COSMO

COVID Social Mobility
& Opportunities Study

Wave 2 November 2023

Data User Guide

Version 1

Tugba Adali, Jake Anders, Lisa Calderwood, Carl Cullinane, Becky Hamlyn, Jonathan Kennett, Xin Shao, Luke Taylor, David Xu

Contact

Data queries: help@ukdataservice.ac.uk

Authors

Tugba Adali (CLS), Jake Anders (CEPEO), Lisa Calderwood (CLS), Carl Cullinane (Sutton Trust), Becky Hamlyn (Kantar Public), Jonathan Kennett (Kantar Public), Xin Shao (CEPEO), Luke Taylor (Kantar Public), David Xu (Kantar Public).

How to cite this guide

Adali, T., Anders, J., Calderwood, L., Cullinane, C., Hamlyn, B., Kennett, J., Shao, X., Taylor, L., Xu, David. (2023) COVID Social Mobility & Opportunities study (COSMO): Wave 2 User Guide (Version 1). London: UCL Centre for Longitudinal Studies and UCL Centre for Education Policy & Equalising Opportunities.

This guide was published in November 2023 by the UCL Centre for Education Policy and Equalising Opportunities (CEPEO), UCL Centre for Longitudinal Studies (CLS), Sutton Trust, and Kantar Public.

The UCL Centre for Education Policy and Equalising Opportunities (CEPEO) is a research centre that carries out cutting-edge research focused on equalising opportunities across the life course. Its work seeks ways to improve education policy and wider practices to achieve this goal. For more information, visit www.ucl.ac.uk/ioe/cepeo.

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Social Research Institute, University College London. It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ucl.ac.uk.

The Sutton Trust champions social mobility through programmes, research and policy influence. Since 1997 and under the leadership of founder Sir Peter Lampl, the Sutton Trust has worked to address low levels of social mobility in the UK. The Trust works to improve social mobility from birth to the workplace so that every young person – no matter who their parents are, what school they go to, or where they live – has the chance to succeed in life. For more information, visit www.suttontrust.com.

Kantar Public UK is an independent research agency that manages high quality cross-sectional and longitudinal research to inform the development of public policy, public service delivery and public communication on behalf of UK government and public sector organisations, academic institutions, and charities. For more information, visit www.kantarpublic.com/

For questions and feedback about this user guide: help@ukdataservice.ac.uk

This document is available in alternative formats. Please contact the Centre for Longitudinal Studies: clsfeedback@ucl.ac.uk

Please visit the study website for initial findings, questionnaires, and news and updates: www.cosmostudy.uk

Contents

1   Introduction

1.1   Background

The COVID Social Mobility & Opportunities study (COSMO) seeks to generate high-quality evidence to answer the central research question of how the COVID-19 pandemic affects socio-economic inequalities in life chances, both in terms of short-term effects on educational attainment and well-being, and long-term educational and career outcomes. To achieve this aim, a representative sample of young people who were in Year 11 in the 2020/2021 academic year across England were invited to a survey between September 2021 and April 2022, with the intention of following them over time as they progress through the final stages of education and into the labour market. The study also included a survey with a parent or guardian[1] of the young person to complement the young person's data.

All young people and parents who took part in Wave 1 were invited to the second Wave of the study. In Wave 1, there were some young people whose parents did not take part. In Wave 2, we invited these parents again, to help enrich background information for more young people in the sample. Wave 2 focussed on how young people's lives have changed since the first wave in terms of their educational or vocational activities, and their future plans. Parents were also asked how they are getting on, including topics like the cost of living.

COSMO is carried out by a collaboration between UCL Centre for Education Policy & Equalising Opportunities (CEPEO), the UCL Centre for Longitudinal Studies (CLS), the Sutton Trust and Kantar Public. The project is further supported by key stakeholders to ensure co-production of policy-relevant evidence including: the Department for Education (DfE), the Office for Students (OfS), Administrative Data Research (ADR UK), the Education Endowment Foundation (EEF), Transforming Access and Student Outcomes in Higher Education (TASO).

This second wave of the study was funded by UKRI Economic and Social Research Council under grant ES/X00015X/1. In addition, the Sutton Trust invested in an 'add on' to the main study (which we refer to as the Sutton Trust boost sample throughout this user guide), focusing on disadvantaged young people with high prior attainment, which they have been funding since the beginning of the study.

Similar to Wave 1, Wave 2 of COSMO was designed as an online-first sequential mixed mode study, which built on the success of the online-first approach for young people at

---

[1]   Any parent or guardian of a sampled young person was eligible for this survey. "Parents/guardians" and "parents" are used interchangeably in this guide.

Wave 1. The data collection for Wave 2 was carried out between October 2022 and April 2023. Data was collected online first, then via a face-to-face and telephone follow-up, as finally an online 'mop-up' fieldwork stage. A proposed Wave 3 of the study is planned subject to availability of funding.

This User Guide accompanies the initial data deposit of Wave 2 data to UK Data Service. Work is ongoing with the DfE to make linked data from the National Pupil Database (NPD) available with COSMO, most likely through the ONS Secure Research Service.

1.2 Investigators

Decisions around substantive and methodological issues on COSMO were taken by a team of investigators led by Jake Anders (CEPEO) (Principal Investigator), and including Lindsey Macmillan (CEPEO), Gill Wyness (CEPEO), Claire Crawford (CEPEO), Lisa Calderwood (CLS), Alissa Goodman (CLS), Praveetha Patalay (CLS), and Carl Cullinane (Sutton Trust).

