the initial knock-to-nudge/telephone data collection stage.

Around one-third of fieldwork was completed over the period during which COVID restrictions on data collection were in place. Without an in-home follow-up, the sequential design intended to maximise response to the cohort member self-complete – whereby those who had not completed the CAWI did so via CASI during the interviewer's visit – could not be realised resulting in a lower than anticipated response to this element of the data collection.

Table 1.4 Sweep 11 response by sample type						
	Total cross-sectional sample (Main and Boost samples)		**Main sample**		**Boost sample**	
	No. of cases	% of issued in-scope	No. of cases	% of issued in-scope	No. of cases	% of issued in-scope
Total in-scope/issued	3637	-	3145	-	492	-
Total achieved (any element completed)	2649	73%	2384	76%	265	54%
Young person elements						
Young person achieved (any element completed)	2484	68%	2235	71%	249	51%
Total face-to-face/telephone interviews	2368	65%	2132	68%	236	48%
Telephone	*1198*	*33%*	*1167*	*37%*	*31*	*6%*
Face-to-face	*1170*	*32%*	*965*	*31%*	*205*	*42%*
Total self-complete (either CAWI or CASI)*	1971	54%	1773	56%	198	40%
Total CAWI interviews	*1536*	*42%*	*1432*	*45%*	*104*	*21%*
Total CASI interviews	*563*	*15%*	*456*	*14%*	*107*	*22%*
Total height and weight measurements	995	27%	819	26%	176	36%
Total cognitive assessments	1076	30%	898	29%	178	36%
Total consenting to SPACES2 follow-up	959	26%	887	28%	72	15%
Main carer elements						
Main carer achieved (by any mode)*	1979	54%	1824	58%	155	32%
Total CAWI interviews	*1739*	*48%*	*1628*	*52%*	*111*	*23%*
Total CASI/CAPI interviews	*340*	*9%*	*291*	*9%*	*49*	*10%*

*Due to fieldwork complications, some cohort members and main carers completed both a CAWI and a CASI. Only one such completion is counted here so this figure is not equivalent to the sum of the subsequent two rows.

The following variables in the dataset can be used to determine which cases responded to the different elements and thus provided different types of data:

- Dkypwebo: Cohort member web outcome
- Dkmcwebo: Main carer web outcome
- Dkypf2fo: Cohort member face-to-face/telephone interview outcome
- Dkypsco: Whether young person provided self-complete data (either via web or CASI during face-to-face visit)
- Dkmcinto: Whether the main carer provided data in any mode
- Dkanyprd: Whether any data was provided by either the cohort member or the main carer (i.e. the household/case was productive)

Table 1.5 provides further details on the reasons for non-response amongst those cases which did not result in a productive CAPI interview with the cohort member in either the knock-to-nudge or in-home phases.

As the table shows, the vast majority of unproductive cases were as a result of a refusal by the cohort member or their main carer to take part.

Table 1.5 Sweep 11 CAPI non-response by sample type

	Total cross-sectional sample (Main and Boost samples)		**Main sample**		**Boost sample**	
	No. of cases	% of issued in-scope	No. of cases	% of issued in-scope	No. of cases	% of issued in-scope
Total in-scope/issued	3637	-	3145	-	492	-
Total face-to-face/telephone interviews	2382	65%	2122	67%	235	48%
Total with no face-to-face/telephone interview	1272	35%	1018	32%	254	52%
Non-contact	*153*	*4%*	*122*	*4%*	*31*	*6%*
Movers	*136*	*4%*	*94*	*3%*	*42*	*9%*
Refusal	*885*	*24%*	*722*	*23%*	*163*	*33%*
Other	*98*	*3%*	*80*	*3%*	*18*	*4%*

1.9 Length of interview

The median interview length, with inclusion of all potential elements - cohort member and main carer interviews, cognitive assessments and height and weight measurements – though only where relevant, lasted 51 minutes. This varied over the course of fieldwork. Interviews conducted remotely, before in-home data collection restrictions were lifted, lasted around 45 minutes whereas those conducted in-home, which could incorporate CASI self-completions, height and weight measurements and cognitive assessments, lasted around 60 minutes.

2 Sweep 11 data collection elements

2.1 Interview with the cohort member (young person)

The cohort members were interviewed directly for the fifth time at sweep 11. This sweep was the first where cohort members were invited to interview independently of their parent, with separate advance letters mailed directly to them. For details about consent procedures and contact procedures please see the interviewer instructions. Copies of the invitation letters and other documents are provided in the data documentation.

Cohort members were invited to take part in two main interview elements:

1) a self-complete CAWI questionnaire issued in advance of the interviewer visit which could be completed via CASI during the visit if not done so beforehand;
2) an interviewer-led interview conducted either via telephone (where in-home data collection was not permitted) or face-to-face/in-home

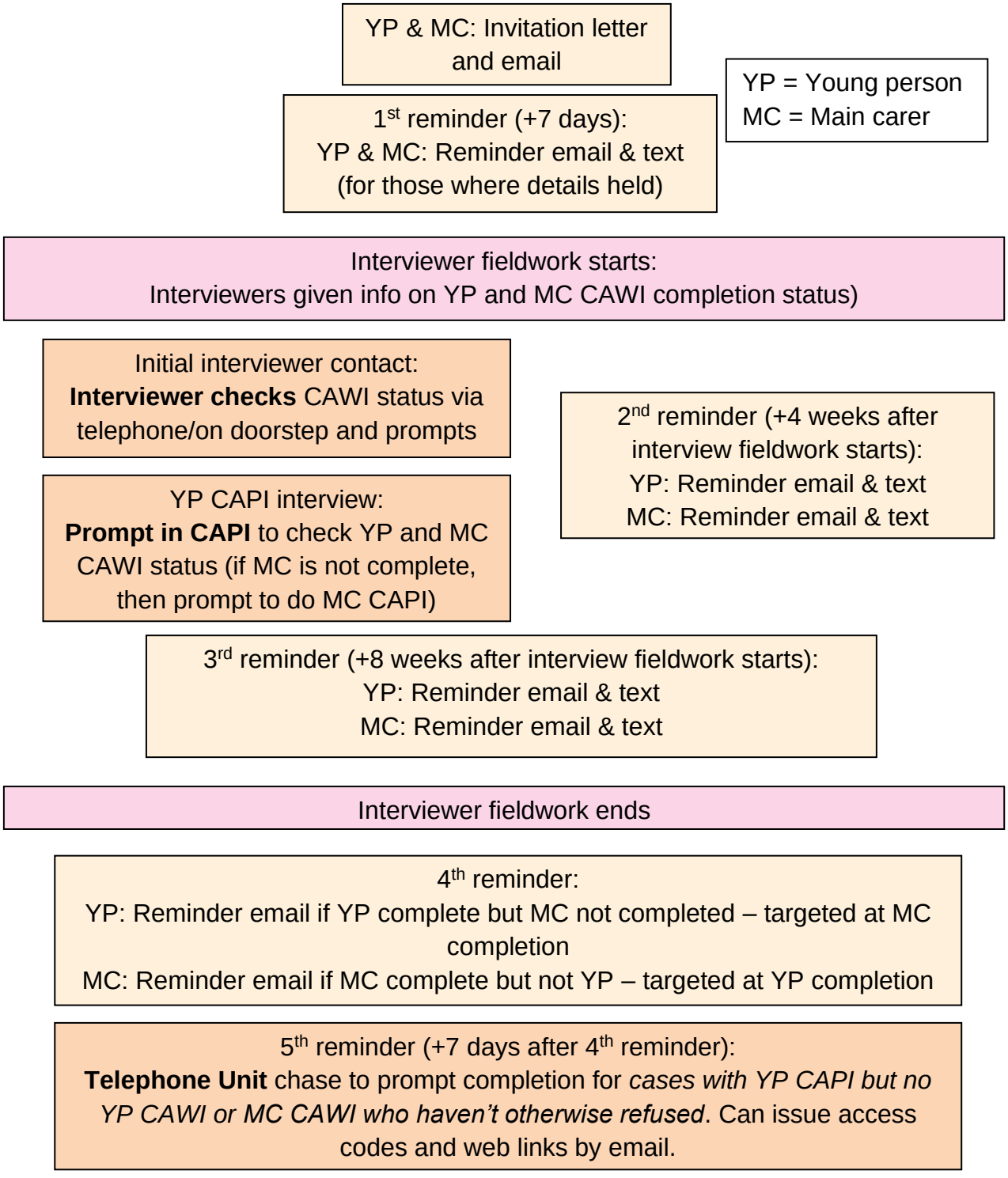
Invitations to the CAWI questionnaire were issued via the advance letter and, if participants had previously provided their email address, by email.