1.3 Ethics

The study design and the tools to be used for COSMO were approved by the UCL IOE Research Ethics Committee (REC1660). This application covered sampling, incentive approach, data linkage consents, participant information, privacy notice, signposting to sources of support, survey mode, questionnaires and any other relevant dimensions of the study.

2 Sampling

In this section we provide a broad overview of the target population for the study and the sampling frames used. Following on from this, a summary is provided outlining how the sample was drawn from each frame ahead of Wave 1, and the sample which was issued for Wave 2.

2.1 Target population, sampling frame and coverage

The estimation population consists of all children in England studying in Year 11 in the 2020/2021 academic year.

Two sample frames were used:

- the DfE National Pupil Database (NPD) of Year 11 children in state schools, as recorded in the Spring 2020/2021 pupil-level census[2]

- A subset of the publicly available DfE Get Information About Schools database (GIAS)[3] covering independent schools with Year 12 pupils in the 2021/2022 Academic Year

Some children appeared in both sample frames: specifically, those that moved from a state school in Year 11 to an independent school in Year 12. These respondents were identified via data collected in the survey questionnaire and the weighting compensated for this (see section 6 of the Wave 1 user guide).

Those studying in very small schools were excluded from both sample frames. The total non-coverage rate among state school children was 0.8%, although it was slightly higher for children in alternative provision or special schools. The non-coverage rate among independent school children was higher: estimated[4] at 9%.

---

[2] The fieldwork timings (beginning in September 2021) did not allow the 2021/2022 NPD to be used for the state school sampling.

[3] https://www.get-information-schools.service.gov.uk/

[4] This is an estimate only because the number of Year 12 children in each school is not recorded in the GIAS database but is inferred by taking the total number of pupils in each school and dividing that by the number of school years covered by the school. There is evidence from the Independent Schools Council (ISC) that Year 12 tends to have fewer pupils than other school years, so this inferential method may well lead to an over-estimate of the number of Year 12 pupils in the school.

These non-covered children remain part of the estimation population and the weighting design (outlined in section 6 of the Wave 1 user guide) is designed to compensate for this non-coverage.

One other group – those children who were in an independent school in Year 11 but moved to a state school for Year 12 – are entirely uncovered. In theory, this group is part of the estimation population but, because it is missing from both sample frames, there is no way to weight the data to compensate for this non-coverage (thought to be <1%).

2.2 Wave 1 Sample design

A summary of the wave 1 sample design is provided below. Further details can be found in the Wave 1 User Guide.

2.2.1 State schools

In drawing the sample, we oversampled pupils from disadvantaged backgrounds (those eligible for free school meals (FSM) at any time in the last six years) and those from the six main minority ethnic groups (Indian, Pakistani, Bangladeshi, Black Caribbean, Black African and Mixed).

We used a multi-staged sampling approach. At stage one, 750 schools (Primary Sampling Units (PSUs)) were sampled using a Probability Proportionate to Size (PPS) approach.

At the second stage of sampling, a stratified random sample of students was drawn from each sampled school, with sampling fractions varying between types of students.

33,719 pupils were issued into field, and there were 3,275 reserve cases that were not issued into field.

The Sutton Trust boost sample was drawn after the main study sample was selected. The sample for the boost sample was drawn from the schools selected as original issue for the main study.

The definition of pupils included in the boost sample was as follows:

• Eligible for FSM in last 6 years AND

• In the top 33% in the combined reading, maths, and GPS (Grammar Punctuation and Spelling) KS2 score (the score weighted as follows: maths 50%, reading 25% & GPS 25%)

Within the original issue PSUs, there were 2,868 pupils that were eligible for the Sutton Trust boost that had not been selected for the main study. From these pupils, a further random sample of 2,000 were selected for the boost.

The use of the NPD as a sampling frame for state schools was made possible through a Data Sharing Agreement[5] between UCL, Kantar Public and the DfE, following an application.

2.2.2 Independent schools

A systematic random PPES (Probability Proportionate to Estimated Size) sample of 240 schools was drawn from the DfE Get Information About Schools database. School sampling probabilities were proportionate to the estimated number of Y12 pupils in the school. There were two explicit strata: (i) independent schools (228 selected) and (ii) independent special schools (12 selected).

Cooperating schools were then asked to distribute the survey invitations to pupils and their parents/carers using a random sampling protocol agreed with Kantar Public.

Cooperating schools also provided information for us to use at the weighting stage to calculate the within-school pupil sampling probability.

2.4 Wave 2 Sample design

As also covered in Chapter 4, the issued sample at Wave 2 consisted of any households where a young person responded at Wave 1 (from the main survey or Sutton Trust boost). Any households where *only* a parent had responded at Wave 1 were not included in Wave 2.

In total, 13,786 records were selected and issued into field at Wave 2. This consisted of 13,112 records sampled from the NPD (state school) sample and 674 from the independent school sample.

For 10,050 households we were looking to achieve repeat interviews with both a young person and parent who also took part in Wave 1. For the remaining 3,735 households, only the young person took at Wave 1. In these households we were seeking a repeat interview with the young person and a first-time interview with a parent.

---

[5] DSAP number DS 00554.

3 Overview of questionnaires

3.1 Development

As will be covered further below, two questionnaires were designed as part of the Wave 2 COSMO study: Young Person and Parent/guardian. These were developed over the course of May–August 2022, and were programmed into Kantar Public's scripting software by October 2022.

To inform development of questionnaire content, meetings were held with various stakeholders, and input was received from researchers, governmental organisations and funders. The scientific and technical development of the questionnaires was supported by the investigators of COSMO (Lisa Calderwood, Claire Crawford, Carl Cullinane, Alissa Goodman, Lindsey Macmillan, Praveetha Patalay, Gill Wyness) led by Jake Anders, working with Kantar Public.

In developing the questionnaires, other relevant surveys were consulted and pre-existing questions including established scales like the 12-item General Health Questionnaire (GHQ-12) or the 2-item Generalised Anxiety Disorder (GAD-2) were used or adapted where possible, to build on prior experience and ensure comparability. These surveys include, but are not limited to, the Longitudinal Survey of Young People in England: Cohort 2 (LSYPE 2, also known as "Our Future"), LSYPE 1 (also known as "Next Steps"), the Millenium Cohort Study, and the Your Life, Your Future Survey. A number of new questions were also developed.

To test comprehension and validity of questions, cognitive testing was carried out with both young people and parents, focussing on a selection of proposed questions from both the young person and parent/guardian questionnaires. This informed decisions around final wording and content of these questions.

Similar to Wave 1, COSMO Wave 2 had very tight timescales, as the aim was to catch-up with study members about a year following their first interviews, around the same time of the academic year. These timescales did not allow for a pilot stage to test questionnaire flow, fieldwork processes and interview length. Therefore, a small number of informal pilot interviews were carried out by the research team using informal networks to ensure the questionnaire worked well, and to derive approximate timing estimates whilst also drawing from the experience of Wave 1.

All questions for the Young Person and Parent questionnaires were designed to work in both web and face-to-face modes. For the web survey, the entire questionnaire was self-completed online. For the face-to-face survey, the more sensitive questions were administered as self-completion (CASI) which respondents completed via the

interviewer's tablet. For the telephone mode (CATI), some questions in these CASI modules were left out from the script due to concerns around privacy and sensitivity in the young person questionnaire. These are marked in Table 1. There were no questions excluded from the parent questionnaire for CATI.

3.2 Overview of content

The overarching aim of COSMO is to provide a representative data resource to support research into the lives and life chances of young people with different characteristics in the wake of the COVID-19 pandemic and its aftermath, in terms of short-term effects on educational attainment and wellbeing, and long-term educational and career outcomes. The unit of analysis is young people, but as mentioned earlier, parents were interviewed as well to complement the data collected from young people, enriching the data with information on socio-economic background, and providing direct reports of parents' experiences during the pandemic.

In Wave 2 questionnaires for young people and parents, we used some responses given at Wave 1 as feed forward variables, allowing participants to review and build on their previous response (for instance about their activities at the time of Wave 1). We also used feed forward variables to reduce respondent burden (for instance, if a respondent reported their ethnicity in Wave 1, they were not asked again in Wave 2). These variables are listed on the questionnaires accompanying this deposit.

Below the two questionnaires are summarised further.

3.3 Young People Questionnaire

By the time Wave 2 started, even though COVID-19 infections were still common, its disruption to everyday life was significantly reduced compared to Wave 1. With Wave 1 data about the COVID-19 disruptions on young people's education retaining its longitudinal significance, there was no further need in Wave 2 for a strong emphasis on COVID-19 in the questionnaire.

Rather, the Wave 2 questionnaire focussed on different paths young people might be taking at this stage of their lives: they were asked whether their main activity had changed since the last interview, and details about their current main activity is at the time of the interview. This involved further details about vocational training, labour market experience, university applications and apprenticeship experience.

Also, this time around, new questions were included which covered young people's sexual orientation, any chronic illnesses they may have, mental health seeking behaviour,

whether young people were UK-born and their religion, to provide further context on young people's experiences.

As in Wave 1, two questions in Section H: Attitudes to Education and Future Careers (ZACCESS and ZAPPLY) were only asked to a specific subgroup of the sample, which was a *boost sample* funded by Sutton Trust. Please refer to the Wave 1 User Guide sample design section for the details of this sample.

All young people were asked for their consent to link named sources of administrative data to their survey data, including re-asking for linkage consent for a sub-set of sources of administrative data requested at Wave 1. Details of these are provided in section 3.6.

As mentioned earlier in this section, some questions in the young people questionnaire were not asked when interviews were done by CATI. These are indicated in Table 1, where a summary of the content is provided. The full questionnaires, annotated with variable names, are available within this same data release and are also available on the COSMO study website.

Table 1. Young Person questionnaire content at COSMO Wave 2

Section	Topic
A. Introduction and verification checks	Verification of name and address from the Wave 1 sample
	Verification of date of birth
B. Household grid	Number of household members
	Gender of household members
	Age of household members
	Relationship of household members to YP
	If not living with parents, reasons for this and length of time lived apart from them
C. Current status	Whether YP's main status is same as Wave 1
	If main activity has changed or has been non-continuous, main status between Waves 1 and 2
	If changed since wave 1: Current main activity
	Current status: All other current activities
	Whether doing traineeship, internship or training course
	If left school/college: Whether completed courses, reasons to leave if not & main reason
	If left work: Whether chose to leave, reasons to leave & main reason
	Verification of school year and name of current school (if no longer at same school as in Wave 1)
	If in education, whether studying full or part time