An initial reminder, to prompt the young person to complete their CAWI questionnaire ahead of the interviewer visit, was sent by email and text (where young people had previously provided these details) around one week after the invitation mailings.

Where interviews were conducted in-home, if the cohort member had not completed their CAWI before the interviewer visit this was completed as a CASI during the interview. However, this was not possible for those interviews conducted by phone over the initial period of fieldwork. As such, a comprehensive programme of reminders was agreed to encourage response to the CAWI questionnaire amongst those cohort members. This is summarised in Figure 1.1.

The CAWI and CAPI elements were programmed separately and not linked. Ahead of and on an ongoing basis during fieldwork, interviewers were provided with manual updates via email on whether or not cases they were covering had completed their CAWI. There were also prompts in CAPI for interviewers to check with participants whether or not they had completed their CAWI. In practice, this approach was not foolproof and in some instances participants completed a CAWI and a CASI. Where this was the case, one set of data was selected for inclusion in the final dataset. This was the CASI by default, by virtue of it being collected at the same time as the CAPI and thus better aligning the temporal nature of the two sources of data, unless the CAWI data was more complete (i.e. more questions had been answered) in which case it was used.

Figure 1.1 Schedule of CAWI reminders for cohort members and main carers



Where the interviewer-led element was conducted by phone, to mimic the face-to-face interview as closely as possible, participants were asked to refer to a copy of showcards which were made available online. If the participant did not have internet access, they were provided with a paper copy of the showcards.

2.2 Interview with the cohort member's main carer

At sweep 1, primarily because of the inclusion of questions on the mother's pregnancy and birth of the sample child, interviewers were instructed as far as possible to undertake the interview with the child's mother. Where the child's mother was not available, interviews were undertaken with the child's main carer. At the following sweeps, interviewers have been instructed to undertake the interview with the same respondent as in the previous sweep, where possible and appropriate. At sweep 11, this means the same respondent as sweep 10 (or sweep 9/sweep 8 etc if the household skipped one or more sweeps). Where this was not possible or appropriate, interviews were conducted with another parent or carer. In practice, most parent/carer interviews were undertaken with the adult who took part in the previous sweep (95% of adult interviews were with the adult respondent who took part in the previous sweep) and this was usually the young person's mother (94% of adult interviews were with the cohort member's mother). Further details about contact procedures are available in the **project instructions**.

Main carers were initially invited to take part by completing a CAWI questionnaire. Invitations to the CAWI were issued via a letter addressed to the parent/carer respondent from the previous sweep. If an email address was held, invites were also issued by email. Two versions of the invitation letter/email were used: version A - which contained a brief finding from sweep 10 on what young people wanted to do after leaving school - and version B - with a brief finding on the perceived financial impact of COVID-19. Version B was used with families whose household income was in the bottom two quintiles, as reported at either sweep 9 or sweep 10. Version A was used for all other families.

Whilst the letter/email encouraged response from the parent/carer who had participated at the previous sweep, the questionnaire could be completed with any adult who lived with and had caring responsibilities for the cohort member. A reminder to complete their CAWI questionnaire ahead of the interviewer visit was sent by email and text (where main carers had previously provided these details) around one week after the invitation mailings.

If the main carer had not completed their CAWI before the interviewer visit this was completed either as a telephone CAPI – if interviews were not being conducted in home – or as a CASI (self-complete on the interviewer's laptop).

If the parent/carer's questionnaire was still outstanding following the interviewer visit, several further reminders were issued by email and text, as shown in Figure 1.1.

As with the young person questionnaires, the CAWI and CASI/CAPI elements were programmed separately and not linked. Thus interviewers were reliant on regular manual updates on the parent's CAWI status (outstanding or complete) via email and prompts in the CAPI to check whether or not the CAWI had been completed. In some cases, parents/carers completed both a CAWI and a CASI. Where this was the case, the CASI data has been used as is the case for cohort member data. As noted in Table 1.4, the majority (79%) of parents/carers completed their questionnaire online.

2.3 Cognitive assessments

Cognitive assessments were carried out with cohort members as part of the main face-to-face interviews. **No cognitive assessments were carried out as part of the knock-to-nudge/telephone data collection.**

Cognitive assessments were previously carried out with the cohort members in BC1 at sweeps 3, 5, 8, 9 and 10. At sweep 11, mirroring the approach at sweeps 8-10, cohort members were assessed using the 'Listening Comprehension' subtest of the Weschler Individual Achievement Tests, 2nd Edition (WIAT-II).

WIAT-II is an educational assessment tool which is widely used by educational psychologists to examine cognitive development and educational ability. The assessments carried out with GUS cohort members were adapted for use in a survey setting and modified to be administered in CAPI.

The Listening Comprehension subtest is designed to measure the ability to listen for detail by selecting the picture that matches a word or sentence (e.g. 'point to the dog') and generating a word that matches a picture and an oral description (e.g. 'what is this?'). There are strict protocols which must be adhered to when administering assessments. These ensure that the resultant data can be confidently compared with the normative data used to produce the various derived scores necessary for analysis.

The Listening Comprehension test includes three sub-assessments: Receptive Vocabulary, Sentence Comprehension and Expressive Vocabulary (see table 2.1 below).

Table 2.1 Child cognitive assessments: WIATT-II Listening Comprehension

Assessment name	Assesses	Method	Max no. of items
Receptive vocabulary	Ability to listen for details and knowledge of words	Young person is asked to select a picture that matches a word	16
Sentence comprehension	Ability to listen for details and knowledge of words	Young person is asked to select a picture that matches a sentence	10
Expressive vocabulary	Knowledge of words	Young person is asked to generate a word that matches a picture and oral description	15

For each assessment, the starting point is determined by the child's age. The assessment continues until the last item or until six consecutive incorrect responses are given.3 At GUS sweep 11, all children started at the same point (note that this was not the first item in each sub-test) however, some children may have subsequently been asked earlier items depending on their progress through the assessment. Where children were not asked those earlier items, they were scored positively. Understanding which set of items were administered to the cohort member is important when analysing the results.

The following scores are available in the dataset:

[3] Further details are available in the project instructions.

- ***Receptive Vocabulary Adjusted Raw Score:*** A count of all the items on Receptive Vocabulary the child answered correctly (including where early items were automatically scored).
- ***Sentence Comprehension Adjusted Raw Score:*** A count of all the items on Sentence Comprehension the child answered correctly (including where early items were automatically scored).
- ***Expressive Vocabulary Adjusted Raw Score:*** A count of all the items on Expressive Vocabulary the child answered correctly (including where early items were automatically scored).
- ***Listening Comprehension Raw score:*** The raw score is a count of the number of items the child answered correctly. The total raw score for the Listening Comprehension subtest is derived by adding up the adjusted raw scores for each of the three sub-assessments (Receptive vocabulary; Sentence comprehension and Expressive vocabulary).
- ***Listening Comprehension Standard Score***: A normalised transformation of the raw score which uses an external standard or 'norming' sample and takes into account the child's age in months at the time the assessment was undertaken. The standard score can be used as a measure of how far a child's score from the mean (and median) score for a child their age, measured in standard deviations. The Listening Comprehension standard score can also be compared to other types of normalised derived scores, like subtest scaled scores from the Wechsler intelligence scales.

For each raw score outlined above it is possible to derive *within-sample standardised z scores* which allow for comparisons to be made across sub-assessments (measures in standard deviations from the mean).

Note that the exercises are designed to provide a picture of the range of skills across a number of young people, not to give a clinical assessment of an individual young person. Formally, WIAT-II is designed to be used with children of a maximum age of 16 years 11 months. This is younger than the age of almost all GUS cohort members at sweep 11 (80% were aged 17, 18% were aged 18 and 2% were aged 16). Notably, the most recent edition of WIAT (WIAT-III) is suitable for use with adults up to age 26[4].

Further information about the WIAT-II measures is available online, at: http://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildAchievementMeasures/WechslerIndividualAchievementTest-SecondUKEdition(WIAT-IIUK).

2.4 Height and weight measurements

The cohort member's height and weight measurements were previously taken in sweeps 2, 4, 6, 7, 8, 9 and 10 and were also included as part of the in-home

[4] See https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Academic-Learning/Reading/Wechsler-Individual-Achievement-Test---Third-UK-Edition/p/P100009274.html?tab=product-details

data collection at sweep 11. **No height and weight measurements were undertaken as part of the knock-to-nudge/telephone data collection.** In these instances, cohort members were asked to provide self-reported height and weight.

Where a face-to-face, in-home visit took place, interviewers were asked to measure the height and weight of all cohort members. However, in some cases it may not have been possible or appropriate to do so, for example if it was clear that the young person was unwilling or that the measurement would be far from reliable.