	If doing apprenticeship, traineeship, internship or training course, whether this is linked to education
	If doing training, whether course linked to job
D. Qualifications studying towards	Types of academic qualifications
	Number of academic qualifications
	Types of vocational qualifications
	Number and level of vocational qualifications
E. Early labour market experience	How many, if any, jobs YP is working
	Characteristics of main job (shift work, holiday job, etc.)
	Full/part time distinction for paid work
	If work/apprenticeship/ training course is main activity, hours per week spent in
	If apprenticeship/traineeship/internship/training course is main activity, whether this involved study at college or training provider or remote study from home and hours spent per week studying
	If main status is work, whether received training recently from an external provider,
	Hours supposed to be working if in part time work
	If in part-time work, whether would prefer full-time
	If main status is work, whether employee or self-employed
	If work/traineeship/internship/training course is main activity, how this was found
	If apprenticeship/training is main activity, whether get paid and type of payment
	If main activity is looking for a job, lowest weekly pay YP would consider and how many hours would work for it
	If YP not in education or training and not currently looking for work, whether looked for work in past 4 weeks
	If not in education or work, reasons that make it difficult to work
	If doing apprenticeship., level of apprenticeship working towards, views on current apprenticeship, reasons for doing an apprenticeship, plans after apprenticeship
F. Residual educational disruption due to COVID	Problems related to studies in Year 12 (for example class cancellations due to COVID-19, missing school due to COVID-19, lacking devices and study spaces, problems with motivation, etc.)
G. University applications	Perceived likelihood of applying to go to university
	Whether started or submitted an application for university
	Perceived likelihood of getting into university if they apply/applied
	Whether young person has an idea about courses/subjects to study at university, what it is, how they decided on it

	Whether young person has an idea about which university to study at, which one it is, reasons for choice of university
	Where YP intends to live if gets into preferred university and if at home, reasons for this
	Reasons for being unlikely to apply to university
H. Attitudes to education and future careers	Thoughts on the area YP lives in in terms of job opportunities
	Attitudes towards studying at university
	Attitudes towards training and qualifications
	If studying for vocational qualifications, highest level of vocational qualification eventually aimed for
	If not doing vocational qualifications, any future plans to study vocational qualifications
	Whether YP as an idea on jobs/careers they're interested in, and what they are
	Confidence around achieving future jobs YP wants
	Participation in activities about careers advice (careers advisors, careers fairs, university open days, etc.)
	Whether YP's schools/colleges offered help with issues around job/internship search, applications, scholarships etc.
	Informal careers advice (family members, teachers, friends, etc.) and whether this has influenced YP's decisions
	Sutton Trust boost sample questions: Awareness of educational access and support programs, and whether has applied to them
	Locus of control
	Whether YP is in favour of income redistribution
I. Spare time/leisure activities and homelessness	Participation in extra-curricular activities in the last 12 months outside of school: • Sports and exercise • Other clubs (arts, crafts, music, drama, etc.) • Classes associated with church/religion • Activities that involved overnight stays (such as Duke of Edinburgh expedition)
	Ever been homeless, if so, whether been homeless on their own or with family
	Whether currently homeless
	Whether ever lived in an institution, with foster parents, etc.
J. Health and wellbeing (CASI)	Whether had at least one COVID vaccine
	Ever had COVID-19
	Whether have had long COVID
	Whether long COVID reduced abilities to carry out day-to-day activities
	Experience of major life events in the last 12 months
	Whether YP finds COVID still has impacts on various aspects of their life, and whether these are positive or negative

	Gender
	Sexual orientation
	Mental health and wellbeing scales, please see section 3.5 for details (GHQ-12, GAD2, PHQ-2)
	Life satisfaction
	Self-assessed general health
	Chronic illnesses, in which areas these affect the YP, and whether they reduce their ability to carry our day-to-day activities
K. Friends, peers and family support (CASI)	Social Provisions scale (not asked in CATI)
	Cyber harassment (not asked in CATI)
	Discrimination (not asked in CATI)
	Evaluation of school provision of support on wellbeing and mental health
	Mental health seeking behaviour (sources reached out to and whether they received help) (not asked in CATI)
	Whether YP cares for someone who is ill, disabled or elderly and in need of care
L. Health Related Behaviours (CASI)	Number of times per week when young person exercised to break into a sweat (lasting at least 30 minutes, typical week over the last 4 weeks)
	If young person hurt themselves on purpose in anyway in the last 12 months (not asked in CATI)
	If young person ever hurt themselves on purpose to attempt to end their lives (not asked in CATI)
M. Closing demographics	Ethnicity
	Whether born in the UK, and country of birth if not UK
	Religion
	Whether has access to driving a vehicle owned in the household
N. Linkage	Linkage consent asked to link records from the below if not already consented in Wave 1: • Department for Education (DfE) • Department for Work and Pensions (DWP) • HM Revenue and Customs (HMRC) Linkage consents asked anew of everyone: • Universities and Colleges Admissions Service (UCAS) • Higher Education Statistics Agency (HESA) • Student Loans Company (SLC)
O. Recontact, signposts and closing screens	Updating of young person's contact details for future waves, intentions to move house and new address, signposting to sources of support If considering applying to university or already applied, consent to be contacted for further research (asked to YP who took part before the closing of the CAWI fieldwork in December 2022) Closing

3.4 Parent Questionnaire

The main focus of the parent/guardian questionnaire is to complement the information obtained from young people and to provide more context on household demographics. At Wave 2, questions included parents' attitudes to education, thoughts on young people's short-term plans, thoughts on young people's interest in apprenticeship and/or university, as well as parents' own experiences of school and learning in the past. There were also questions about parenting, and tuition.

In Wave 2, while some parents interviewed in Wave 2 were also surveyed at Wave 1, some were recruited in Wave 2 for the first time. The questionnaire was designed to accommodate both these groups, with 'new' parents being asked some of the key demographic questions which Wave 1 parents had already answered at the first wave. For example, some background questions like ethnicity or education were not asked again to Wave 1 respondent parents (unless their information was missing in Wave 1), but were only asked to new parents at Wave 2.

Similarly, the activity history section varied by type of parent respondent. New parents were asked about the period starting from the beginning of the pandemic in Wave 1 until the date of interview. For Wave 1 parents, this activity history was shorter, covering the period between the last recorded activity in Wave 1 and the date of interview.

On the documented questionnaires, the routing is clearly marked to help data users identify which questions were asked to Wave 2 parents only.

For parental occupation, similar to Wave 1, a *household reference person* approach was used so that this measure would be less dependent on the responding parent.

A summary of the parent questionnaire is provided below in Table 2. As mentioned in the previous section, the full questionnaires, annotated with variable names, are available within the data release on the UK Data Service and are also available on the COSMO study website.

Table 2. Parent questionnaire content at COSMO Wave 2

Section	Topic
A. Introduction, verification and opening demographics	If parent interviewed in Wave 1: Verification of being the same parent/guardian interviewed in Wave 1 (name, address, young person's name) If parent not interviewed in Wave 1: Verification of being a parent/guardian of named young person (young person's name, address)
	Relationship of parent/guardian to YP
	Gender of YP
	Date of birth of YP
	YP's school year

	Gender	
	Age	
	Number of household members	
	Whether YP lives at same address as their other parent	
	Relationship status (legal marital status and whether lives with someone as a couple)	
B. Attitudes to Education/ educational and career aspirations	Whether parent talks about school reports/progress reviews with YP	
	Current status: All current activities YP does, and main status	
	Whether parent talked to YP about plans for the near future, what these plans are, whether they are happy with it, and if not what they would prefer YP to do	
	Parents' evaluation of whether YP will do an apprenticeship, and reasons if not	
	Whether parent feels confident in supporting YP with a potential apprenticeship application	
	Parents' evaluation of whether YP will go to university, and reasons if not	
	Whether YP started applications for university	
	Whether parent feels confident in supporting YP with a potential university application	
	Attitudes towards statements on future life of YP (importance of having a job/career, raising a family, etc.)	
	Locus of control	
	Parents' own experiences of school and learning	
	Parents' views on things that influence young people's opportunities in life, and how much	
C. Parenting, home learning, tuition & catch-up	Parenting questions: Whether parents know where their child is going, whether they use online location services, whether they set a time for them to be back by, and how close they are to them, whether parent and YP get on each other's nerves, how often parent talks to YP about important things	
	Whether YP had to learn at home in the 2021/22 academic year, and on how many occasions	
	Whether YP had a tutor in the 2021/22 academic year, and its purposes, amount spent in private tuition	
	Evaluation of the effect of the educational disruption of the pandemic on YP	
D. Working status across the pandemic	If parent interviewed in Wave 1: main status history since Wave 1 (whether continuously the same main status since Wave 1, whether same job since Wave 1, each unique status and date they ended if anything else in between)	
	If parent not interviewed in Wave 1: Main status of parent before the pandemic, main location of work (home, office, etc.) before the pandemic	
	If parent not interviewed in Wave 1: Main location of work (home, office, etc.) before the pandemic, main status history covering from before the beginning of the pandemic until survey date (each unique status and date they ended)	

	A derived variable of parent's current main status and current work location if working
	If parent not interviewed in Wave 1: Whether parent was classified as a key or critical worker during the pandemic
	Main economic status of parent's partner
	Whether parent's partner is working full or part time
	If parent's partner if out of work for health reasons whether they have a long-term sickness or disability
	A derived variable of parent's partner's current main status
	Whether parent's partner was put on furlough since the beginning of the pandemic
E. Parental tenure, HRP and occupational details	Whether parent or household rents or has another arrangement
	Steps to determine household reference person: Whose name the property is owned or rented, whether parent or their partner has the highest income, whether parent or their partner is older. A derived variable on who the household reference person (HRP) is.
	For the HRP: Since when they have been in their current status, if not in work whether ever worked. Details of last job for those who had a job before: whether employee or self-employed. Main job title (open text). Open text descriptions of job, and what employer/business mainly does, for occupational coding. Whether the job required special qualifications and open text descriptions of them. Whether the job entailed managerial duties or supervision of other employees, whether more than 25 people are supervised, how many people work where HRP works as an employee, and how many employees HRP has if self-employed (not asked if same parent as W1 is HRP who has been holding the same job continuously since Wave 1)
	If home owners, time of purchase, and purchase value of home
F. Parental education	If parent not interviewed in Wave 1 or missing in Wave 1: Highest academic qualifications, and vocational qualifications
	If parent not interviewed in Wave 1 or missing in Wave 1: Other parent's highest academic qualifications, and highest vocational qualifications
G. Parental income	Sources of income for parent and parent's partner, receipt of universal credit and other benefits, receipt of additional universal credit due to circumstances (having children, a disability or health condition, etc.)
	Banded income over a year, month or week (22 bands)
H. Grandparents	Whether any grandparents of YP got a university degree
	Whether any grandparents of YP born outside of UK, and country of birth
I. COVID History and vaccination (CASI)	Vaccination status (whether vaccinated, number of doses received)
	Reasons if not vaccinated
	Whether all eligible persons in HH vaccinated
	Ever had COVID-19
	Whether have had long COVID
	Whether long COVID reduced abilities to carry out day-to-day activities