It was recommended that height and weight measurements be taken on a floor which was level and not carpeted. If all the household was carpeted, a floor with the thinnest and hardest carpet was chosen (usually the kitchen or bathroom). The interviewer was asked to code whether they experienced problems with the height and/or weight measurements and, if they did, to indicate whether they felt the end result was reliable or unreliable at variable WkXhei14 and WkXwei19. As a rough guide, if the measurement was likely to be more than 2 cm (3/4 inch) from the true figure for height or 1 kg (2 lbs) from the true figure for weight, it was coded as unreliable.

If the cohort member was not willing to have their height or weight measured, for example saying that they were too busy or already knew their measurements, a refusal code was entered for the measurements at variable WkXhei01, with the reason for refusal at WkXhei021-8 or WkXwei021-7.

If the height or weight was refused or not attempted, including because the interview was being conducted by telephone, the cohort member was asked for an estimated height or weight, in metric or imperial measurements.

Detailed protocols of how height and weight measurements are taken are included as appendices to the main interviewer instructions deposited with the dataset and available from the data archive website.

The data has been used to estimate an approximate BMI (Body Mass Index) score for each cohort member. Separate height, weight and BMI variables are available for interviewer-measured and self-reported measurements. Further details on the data and variables associated with the height and weight measurements can be found in section 0.

2.5 Physical activity and GPS data

Objective physical activity and GPS data was collected from cohort members as part of the Studying Physical Activity in Children's Environments across Scotland (SPACES2) project run by researchers at the MRC/CSO Social and Public Health Sciences Unit at University of Glasgow.

All cohort members eligible for participation in GUS sweep 11 were provided with information about the SPACES2 data collection via a dedicated GUS branded leaflet issued alongside their main GUS advance mailing/interview invite. The leaflet informed cohort members that to participate in this part of the project, they would wear a waist-worn accelerometer and GPS device during waking hours for 8 days.

Consent for participation was gathered during the CAPI (telephone or in-home) interview. Where consent was given, a participation pack was sent from the

NatCen office containing the devices and related equipment (e.g. belt, charging cable), participant instructions, a postage-paid return envelope and a logbook to record times when the devices were put on and taken off each day.

A series of text, email and telephone reminders were put in place to support cohort members taking part:

- Reminder 1 – email/text to cohort member: 1 day after devices despatched. Provided information on starting the data collection.
- Reminder 2 – email/text to cohort member: 11 days after devices despatched. Reminder to return devices.
- Reminder 3 – email/text to cohort member where device had not been returned: 18 days after devices despatched. Reminder to return devices.
- Reminder 4 – telephone calls to cohort member and/or main carer where devices had not been returned: +23 days after device despatched. Reminder to return devices.
- Reminder 5 – email to cohort member and/or main carer where devices remained outstanding: Reminder to return devices.

Where devices remained outstanding after all reminders, cases were batched and issued to survey interviewers for an in-person visit to retrieve them.

As noted in Table 1.4 above, of those cohort members with a CAPI interview, 959 agreed to participate in the SPACES2 data collection and had devices issued. A total of 708 participants (74% of those with devices issued) returned at least one of their devices.

3 Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted. The enclosed **CAPI edit instructions** provide details of the tasks that were conducted.

Coding and editing tasks were carried out on data collected across all modes.

4 Weighting the data

4.1 Background

Sweep 11 of GUS is the first in which the cohort young person was the principal respondent. Cohort members and their main carers participated in sweep 11 independently, thus not all responding young people also had main carer data and not all responding main carers also had young person data. Consequently, it was decided to split the weights, so that young person and main carer data could be analysed separately. The methods used for the young person and main carer weights were kept completely consistent.

4.1.1 Weights developed for sweep 11

Four sets of weights were generated for the sweep 11 data:

- A longitudinal weight for responding cohort young people whose main carer had responded at every other sweep of GUS up to and including sweep 10.
- A cross-sectional weight for all responding cohort young people.
- A longitudinal weight for responding main carers who had also responded at every other sweep of GUS up to and including sweep 10.
- A cross-sectional weight for all responding main carers.

Data collection from cohort young people consisted of both an interview and a self-complete questionnaire (CASI). Not all young people who were interviewed also did the CASI, however 78% of young people with interview data also had CASI data. Weighted and unweighted profiles of all responding young people were compared with CASI responding young people, to determine whether a separate CASI weight was needed. Both unweighted and weighted profiles were very similar, with an absolute mean difference of 0.3 percentage points across 23 key demographic variables used in weighting. It was therefore decided that separate young person CASI weights were not required, and the young person cross-sectional weights can be used for analysis of young person CASI responses.

4.2 Weights for young person interview data

4.2.1 Young person sample

The sweep 11 sample of young person respondents can be split into three groups. For the purposes of describing the weighting, these have been named Sample A, Sample B, and Boost Sample. They are defined as follows:

- Sample A are young people whose main carers responded to all previous sweeps and who took part in sweep 11 themselves.

- Sample B are young people who took part in sweep 11 themselves and whose main carers responded to sweep 1 but did not participate in one or more of sweeps 2 to 10.
- Boost sample are young people who were part of the refreshment sample added at sweep 9 and took part in sweep 11 themselves.

The three samples were treated separately during the weighting, for consistency with the weighting of previous sweeps.

- There were 2227 young people in Sample A, 1830 of whom responded at sweep 11. The response rate was 82%.
- There were 926 young people in Sample B, 405 of whom responded at sweep 11. The response rate was 44%.
- There were 492 young people in the boost sample, 249 of whom responded at sweep 11. The response rate was 51%.

The issued and responding sample sizes for the three groups in the young person data are given in Table 4.1 below.

Table 4.1 Young person sample groups for weighting and corresponding response figures

Sample	No. issued	No. responding	Response rate
Sample A	2227	1830	82%
Sample B	926	405	44%
Boost	492	249	51%
Combined (A+B+Boost)	3645	2484	68%

Two sets of weights were created for the responding young people: a longitudinal weight and a cross-sectional weight. Only members of Sample A, whose main carers responded to all previous GUS sweeps, received a longitudinal weight. This weight is described in greater detail in section 4.2.2 below.

All young people that responded in sweep 11 were given a cross-sectional weight, including Sample A, Sample B, and Boost Sample cases. This weight is described in greater detail in section 4.2.3 below.

4.2.2 Longitudinal weights for young person interview data

Young person longitudinal weights were only generated for responding young people in Sample A. The weights were generated using a model-based method with data from previous sweeps. This same method was used for the main carer longitudinal weights. Ineligible households (deadwood) were not included in the modelling. The base for the model was the sweep 10 longitudinal respondents (n = 2238), including 11 cases that were not issued in sweep 11 as they withdrew from the study.

A model of whether the young person responded to sweep 11 was fitted using forward and backward stepwise logistic regression. The potential predictors tested were a set of socio-demographic variables collected from previous sweeps of the study, including both characteristics of main carers and of the

household. The non-response model was run weighted by the sweep 10 longitudinal 100% sample weights.

The following variables were significant in the stepwise regressions and therefore included in the final model: household tenure, number of children in household, highest qualification of main carer, whether main carer currently has a job, mother's employment status, main carer closeness to child, frequency that main carer helped child with their homework, main carer health, and whether main carer has a disability or limiting illness.

The non-response weight was calculated as the inverse of the modelled probabilities of response. This was trimmed at the 99th percentile to improve efficiency and multiplied by the sweep 10 longitudinal 100% sample weight. The combined weight was checked for outliers and the top 3 weights trimmed. The final weights were scaled to the sweep 11 young person Sample A size of 1830. They have an efficiency of 68% and an effective sample size of 1240.

4.2.3 Cross-sectional weights for young person interview data

Cross-sectional weights were generated for all young people who responded to sweep 11, whether they completed the interview, the CASI, or both elements. The cross-sectional responding sample included Sample A, Sample B, and Boost Sample cases.

Calibration weighting was applied to the young person combined sample to create the cross-sectional weights. This method adjusts a set of start weights using an iterative procedure so that they match pre-defined population totals. The resulting weights, when applied to the young person data, produce survey estimates that match the population estimates.

Population estimates for calibrating the combined young person responding sample were derived from weighting young person Sample A by the young person longitudinal weight. Variables included in the calibration were selecting by fitting a model using forward and backward stepwise regression. The base of the model was the young person cross-sectional responding sample and the outcome variable whether the young person's main carer was interviewed in all previous waves. The models were run weighted by a start weight, consisting of the young person longitudinal weight for Sample A cases and the cross-sectional weights from the wave in which the main carer last responded for Sample B and Boost cases.