	Parent or anyone in the HH asked to shield by health professionals
J. Parent health and wellbeing (CASI)	Whether the household experienced major life events in the last 12 months (loss of a job, death of someone close, moving, etc.)
	Mental health and wellbeing scales, please see section 3.5 for details (GHQ-12, GAD-2, PHQ-2)
	Life satisfaction
	Self-assessed general health
K. Disadvantage (CASI)	Comparison of current financial situation to this time last year
	Whether fallen behind on rent or mortgage in the last 12 months
	Whether fallen behind on energy bill payments in the last 12 months
	Self-assessment of financial situation
	Issues with housing (mould, heating issues, etc.)
	Number of bedrooms
	Food poverty in the last 12 months and who was affected by it
	Use of a food bank in the last 12 months, and frequency of use
	Whether any measures taken to save money in the last 12 months (cut down on electricity use, driving, etc.)
	If parent not interviewed in Wave 1 or missing in Wave 1: YP's eligibility to FSM between Year 7 and Year 11
L. Closing demographics	If parent not interviewed in Wave 1 or missing in Wave 1: Ethnicity
	If parent not interviewed in Wave 1 or missing in Wave 1: Whether parent is born in the UK and which country
	Ethnicity for the other parent of YP
	Whether other parent of YP is UK born, and country of birth
	If parent not interviewed in Wave 1 or missing in Wave 1: Religion
M. Contact details, signposting and closing screens	Name and contact information for new parents and updating for Wave 1 parents, whether parent lives in the same address as the YP and updating of either if necessary for future waves, intentions to move house and new address, signposting to sources of support and closing

3.5 Scales

The COSMO Wave 1 questionnaires included several established scales which are listed below.

3.5.1 GHQ-12 (12 items) (Young Person questionnaire and Parent questionnaire)

Goldberg D, Williams P. A user's guide to the general health questionnaire. London: Nfer-Nelson; 1988.

The General Health Questionnaire (GHQ) is used as a screening tool of probable mental ill health. The 12 item screening instrument measures general, non-psychotic and minor psychiatric disorders; and concentrates on the broader components of psychological ill health and characteristics as general levels of happiness, depression and self-confidence. Each of the 12 GHQ items, six positively and six negatively phrased, are rated on a four-point scale to indicate whether symptoms of mental ill health are present.

Variable name	Question
W2_ZGHQ1	Have you recently been able to concentrate on what you're doing?
W2_ZGHQ2	Have you recently lost much sleep over worry?
W2_ZGHQ3	Have you recently felt that you are playing a useful part in things?
W2_ZGHQ4	Have you recently felt capable of making decisions about things?
W2_ZGHQ5	Have you recently felt constantly under strain?
W2_ZGHQ6	Have you recently felt you couldn't overcome your difficulties?
W2_ZGHQ7	Have you recently been able to enjoy your normal day to day activities?
W2_ZGHQ8	Have you recently been able to face up to your problems?
W2_ZGHQ9	Have you recently been feeling unhappy or depressed?
W2_ZGHQ10	Have you recently been losing confidence in yourself?
W2_ZGHQ11	Have you recently been thinking of yourself as a worthless person?
W2_ZGHQ12	Have you recently been feeling reasonably happy, all things considered?

The cohort member's score on the General Health Questionnaire 12 point scale (GHQ12) is derived by summing responses to the twelve GHQ12 questions (W2_ZGHQ121 to W2_ZGHQ1212). This is scored according to the 0-0-1-1 method, in which the first two possible responses to each question are assigned a value of 0 and the third and fourth responses with a value of 1, resulting in a maximum possible score of 12 for this variable. A higher score on this scale indicates a greater likelihood of mental ill health. This scale was included in COSMO Wave 1 as well.

3.5.2 GAD-2 (Generalised Anxiety Disorder 2-item) (Young Person questionnaire)

Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med. 2007;146:317-25.

The GAD-2 was based on the GAD-7, which was developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues, with an educational grant from Pfizer Inc. No permission required to reproduce, translate, display or distribute.

GAD-2 was used in COSMO Wave 1, and also recently used in the COVID-19 Surveys conducted by CLS on the Millennium Cohort Study, Next Steps, 1970 British Cohort Study, 1958 National Child Development Study, and MRC National Survey of Health and Development.

The Generalized Anxiety Disorder 2-item (GAD-2) is a brief initial screening tool for generalized anxiety disorder.

Respondents are asked whether they have been bothered by problems over the last 2 weeks, with the following response options:

1. Not at all
2. Several days
3. More than half the days
4. Nearly every day

The GAD-2 score is obtained by adding the score for each question (Total points). The score for each question is:

0 = Not at all

1 = Several days

2 = More than half the days

3 = Nearly every day

Variable name	Question
W2_ZGAD2PHQ2_01	Feeling nervous, anxious or on edge
W2_ZGAD2PHQ2_02	Not being able to stop or control worrying

3.5.3 PHQ-2 (Patient Health Questionnaire 2-item) (Young Person questionnaire)

Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener. Medical Care. 2003;41:1284-92.

The PHQ-2 enquires about the frequency of depressed mood and anhedonia over the past two weeks. The PHQ-2 includes the first two items of the PHQ-9. PHQ-2 was recently used in the COVID-19 Surveys conducted by CLS on the Millennium Cohort Study, Next Steps, 1970 British Cohort Study,1958 National Child Development Study, and MRC National Survey of Health and Development. It was also included in COSMO Wave 1.

Respondents are asked whether they have been bothered by problems over the last 2 weeks, with the following response options:

1. Not at all

2. Several days

3. More than half the days

4. Nearly every day

The PHQ-2 score is obtained by adding the score for each question (Total points). The score for each question is:

0 = Not at all

1 = Several days

2 = More than half the days

3 = Nearly every day

Variable name	Question
W2_ZGAD2PHQ2_O3	Little interest or pleasure in doing things
W2_ZGAD2PHQ2_O4	Feeling down, depressed or hopeless

3.5.4 Locus of control (Young Person questionnaire and Parent questionnaire)

Young people were asked how much they agree or disagree with five items to derive a variable to indicate the extent to which they believe that they have control over events in their lives, from the following responses:

1. Strongly agree

2. Agree

3. Disagree

4. Strongly disagree

Variable Name	Questions
W2_ZSCHOOLATT2_1	If someone is not a success in life, it is usually their own fault
W2_ZSCHOOLATT2_2	People like me don't have much of a chance in life
W2_ZSCHOOLATT2_3	I can pretty much decide what will happen in my life
W2_ZSCHOOLATT2_4	How well you get on in this world is mostly a matter of luck
W2_ZSCHOOLATT2_5	If you work hard at something you'll usually succeed

The cohort members' total score on the locus of control scale was derived by summing the responses to the locus of control questions to generate a total score ranging from 5 to 20. A low value of 5 to 9 indicates an internal locus of control, a score ranging between 10 and 14 indicates either a moderate internal or moderate external locus of control, and a score between 15 and 20 suggests external locus of control.

These items have previously been asked in Next Steps Age 25 survey, as well as Next Steps Waves 7, 4 and 2 (then LSYPE1) and also COSMO Wave 1.

In Wave 2, parents were also asked how much they agree or disagree with a similar set of seven items to derive a variable to indicate the extent to which they believe that they have control over events in their lives, with the same set of responses as above. The items they were asked about were:

Variable Name	Questions
W2_XYPJOBATT2_01	If a person works hard at something they will usually succeed
W2_XYPJOBATT2_02	How well you get on in this world is mostly a matter of luck
W2_XYPJOBATT2_03	Some children just do better at school than others
W2_XYPJOBATT2_04	It is important to me that [XCHILDNAME] does as well as or better than me in life
W2_XYPJOBATT2_05	It's more important to go out and get a job than to take time gaining lots of qualifications
W2_XYPJOBATT2_06	How well a child does in their education will affect how well they do in life
W2_XYPJOBATT2_07	I don't think much about what my child might be doing in a few years' time

3.6 Data Linkage

As mentioned in section 3.3, young people were asked for their consent to link administrative data to their survey data, held by a variety of organisations.

Young people who consented to a particular administrative data linkage (listed below) in Wave 1 were not asked again in Wave 2 about these linkages. Those who did not consent in Wave 1 were asked again at Wave 2 if they would consent to link administrative data to their survey data from the below sources:

- Education records, held by the DfE, including the National Pupil Database (NPD) and Individualised Learner Records (ILR) – covering achievement in school and further education as well as details about the school, college or training centre young people attended;

- Information on benefit and employment programs, kept by Department for Work and Pensions (DWP);

- Information on employment, earnings, tax credits, occupational pensions and National Insurance Contributions, kept by HM Customs and Revenue (HMRC).

    Taken together, consent to the linkage to NPD, ILR, DWP and HMRC records allows for linkage to the UK Government's combined Longitudinal Educational Outcomes (LEO) dataset, which is based on a combination of these administrative datasets.

In addition to the above, there were three new linkage consents asked in Wave 2 which were not included in Wave 1, therefore were asked to all young people:

- Information about higher education applications and offers held by the Universities and Colleges Admissions Service (UCAS),

- Information about participation and achievement in school, further and higher education as well as details about the school, college, university or training centre young people attended, via records kept by the Higher Education Statistics Agency (HESA),

- Information about payments of student support and loan repayments, held by the Student Loans Company (SLC)

These new consents are envisioned to enrich future analysis of the COSMO data sets with higher education data.

The procedures for explaining and obtaining these consents from young people were approved through the procedures set out by the UCL IOE Research Ethics Committee.

3.7.1 Data linkage consent process

When young people were invited to participate in COSMO, they were sent a leaflet, which explained that they would be asked for data linkage consent and it was emphasised that this was entirely their choice. Moreover, in the respondent facing website, there was a separate page on data linkage, where young people could access some frequently asked questions on data linkage. These made clear how the linkage process worked, which data holders they would be asked about and the purpose of data linkage. The webpage also emphasised that they may choose to consent to some and not other linkages, that they can complete the survey without consenting to any of them, and young people were also informed about issues like data retention and withdrawing their data.

As the young people were over the age of 16 at the time of the interview, there was no parental consent necessary for data linkage. However, on the website, it was emphasised that young people could want to discuss these with their parents if they wish to do so, and parents also received a copy of the survey leaflet which outlined this process.

Within the survey, at the beginning of the consent module, young people were informed of the steps of data linkage, that information on them will be collected on an ongoing basis unless they told the study team to stop, and that they could change their permissions at any time.