The potential predictors tested were a set of socio-demographic variables collected from previous sweeps of the study, including both characteristics of main carers and of the household. The following variables were significant in the stepwise regressions and therefore included in the calibration: number of parents in family, whether main carer currently has a job, employment in the household, mother's employment status, young person health, main carer health, whether main carer has a limiting illness, household tenure, whether cohort child was mother's firstborn, key subgroup5, quintiles of SIMD education domain scores, and quintiles of SIMD crime domain ranks. Three additional key

[5] A four category variable identifying whether the cohort child's parent was under 25 when they were born and whether their address was in the most deprived 15% of areas in Scotland

variables were included in the calibration despite not being found significant during modelling, because of their importance for analysis and for continuity with previous sweeps: household income, urban-rural status, and deprivation quintiles.

After calibration the weights were checked for outliers and the top five weights trimmed. The final weights were scaled to the sweep 11 young person responding sample size of 2484. They have an efficiency of 74% and an effective sample size of 1826.

4.3 Weights for main carer interview data

4.3.1 Main carer sample

The sweep 11 sample of main carer respondents can be split into three groups. For the purposes of describing the weighting, these have been named Sample A, Sample B, and Boost Sample. They are defined as follows:

- Sample A are main carers who responded to all previous sweeps and who took part in sweep 11 themselves.
- Sample B are main carers who took part in sweeps 1 and 11 but did not participate in one or more of sweeps 2 to 10.
- Boost sample are main carers who were part of the refreshment sample added at sweep 9 and took part in sweep 11.

The three samples were treated separately during the weighting, for consistency with previous sweeps.

- There were 2227 main carers in Sample A, 1528 of whom responded at sweep 11. The response rate was 69%.
- There were 926 main carers in Sample B, 296 of whom responded at sweep 11. The response rate was 32%.
- There were 492 main carers in the boost sample, 155 of whom responded at sweep 11. The response rate was 32%.

The issued and responding sample sizes for the three groups in the young person data are given in Table 4.2 below.

Table 4.2 Main carer sample groups for weighting and corresponding response figures

Sample	No. issued	No. responding	Response rate
Sample A	2227	1528	69%
Sample B	926	296	32%
Boost	492	155	32%
Combined (A+B+Boost)	3645	1979	54%

Two sets of weights were created for the responding main carers: a longitudinal weight and a cross-sectional weight. Only members of Sample A, who responded to all previous GUS sweeps, received a longitudinal weight. This weight is described in greater detail in section 4.3.2 below.

All main carers that responded in sweep 11 were given a cross-sectional weight, including Sample A, Sample B, and Boost Sample cases. This weight is described in greater detail in section 4.3.3 below.

4.3.2 Longitudinal weights for main carer self-complete data

Main carer longitudinal weights were only generated for responding main carers in Sample A. The weights were generated using a model-based method with data from previous sweeps. This same method was used for the young person longitudinal weights. Ineligible households (deadwood) were not included in the modelling. The base for the model was the sweep 10 longitudinal respondents (n = 2238), including 11 cases that were not issued in sweep 11 as they withdrew from the study.

A model of whether the main carer responded to sweep 11 was fitted using forward and backward stepwise logistic regression. The potential predictors tested were a set of socio-demographic variables collected from previous sweeps of the study, including both characteristics of main carers and of the household, the same set that was used for the young person longitudinal weighting. The non-response model was run weighted by the sweep 10 longitudinal 100% sample weights.

The following variables were significant in the stepwise regressions and therefore included in the final model: household tenure, mother's age group at birth of cohort child, highest qualification of main carer, NS-SEC category of main carer, employment in the household, main carer health, closeness of main carer to child, and household income.

The non-response weight was calculated as the inverse of the modelled probabilities of response. This was trimmed at the 99th percentile to improve efficiency and multiplied by the sweep 10 longitudinal 100% sample weight. The combined weight was checked for outliers and the top 6 weights trimmed. The final weights were scaled to the sweep 11 main carer responding sample size of 1528. They have an efficiency of 65% and an effective sample size of 988.

4.3.3 Cross-sectional weights for main carer self-complete data

Cross-sectional weights were generated for all main carers who responded to sweep 11, included Sample A, Sample B, and Boost Sample cases.

Calibration weighting was applied to the main carer combined sample to create the cross-sectional weights. This method adjusts a set of start weights using an iterative procedure so that they match pre-defined population totals. The resulting weights, when applied to the combined data, produce survey estimates that match the population estimates.

Population estimates for calibrating the combined young person responding sample were derived from weighting main carer Sample A by the main carer longitudinal weight. Variables included in the calibration were selecting by fitting a model using forward and backward stepwise regression. The base of the model was the main carer cross-sectional responding sample and the outcome variable whether the main carer was interviewed in all previous waves. The

models were run weighted by a start weight, consisting of the main carer longitudinal weight for Sample A cases and the cross-sectional weights from the wave in which the main carer last responded for Sample B and Boost cases.

The potential predictors tested were a set of socio-demographic variables collected from previous sweeps of the study, including both characteristics of main carers and of the household. The same set of variables was used as for the young person cross-sectional weights. The following variables were significant in the stepwise regressions and therefore included in the calibration: number of parents in family, whether main carer currently has a job, mother's employment status, young person health, main carer health, whether cohort child was mother's firstborn, and key subgroup [6]. Three additional key variables were included in the calibration despite not being found significant during modelling, because of their importance for analysis and for continuity with previous sweeps: household income, urban-rural status, and deprivation quintiles. These three variables were also included in the calibration of the sweep 11 young person cross-sectional weights.

After calibration the weights were checked for outliers and the top four weights trimmed. The final weights were scaled to the sweep 11 young person responding sample size of 1979. They have an efficiency of 70% and an effective sample size of 1381.

4.4 Applying the weights

For each sample, the cross-sectional weights should be used for any cross-sectional analysis, for example any analysis of sweep 11 young person data only. All respondents, including those from the Boost Sample, that responded at sweep 11 have either a young person or main carer cross-sectional weight. Analysis of data collected from young people in sweep 11 should use the young person cross-sectional weight and analysis of data collected from main carers in sweep 11 should use the main carer cross-sectional weight.

The longitudinal weights may be used for analyses of more than one sweep of data. Main carers who responded to sweep 11 every previous sweep of GUS have a main carer longitudinal weight. Young people who responded to sweep 11 and their main carer responded to every previous sweep of GUS have a young person longitudinal weight.

The application of longitudinal and cross-sectional weights for analysis of multiple sweeps of data is discussed in the GUS weighting review published alongside the sweep 10 data[7] (Lubian, Tipping and Bradshaw, 2021).

Table 4.3 Description of weight variables in the data file

Variable name	Label
DkWTchd2	Dk: Longitudinal weight – Young Person
DkWTchld	Dk: Cross-sectional weight – Young Person
DkWTbth2	Dk: Longitudinal weight – Main Carer

[6] A four category variable identifying whether the cohort child's parent was under 25 when they were born and whether their address was in the most deprived 15% of areas in Scotland

[7] Lubian, K., Tipping, S. and Bradshaw, P. (2021) Survey weights and longitudinal analysis – Summary findings from a Growing Up in Scotland Working Paper, Edinburgh: Scottish Centre for Social Research. Access at https://growingupinscotland.org.uk/data-documentation

DkWTbrth	Dk: Cross-sectional weight – Main Carer

5 Using the data

The GUS sweep 11 data collected from cohort members and their main parent or carer consists of the following SPSS file:

GUS_SW11_B.sav	2649 cases	Birth cohort 1

5.1 Variables on the data file

The data file contains questionnaire variables (excluding variables used for administrative purposes) and derived variables. The variables included in the file are detailed in the **variable list**. As far as possible they are grouped in the order they were asked in the interview/questionnaire.

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 interview documentation.

Data from each of the different modes and instruments has been captured in a single dataset. Where a question asked in the CAWI and the CASI were the same, these have been merged into single variable. A number of variables are included to enable identification of the completion mode. There were no notable differences between questions asked across the different modes.

As noted above, in a number of cases, participants provided the same data twice – completing both the CAWI and, later, in-home versions of a questionnaire. Where this was the case, only one set of data was retained for inclusion. This was usually the in-home/CAPI/CASI version - for the reason that it would be more contemporaneous with the main interview data – unless the CAWI data was more complete (i.e. there were significantly fewer variables with missing data).

For variables with answers following a scale, such as ‘Strongly agree’ to ‘Strongly disagree’ for instance, it must be noted that the order of the answer categories may not follow systematically an ascending or descending scale throughout the list of variables. Also, the answers may equally refer to positive or negative statements as in the Strength and Difficulties questions. The phrasing of the question and the list of answers provided on the showcards - if any - shape the variables. The user must therefore take these variations into account when creating derived variables.

5.2 Variable naming convention

Variables names are normally made up of 8 characters, the first indicates the source of the variable, the second the year of collection and the rest is an indication of the question topic. Therefore, where the same question was asked in the different sweeps the names will usually be the same apart from the second character. If a variable name has changed substantially between sweeps this is marked in the variable list.

The naming convention is summarised in Table 5.1

Table 5.1 GUS variable naming conventions – BC1			
Character no.			
1		2	
Source of data		**Sweep**	
Non-sequential capitals: A, D,M, P, C		Sequential lower case: a, b, c, etc...	
Source code	**Details**	**Sweep code**	**Child's age**
AL	Area level variable	a	10 months
D	Derived variable	b	Almost 2 years
DP	Derived variable from partner int	c	Almost 3 years
DWP	DWP variable	d	Almost 4 years
M	Main carer/adult interview	e	Almost 5 years
P	Partner interview	f	Almost 6 years
C	Cohort member (young person) interview	g	Almost 8 years
W	Young person height/weight	h	Around 10-11 years (in Primary 6)
Z	Z-score variable	i	Around 12-13 years (in Secondary 1)
		j	Around 14-15 years (in Secondary 3)
		k	Around 17-18 years (in Secondary 6)

5.3 Variable labels

In the sweep 11 dataset the variable labels have been shortened to 40 characters as far as possible; the first 2 show the source and year of the data (as in the variable name). Although the labels give an indication of the topic of the question **it is essential to refer to the questionnaire documentation to see the full text of the question and the routing applied to that variable**.

5.4 Derived variables

Derived variables included in the dataset are listed with the questionnaire variables for the same topic. The SPSS syntax used to create them is available on request by contacting the GUS research team on gus@scotcen.org.uk.

5.5 Multicoded questions

Some questions in the survey enabled participants to give more than one answer. In the dataset each of the answer options has been converted into a binary variable with the people who selected that option coded 1 and the rest coded 0.

5.6 Indicators and summary variables

5.6.1 Household details collected at sweep 11

In all cases where a household interview took place either in-home or by telephone and with either the cohort member or their parent/carer, details about each member of the household such as their gender, age and relationship to other members of the household were collected, as were details such as employment, income, education and country of birth of the main adult respondent and (where applicable) their resident partner.

In the small number of cases where only a web interview was carried out, no household details were collected at sweep 11. For proxies for these details, users are referred to the GUS BC1 sweep 10 dataset which is also available through the UK Data Service.

5.6.2 Household data

Similar to previous sweeps, the household interview (HI) respondent was asked about each member of the household. The gender, age and marital status of each household member was collected along with their relationship to each other and to the cohort member (young person). Each person in the household was identified by their person number, which they retain through each sweep of the survey. The variable MkHGSl(n) can be used to see whether a person who was in the household at a previous sweep is still in the household at sweep 11.

A set of CAPI derived summary household variables is also included in the data. Amongst other things these detail the number of adults, number of children or number of natural parents in the household. A list of these variables is included in Table 5.2. A set of variables which allow identification of the parent/carer respondent and their partner (if present) in the household grid are also included. These permit easier analysis of adult respondent's and partner's age, marital status and relationship to other people in the household. The age variables have been banded for all persons in the household except the cohort member.

Table 5.2 Key household derived variables

Variable name	Description
MkRespID	Main carer's ID
MkRsex	Main carer's sex
MkPartID	Main carer's partner ID
MkRPsex	Main carer's partners sex
MkHGnp01	Number of biological parents in hhold
MkHGnp02	Biological mother in household
MkHGnp03	Biological father in household
MkMothID	Biological mother's ID
MkFathID	Biological father's ID
Dkhgrsp05	Whether cohort member's main carer is lone parent or living as a couple
MkHGsx1	Study child's sex
Dkhgagc	Study child's age at interview (months)
DkHSize2b	Number of people in the household (banded)

5.6.3 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 sweep 11 dataset includes the five-category system in which cohort members, main carer respondents and the resident partners of main carers, 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. A sixth category 'never worked' is also coded on this variable. The decision on whether or not this category should be included as a separate category, incorporated with category 5 'semi-routine or routine' or set to 'missing' is dependent on the particular analysis to which it is being applied.

Further information on NS-SEC is available from the National Statistics website: SOC 2020 Volume 3: the National Statistics Socio-economic Classification (NS-SEC rebased on the SOC 2020) - Office for National Statistics (ons.gov.uk)

Cohort member employment details were collected for the first time at sweep 11. This data was used to classify their employment details using NS-SEC. The variables included on the dataset are listed in Table 5.3.

Table 5.3 NS-SEC variables

Variable name	Description
DkMsec01	Main carer respondent's NS-SEC (6 category)
DkYsec01	Main carer respondent's partner's NS-SEC (6 category)
DkCsec01	Cohort member's NS-SEC (6 category)

5.6.4 Equivalised household annual income

The income that a household needs to attain a given standard of living will depend on its size and composition. For example, a couple with dependent children will need a higher income than a single person with no children to attain the same material living standards. "Equivalisation" means adjusting a household's income for size and composition so that we can look at the incomes of all households on a comparable basis. Official income statistics use the 'Modified OECD' equivalence scale, in which an adult couple with no dependent children is taken as the benchmark with an equivalence scale of one. The equivalence scales for other types of households can be calculated by adding together the implied contributions of each household member from the table below.

Table 5.4 Income equivalence scales for household members	
Household member	**Equivalence scale**
Head	0.67
Subsequent adults	0.33
Each child aged 0-13	0.20
Each child aged 14-18	0.33

For example, a household consisting of a single adult will have an equivalence scale of 0.67 - in other words he or she can typically attain the same standard of living as a childless couple on only 67 percent of its income. In a household consisting of a couple with one child aged three, the head of the household would contribute 0.67, the spouse 0.33, and the child 0.20, giving a total equivalence scale of 1.20. In other words, this household would need an income 20 percent higher than a childless couple to attain the same standard of living.

GUS collects a banded version of total net household income from all sources in the main carer interview. The midpoint of the band is used to calculate equivalised income. This midpoint income value is adjusted, using the above equivalence scale, according to the characteristics of the household, to produce an equivalised annual household income value. Variables with the full equivalised income scale (DkEqvinc) and quintiles of the scale based on within sample distribution (DkEqv5) are available in the datasets[8].

5.6.5 Area-level variables

Scottish Government Urban/Rural Classification

The dataset includes a binary measure of urban/rural location (ALkurin2). This is based on the Scottish Government's two-fold urban rural classification which is itself derived from the more detailed six-fold classification shown in Table 5.5.

The Scottish Government Urban Rural Classification was first released in 2000 and is consistent with the Government's core definition of rurality which defines settlements of 3,000 or less people to be rural. It also classifies areas as remote based on drive times from settlements of 10,000 or more people. The definitions of urban and rural areas underlying the classification are unchanged.

Table 5.5 Scottish Government Six-fold and Two-fold Urban Rural Classifications		
Classification	**Description – six-fold**	**Description – two-fold**
1. Large Urban Areas	Settlements of over 125,000 people	1. Urban
2. Other Urban Areas	Settlements of 10,000 to 125,000 people	1. Urban

[8] Note previous user guides suggested this variable referred to UK wide income distribution using data from the Family Resources Survey. This is not the case for this sweep nor any previous sweep. Income distribution is considered only amongst the GUS sample.

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	1. Urban
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	1. Urban
5. Accessible Rural	Settlements of less than 3,000 people and within 30 minutes' drive of a settlement of 10,000 or more	2. Rural
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	2. Rural

For further details on the classification see the Scottish Government's website: Scottish Government Urban Rural Classification 2020 - gov.scot (www.gov.scot). A detailed urban/rural variable with all six categories outlined above is available on request under UKDS Secure Licence.

Scottish Index of Multiple Deprivation

The Scottish Index of Multiple Deprivation (SIMD) identifies small area concentrations of multiple deprivation across Scotland. It is based on a number of indicators - 37 indicators in the 2020 version – 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 Crime. SIMD is presented at data zone level, enabling small pockets of deprivation to be identified. The data zones – which for the 2020v2 version have a median population size of 755 – are ranked from most deprived (1) to least deprived (6976) on the overall SIMD and on each of the individual domains. The result is a comprehensive picture of relative area deprivation across Scotland. The GUS sweep 11 dataset contains SIMD 2020v2. It should be noted that analyses in various GUS reports may be based on earlier versions of SIMD.

In the sweep 11 dataset, the data zones are grouped into quintiles in variable ALksimdq2020. 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: https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/.

Further area-level variables (available via UKDS Secure Licence)

Further geographical measures have been derived and are available through UKDS Secure Licence arrangements. These are outlined below.

Data zones

The data zone is the key small-area statistical geography in Scotland providing a common, stable and consistent, small-area geography. The data zone geography covers the whole of Scotland and nests within local authority

boundaries. Data zones are groups of 2001 census output areas and have populations of between 500 and 1,000 household residents. Where possible, they have been made to respect physical boundaries and natural communities. They have a regular shape and, as far as possible, contain households with similar social characteristics[9].

Intermediate geography

Not all statistics are suitable for release at the data-zone level because of the sensitive nature of the statistics, or for reasons of reliability, and it was apparent that a statistical geography between data zone and local authority was required. The intermediate zones are aggregations of data zones within local authorities and contain between 2,500 and 6,000 people[10].

Local authority

Local government in Scotland comprises 32 unitary local authorities, responsible for the provision of a range of public services. Local authority areas (also known as council areas) reflect the geographical diversity within Scotland with wide variations in size (from 60 square miles in Dundee City council area to 25,656 square miles in Highland council area) and population (from under 20,000 people in Orkney Islands council area to over 600,000 in Glasgow City council area).

5.6.6 Strengths and Difficulties Questionnaire (SDQ)

The Strengths and Difficulties Questionnaire (SDQ) is a brief behavioural screening questionnaire designed for use with 3-16-year-olds[11]. The scale includes 25 questions which are used to measure five aspects of the child or young person's development – emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and pro-social behaviour. Further details on the SDQ can be found at sdqinfo.org.

At sweep 11, the full list of SDQ items was asked of the cohort members themselves as part of the self-completion questionnaires (CAWI or CASI). The SDQ items have been asked of the main carer at most sweeps since the cohort member was aged 3 and were asked of the cohort member for the first time at sweep 10.

A score is calculated for each aspect of the young person's development, as well as an overall 'difficulties' score which is generated by summing the scores from all the scales except pro-social. For all scales, except pro-social where the reverse is true, a higher score indicates greater evidence of difficulties.

The dataset includes the constituent items from the cohort member. Variables providing the various composite scores and the total score can be derived with reference to the guidance and syntax provided online at sdqinfo.org. A full list of

[9] Further information on data zones is available from the Scottish Government Scottish Neighbourhood Statistics Guide: https://www2.gov.scot/Publications/2005/02/20697/52626
[10] Further information on intermediate geography is available from the Scottish Government Scottish Neighbourhood Statistics Guide: https://www2.gov.scot/Publications/2005/02/20697/52626
[11] Goodman, R. (1997) "The Strengths and Difficulties Questionnaire: a research note", *Journal of Child Psychology and Psychiatry,* 38, pp581-586

the constituent variables in the cohort member questionnaire is included in Table 5.6.

Table 5.6 Variables associated with the Strengths and Difficulties Questionnaire

Variable name	Description
CkCSDQni	Ck: I try to be nice to other people. I care about their feelings.
CkCSDQrt	Ck: I am restless, I find it hard to sit down for long.
CkCSDQac	Ck: I get a lot of headaches, stomach-aches or sickness.
CkCSDQsh	Ck: I usually share with others for example food or drink.
CkCSDQan	Ck: I get very angry and often lose my temper.
CkCSDQal	Ck: I would rather be alone than with other people.
CkCSDQto	Ck: I am generally willing to do what other people want.
CkCSDQwo	Ck: I worry a lot.
CkCSDQhe	Ck: I am helpful if someone is hurt, upset or feeling ill.
CkCSDQfi	Ck: I am constantly fidgeting or squirming.
CkCSDQfr	Ck: I have at least one good friend.
CkCSDQfg	Ck: I fight a lot. I can make other people do what I want.
CkCSDQun	Ck: I am often unhappy, depressed, or tearful.
CkCSDQli	Ck: Other people generally like me.
CkCSDQdi	Ck: I am easily distracted, I find it difficult to concentrate.
CkCSDQne	Ck: I am nervous in new situations. I easily lose confidence.
CkCSDQki	Ck: I am kind to children.
CkCSDQly	Ck: I am often accused of lying or cheating.
CkCSDQpb	Ck: Other people pick on me or bully me.
CkCSDQvo	Ck: I often offer to help others (family members, friends, colleagues).
CkCSDQth	Ck: I think before I do things.
CkCSDQst	Ck: I take things that are not mine from home, work or elsewhere.
CkCSDQgo	Ck: I get along better with older people than with people my own age.
CkCSDQfe	Ck: I have many fears. I am easily scared.
CkCSDQwk	Ck: I finish the work I'm doing. My attention is good.

5.6.7 Mental wellbeing: selected items from the Students' Life Satisfaction Scale

Life satisfaction is measured through the use of selected items from the Students' Life Satisfaction Scale[12]. These items were asked as part of the self-completion questionnaire for the cohort member and were previously asked at sweep 7, 9 and 10. Relevant variables are listed in Table 5.7.

Table 5.7 Selected items from the Students' Life Satisfaction Scale

Variable name	Description
CkWed	Do you wish your life was different?
CkWer	Do you feel that your life is just right?
CkWea	Do you feel you have what you want in life?
CkWeg	Do you feel you have a good life?

[12] Huebner, E. S. (1991). Initial Development of the Student's Life Satisfaction Scale. School Psychology International, 12(3), 231-240. https://doi.org/10.1177/0143034391123010

5.6.8 Anxiety: Generalised Anxiety Disorder Assessment (GAD-7)

The GAD-7[13] is a brief self-report scale designed as a screen for symptoms of Generalised Anxiety Disorder (GAD). The 7 items are scored 0-3 (total score range 0-21), reflecting the frequency of experiencing symptoms of GAD in the past 2 weeks. The GAD-7 shows acceptable internal reliability and had been validated as a screen for GAD in both clinical and population samples. Total scores of 5, 10 and 15 represent cut-points for mild, moderate and severe anxiety[14].

Questions were asked of the cohort member as part of the self-completion questionnaire and were also included at sweep 10.

Table 5.8 GAD-7

Variable name	Description
CkGadNer	In last 2 weeks, bothered by: Feeling nervous, anxious or on edge?
CkGadWoS	In last 2 weeks, bothered by: Not being able to stop or control worrying?
CkGadWoD	In last 2 weeks, bothered by: Worrying too much about different things?
CkGadRel	In last 2 weeks, bothered by: Having trouble relaxing?
CkGadRes	In last 2 weeks, bothered by: Being so restless that it is hard to sit still?
CkGadAnn	In last 2 weeks, bothered by: Becoming easily annoyed or irritable?
CkGadAfr	In last 2 weeks, bothered by: Feeling afraid as if something awful might happen?

5.6.9 Depression: Selected items from the Patient Health Questionnaire (PHQ-9)

The PHQ-9[15] includes the full 9-item Depression Module of the self-report Patient Health Questionnaire, designed to screen for depression symptoms in the past two weeks. The items reflect the 9 DSM-IV criteria for depressive disorders. Each item is rated on a 4-point scale (score 0-3, total score range 0-27), reflecting the frequency with which symptoms are experienced. The PHQ-9 has been extensively evaluated and found to be valid as both a severity and diagnostic measure in both patient and community samples. On the full scale, a cut-off score of >=10 has been found to maximise combined sensitivity and specificity overall, and for subgroups, by comparison with semi-structured diagnostic interview assessments of depression. Scores of 10 or higher on the full scale are commonly used to identify individuals with depression[16].

[13] Spitzer R. L., Kroenke K., Williams J. B. W., Löwe B. (2006). A brief measure for assessing generalized anxiety disorder the GAD-7. Arch. Intern. Med. 166 1092–1097. 10.1001/archinte.166.10.1092

[14] *Catalogue of Mental Health Measures* (2023). Available at www.cataloguementalhealth.ac.uk

[15] Kroenke K., Spitzer R. L., Williams J. B. (2001). The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16 606–613. 10.1046/j.1525-1497.2001.016009606.x

[16] *Catalogue of Mental Health Measures* (2023). Available at www.cataloguementalhealth.ac.uk

A subset of 8 questions was asked of the cohort member as part of the self-completion questionnaire. Variables on the dataset are set out in Table 5.9. On the full, formal scale items CkPHQMo and CkPHQRe are asked as a single question.

Table 5.9 Patient Health Questionnaire (PHQ-9)

Variable name	Description
	Over the last two weeks, how often have you been bothered by any of the following problems:
CkPHQSl	Trouble falling or staying asleep, or sleeping too much?
CkPHQTi	Feeling tired or having little energy?
CkPHQAp	Poor appetite or overeating?
CkPHQBa	Feeling bad about yourself - or that you are a failure or have let yourself or your family down?
CkPHQCo	Trouble concentrating on things, such as reading the news or watching television?
CkPHQMo	Moving or speaking so slowly that other people could have noticed?
CkPHQRe	Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual?
CkPHQDe	Thoughts that you would be better off dead, or of hurting yourself in some way?

5.6.10 Parent physical and mental wellbeing

Mental wellbeing for parent/carer respondents was measured using selected items from the Medical Outcomes Study 12-Item Short Form (SF-12)[17]. This measure was previously used at multiple previous sweeps in its full 12-item form. At sweep 11, only questions related to mental wellbeing were included along with the general health item.

The published original scoring instructions[18] rely on all 12 items being included. Analysts wishing to compare sweep 11 with previous sweeps should devise an alternative scoring approach. Table 5.10 lists the individual item variables.

Table 5.10 Constituent variables associated with the SF-12

Variable name	Description
MkHpgn01	Mk How is resp health in general
MkHlmt05	Mk Resp mental health limited accomplishments past 4 wks
MkHlmt06	Mk Resp mental health limited quality of accomplishments past 4 wks
MkHpgn02	Mk Time resp felt calm in past 4 wks
MkHpgn03	Mk Time resp felt energetic in past 4 wks
MkHpgn04	Mk Time resp felt down in past 4 wks
MkHpgn05	Mk Time resp health interfered socially in past 4 wks

[17] Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34:220–233. doi: 10.1097/00005650-199603000-00003.

[18] Ware J, Kosinski M, Keller SD. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales (Second Edition). 1998. Available here: https://www.researchgate.net/publication/242636950_SF-12_How_to_Score_the_SF-12_Physical_and_Mental_Health_Summary_Scales.

5.6.11 Alcohol and smoking

As part of their self-completion questionnaire, cohort members were asked a number of questions about alcohol and smoking. Questions were adapted from the Health Behaviour in School Aged Children Survey (HBSC[19]) and are listed in Table 5.11.

Questions about alcohol and smoking were also asked at sweeps 9 and 10. Therefore, not all cohort members answered all questions about alcohol and smoking at sweep 11 (e.g. 'whether ever had alcoholic drink' was not asked of those who had already reported having an alcoholic drink at sweep 10).

Table 5.11 Selected items from HBSC on child's health behaviours - alcohol and smoking (young person questionnaire)

Variable name	Description
CkBSm2	Ck Whether ever tried a cigarette
CkBSn	Ck How often smokes now
CkBSe2	Ck Whether ever tried e-cigarette or vaping device
CkBESnC	Ck How often smoke e-cigarettes or vaping devices now
CkBSa2	Ck Whether ever had alcoholic drink
CkBAl	Ck How often drank alcohol in the last 30 days
CkAlfn	Ck How many times have had five or more alcoholic drinks at a time in the last 12 months?

5.6.12 Drug use

Alongside questions on alcohol and smoking, cohort members were also asked about drug use. These items were adapted from the Longitudinal Study of Australian Children (LSAC20), Growing Up in Ireland (GUI21) and the Scottish Adolescent Lifestyle and Substance Use Survey (SALSUS22). See Table 5.12 below.

Table 5.12 Items on drug use

Variable name	Description
CkDrugMe	Ck Have you ever tried cannabis?
CkDrugMo	Ck How often have you used cannabis?
CkDrugOe	Ck Have you ever tried any drugs other than cannabis?
CkDrugOo	Ck How often have you used drugs other than cannabis?
CkDrugOl..	Ck Which drugs other than cannabis have you ever tried?
CkDrugOl_1	Prescription only painkillers
CkDrugOl_2	Cocaine
CkDrugOl_3	Ecstasy
CkDrugOl_4	Poppers
CkDrugOl_5	Tranquilisers
CkDrugOl_6	LSD
CkDrugOl_7	Other

[19] https://hbsc.org/about/
[20] https://growingupinaustralia.gov.au/
[21] https://www.growingup.gov.ie/
[22] https://www.gov.scot/collections/scottish-schools-adolescent-lifestyle-and-substance-use-survey-salsus/

CkDrugOI_8	None of the above

5.6.13 Anti-social behaviour and offending

Questions on anti-social behaviour and offending are mostly adaptations of questions previously asked in the Edinburgh Study of Youth Transitions and Crime[23].

At sweep 11, for each of the behaviours listed in Table 5.13 below, cohort members were asked how many times they had engaged in a particular form of behaviour in the last year. These questions have also been asked at previous sweeps.

Table 5.13 Items on anti-social behaviour and offending - adapted from the Edinburgh Study of Youth Transitions

Relevant variable names	Description
CkASBsy	Ck: In last year: how many times taken something from a shop or a store
CkASBry	Ck: In last year: how many times been rowdy or rude in public
CkASBmy	Ck: In last year: how many times stolen money or other things
CkCRED	Ck: In last year: how many times online fraud
CkASBky	Ck: In last year: how many times carried a knife or weapon
CkASBpy	Ck: In last year: how many times deliberately damaged or destroyed property
CkASBby	Ck: In last year: how many times broken into a locked place to steal something
CkASBgy	Ck: In last year: how many times written things or sprayed paint on property
CkASBwy	Ck: In last year: how many times used force, threats or a weapon etc.
CkASBhy	Ck: In last year: how many times hit, kicked or punched someone
CkCOMP	Ck: In last year: hacking or malicious IT
CkHRSD	Ck: In last year: harassed or bothered someone by phone, email or social media
CkRMRS	Ck: In last year: sent pictures or spread rumours by phone, email or social media

5.6.14 Parent-Child Attachment: selected items from the People In My Life (PIML) scale

The People in My Life measure is a self-report instrument designed to measure attachment to parents and peers in middle childhood. The sweep 11 main carer and cohort member questionnaires both included selected items from the Parent Attachment scale. Further information about the PIML scale can be found on the Fast Track Project website: http://fasttrackproject.org/techrept/p/pml/

At sweep 11, the cohort member was asked these questions about up to two resident parents and up to two parents living elsewhere. Different participants

[23] https://www.edinstudy.law.ed.ac.uk/

may have answered each set of questions in relation to different parents depending on who the parent respondent was at the previous sweep and any changes in household composition since the previous sweep. The particular parent the cohort member is responding about in any set of questions is defined in variables CkPa1Rel, CkPa2Rel, CkPaNRel, CkPa3Rel and CkPe1ReC

Please also note that the approaches to identifying parents are also different to those applied at previous sweeps. Users should consult the relevant **questionnaire documentation**.

Table 5.14 outlines the relevant variables on the dataset.

Table 5.14 Selected items from People In My Life scale (cohort member, main carer and partner questionnaires)

Variable name	Description
Cohort member questionnaire	
CkPar101, CkPar201, CkPar301, CkParn01, CkPew101C	Ck: … he/she listens to what I have to say.
CkPar102, CkPar202, CkPar302, CkParn02, CkPew102C	Ck: … I can count on him/her to help me when I have a problem.
CkPar103, CkPar203, CkPar303, CkParn03, CkPew103C	Ck: … I talk to him/her when I am having a problem.
CkPar105, CkPar205, CkPar305, CkParn05, CkPew105C	Ck: … I share my thoughts and feelings with him/her.
CkPar106, CkPar206, CkPar306, CkParn06, CkPew106C	Ck: … he/she pays attention to me.
Main carer questionnaire	
MkPall	Mk I listen to what child has to say
MkPalu	Mk I can tell when child is upset about something
MkPAlt	Mk child talks to me when child is having a problem
MkPAlb	Mk If I know something is bothering my child, I ask about it
MkPAla	Mk I pay attention to child, even when I am busy
MkPAls	Mk Child shares thoughts and feelings with me

5.6.15 Peer Attachment: selected items from the People In My Life (PIML) scale

In addition to the items on parent-child communication outlined above, the sweep 11 self-completion questionnaire for the cohort member also included selected items from the PIML Peer Attachment Scale. Relevant items are outlined in Table 5.15.

Table 5.15 Selected items from People In My Life Peer Attachment scale (cohort member questionnaire)

Variable name	Description
CkCrFrl	Ck My friends listen to what I have to say

CkCrFrc	Ck I can count on my friends to help me when I have a problem
CkCrFrt	Ck I talk to my friends when I am having a problem
CkCrFrb	Ck If my friends know something is bothering me, they ask me about it
CkCrFrs	Ck I share my thoughts and feelings with my friends
CkCrFra	Ck My friends pay attention to me

5.6.16 Additional new measures sourced elsewhere

In addition to the items outlined above, Table 5.16 details items new to GUS at sweep 11 which were either directly sourced elsewhere or adapted from existing sources.

Table 5.16 Young person's CAPI interview

Variable name(s)	Description	Source(s)
	Political attitudes and civic participation	Survey of Young Scots[24] Scottish Social Attitudes Survey[25]
	Activities in free time	Adapted from the Millennium Cohort Study (Age 17)
	Diet	Adapted from Food and You, Food Standards Agency
CkSleSc, CkSleNs, CkSleWe	Sleep	Adapted from Growing Up in Australia
CkSSlvy, CkStchw to CkSath	Recent experience of and attitudes towards school	Adapted from Growing Up in Ireland and Scottish School Leavers Study
CkSubChr1[1-13]	Reasons for subject choices	Adapted from Next Steps
CkLcoca to CkCovwjp	Impact of COVID on learning	Adapted from Growing Up in Ireland
CkCasAS	Educational aspirations	Adapted from the Millennium Cohort Study (Age 11)
CkCrad to CkCrinot	Careers advice	Adapted from Scottish School Leavers Study and Skills Development Scotland
CkOInSou	Sources of income	Adapted from the Millennium Cohort Study (Age 17)
CkFutca to CkJobthr	Aspirations and future plans	Adapted from Scottish School Leavers Study and Skills Development Scotland
CkTrsw, CkTRTI	Commuting	Adapted from the Millennium Cohort Study (Age 17)

[24] See Eichhorn, J. (2018) Mobilisation through early activation and school engagement – the story from Scotland, Journal of Youth Studies, 21:8, 1095-1110, DOI: 10.1080/13676261.2018.1450968

[25] See https://ssa.natcen.ac.uk/

Table 5.16 Young person's CAWI/CASI questionnaire		
Variable name(s)	**Description**	**Source**
CkLone	Loneliness	Adapted from Scottish Household Survey
CkFeFo to CkFeff	Fear of failure	Adapted from Programme for International Student Assessment
CkGmin	Growth mindset	Adapted from Programme for International Student Assessment
CkSBel, CkSOut	Sense of belonging	Adapted from Programme for International Student Assessment
CkSoMeAn to CkSoMeTi	Social media	Adapted from Realigning Children's Services Wellbeing Survey
CkSexOr to CkSexPart	Sexual relationships/activity	Adapted from Young Person's Behaviour and Attitudes Survey 2013 Version B
CkDEAN to CkTRDV	Mental health	Adapted from the Millennium Cohort Study
CkCONTRL	Control	CLS COVID-19 Survey
CkOptFu,	Optimism	CLS COVID-19 Survey
CkRISK	Attitudes to risk	CLS COVID-19 Survey

Table 5.16 Main carer questionnaire		
Variable name(s)	**Description**	**Source**
MkVoteSP to MkPolInY	Political attitudes and civic participation	Survey of Young Scots[26] Scottish Social Attitudes Survey[27]
CkOptFu, CkOptCh	Optimism	CLS COVID-19 Survey
MkWcvp	Impact of COVID on employment	Adapted from Understanding Society

5.6.17 Cohort member height and weight measurements: Body Mass Index (BMI) scores

Body Mass Index (BMI), i.e. weight divided by height squared, is a score that adjusts a person's weight for their height. Taken as a number in isolation, the BMI it does not actually represent anything medically. It is only meaningful in the context of a distribution of values for a population. Individuals are placed into bands to show where they stand in relation to the rest of the population, in particular whether they have unusually high or low BMI.

In adults BMI stays fairly constant on average as people get older. Therefore, BMI categories for adults ignore age and calculate the same BMI for two people

[26] See Eichhorn, J. (2018) Mobilisation through early activation and school engagement – the story from Scotland, Journal of Youth Studies, 21:8, 1095-1110, DOI: 10.1080/13676261.2018.1450968

[27] See https://ssa.natcen.ac.uk/

with the same weight and height regardless of the differences in their ages. Thus adults (aged 16 and over) can usually be classified into the following BMI groups:

- Less than 18.5 - Underweight
- 18.5 to less than 25 - Normal/Healthy weight
- 25 to less than 30 - Overweight
- 30 to less than 40 - Obese
- 40 and above - Morbidly obese

However, among children and young people BMI changes as the child or young person ages. Since having a certain BMI at one age may be the norm but be unusually high or low at another age, different centiles are calculated for different ages.

Along with BMI, two principle sets of cohort member overweight and obesity variables have been included on the dataset:

- International Obesity Taskforce (IOTF) cut-offs.
- Public Health Scotland/Information Services Division (ISD) centiles

The IOTF cut-offs are based on BMI reference data from six different countries around the world (over 190,000 subjects in total aged 0 to 25 from UK, Brazil, Hong Kong, the Netherlands, Singapore, and the United States)[28]. In summary, the BMI percentile curves that pass through the values of 25 and 30 kg/m 2 (standard adult cut-off points for overweight and obesity, respectively) at age 18 were smoothed for each national dataset and then averaged.

The averaged curves were then used to provide age and sex-specific BMI cut-off points for children and adolescents aged 2 to 18[29]. By averaging the distribution curves from each reference country, the international cut-offs for children purport to be representative of the countries but independent of the overweight or obesity level in each country.

One of the benefits of using these international standards is the possibility of making international comparisons. However, the international classification is not without problems: international reference data differ from those for the UK population, and this is reflected in the sex-specific overweight and obesity estimates produced by the International classification.

In light of this lack of consensus, variables have also been produced using Public Health Scotland's (PHS, formerly Information Services Division - ISD) method. This produces BMI centiles (Cole's LMS method), taking into account the fact that BMI data does not follow a normal distribution. Cohort member BMIs are converted to standard deviation scores/centiles in order to compare them to the growth reference data and assign children to the various categories of (un)healthy weight. Each participant's BMI is calculated then converted into SD scores/centiles, using the UK 1990 growth reference data based on sex and age in months and Cole's LMS method. Further information on this approach

[28] Cole, T.J. and Lobstein, T. (2012) Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatric Obesity. https://doi.org/10.1111/j.2047-6310.2012.00064.x

[29] Lookup tables can be accessed via this link: https://www.worldobesity.org/about/about-obesity/obesity-classification

can be found in the technical reports which accompany the Primary 1 Body Mass Index (BMI) national statistics publications[30].

Percentile cut-off	PHS description
At or below 5th percentile	Underweight
Above 5th percentile and below 85th percentile	Healthy weight
At or above 85th percentile and below 95th percentile	Overweight
At or above 95th percentile and below 98th percentile	Obese
At or above 98th percentile	Morbidly obese

Note that only those height and weight measurements considered by the interviewer to be reliable were used to calculate the BMIs.

Table 5.17 Cohort member (young person) derived BMI variables

Variable name	Description
Dkbmi	Dk BMI (reliable interviewer measurements only)
Dkbmi_sr	Dk BMI using self-reported height and weight
DkINTbmi	Dk International BMI cut-offs (measured)
DkINTbmi_sr	Dk International BMI cut-offs (self-reported)
Dkisdbmi	Dk BMI - 5 groups ISD classification (measured)
Dkisdbmi_sr	Dk BMI - 5 groups ISD classification (self-report)

5.7 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 transformed into derived variables instead).

The following types of variables have been deleted or replaced with a derived variable coded into broader categories in order to reduce the potential to identify individuals:

1. Those containing text

2. Those which contained a personal identifier (e.g. name/address)

3. Those considered to be disclosive, such as:

    - Detailed ethnicity
    - Detailed religion
    - Detailed geography variables
    - Language spoken at home
    - Full interview date
    - Full date of birth
    - Timing variables

[30] For example see https://beta.isdscotland.org/find-publications-and-data/population-health/child-health/primary-1-body-mass-index-bmi-statistics-scotland/

There are no geographical variables in the archived dataset beyond a binary area urban-rural classification and the Scottish index of multiple deprivation summary variable. As noted in section 5.6.5, access to more detailed geographic variables is possible via the UKDS Secure Licence facility. Access to variables with text/string content is available on request at the discretion and agreement of Scottish Government as the data controller.

5.8 Missing values conventions

The following missing values conventions have been observed:

-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.

6 Documentation

The documentation includes the following:

- Questionnaires (with variable names added)
- List of variables in the dataset
- Showcards
- Survey invitation letters and information leaflet
- Interviewer (project) instructions
- CAPI edit spec

7 Contact details

Further details about the study and a list of publications using the data can be found on the study website: growingupinscotland.org.uk.

Queries should be directed to the GUS team at the Scottish Centre for Social Research:

gus@scotcen.org.uk

Scottish Centre for Social Resesarch

Scotiabank House

6 South Charlotte Street

Edinburgh, EH2 4AW

T: 0131 240 0210