As mentioned in the previous section, young people who consented to linkages asked in Wave 1 (for NPD, ILR, DWP and HMRC) were not asked for consent again. These repeat questions were only asked to young people who did not consent in Wave 1. This is in line with the information we provided young people in Wave 1, which informed them that we would add information from their records unless they asked us to stop.

The proportion of young people who consented to linkage are presented in section 4.6.

4 Fieldwork

Wave 2 fieldwork involved recontacting all young people who took part at Wave 1, and all parents who took part at Wave 1 if they were in the same household as a responding young person at Wave 1.[6] In addition, for households where only a young person interview had been achieved at Wave 1, we attempted to obtain a Wave 2 interview with the parent for the first time. This allowed us to supplement the young person's data with important parental socio-demographic information where this had previously been missing.

Wave 2 fieldwork with young people and parents was conducted between 18 October 2022 and 16 April 2022, at which stage the cohort of young people was aged 17–18 and in Year 13.

The time interval between Wave 1 and Wave 2 varied from 6 months to 19 months depending on fieldwork dates at each wave.

All fieldwork at Wave 2 was conducted by Kantar Public.

4.1 Summary of Wave 1 fieldwork

4.1.1. Overview of fieldwork at Wave 1

A summary is provided here, with more extensive details in the [Wave 1 user guide](#) and [Wave 1 technical report.](#)

Wave 1 fieldwork ran between September 2021 and April 2022, at which stage the cohort of young people was aged 16–17 and in Year 12. All fieldwork was conducted by Kantar Public, with support from NatCen during the face-to-face stage of the study.

Throughout fieldwork, efforts were made to maximise the number of households where both the young person and a parent participated, as this provided a more complete picture of household characteristics. Within each household, only one parent was asked to complete the questionnaire, and any resident parent could choose to do this.

Wave 1 used a sequential mixed-mode design which comprised an initial online data collection phase and a subsequent face-to-face stage. The online phase comprised a launch mailing followed by up to 4 reminders. As the only contact information available

---

[6] Parents who responded at Wave 1 without a matching young person interview from the same household were not weighted for analysis at Wave 1 and were removed from the longitudinal panel going forwards. Only parents with a matching young person household interview were included at Wave 2.

about issued sample members at Wave 1 was a postal address, all communication was conducted by post.

The main objectives of the subsequent face-to-face stage were to improve response rates and sample balance, with a subset of non-responders issued to the face-to-face stage based on addresses which were least well represented after the online phase, using a regression non-response model to identify households with the lowest response probabilities across the whole sample. The face-to-face stage was also used to help increase the rate of complete households (i.e., to achieve an interview with the young person or parent where only one of them had completed online to date).

However, industry-wide COVID-19 related challenges with interviewer capacity, coupled with respondent reluctance for in-home interviews during this time, meant that the yield from the face-to-face stage was very low. As a result, a decision was made to issue additional fresh sample from a reserve sample of addresses which had been selected at the outset alongside the main sample of addresses. For the reserve sample, data were collected via online methods only between March 2022 and April 2022.

Throughout fieldwork young people and their parents were offered a voucher conditional on survey completion to the value of either £10 or £20. Higher incentivisation was targeted at students and their parents expected to be from more disadvantaged backgrounds, based on information from the sampling frame.

The main fieldwork was based on students sampled from the National Pupil Database (NPD) which covers students in state schools within England. However, alongside this a small supplementary sample of students was sampled from a randomly selected sample of 33 independent schools which sent invites to students in the relevant cohort on Kantar Public's behalf.

A 'Keeping in touch' (KIT) exercise was implemented between Wave 1 field closure and Wave 2 field start (during July/August 2022) for the original sample issue of young people to help collect some missing data and to provide updated contact information.

4.1.2. Achieved sample sizes and response rates at Wave 1

The final achieved Wave 1 sample included 10,050 cases which comprised data from a matching young person and one of their parents, and a further 3,735 cases where the data only included a young person with no matching parent interview. This provided a total sample of 13,786 young people and 10,050 parents.[7]

The total Wave 1 sample of 13,786 young people was made up of 13,112 young people sampled from the NPD (state school) sample and 674 from the independent school sample.

---

[7] The original figures were 13,787 young people interviews, and 10,051 parents. However, a young person and a parent interview were later excluded to respect their wishes in a data deletion request.

The total Wave 1 sample of 10,050 parents with a matching young people interview was made up of 9,845 parents of young people sampled from the NPD (state school) sample and 206 parents of young people from the independent school sample.

At Wave 1, the large majority of data was collected online; only 3.5% of all individual interviews as part of the NPD sample were completed face-to-face. The core reason for the low face-to-face yield was industry-wide COVID-19 related challenges as explained further in section 4.2.2.

A summary of the overall response rates achieved at Wave 1 is provided in Table 3 below. This documents the response rates, and design weighted response rates, for the main sample and boost sample combined. The independent school sample was sampled and fielded in a different way at Wave 1 and therefore these are not included in the figures below.

The design weighted response rates are included as we used a disproportionate sampling design at Wave 1 (see section 2) and the unweighted response rates are affected by this sample design.

The overall response rates at Wave 1 were 37% for young people, 32% for parents, and complete household data was obtained at 28% of addresses. The parent-only response rates include 1,680 parents were interviewed at Wave 1 where there was no matching young person interview. Although all surveyed cases have been included in the Wave 1 dataset, these cases have been given a zero-weight value in the dataset and are not included as part of the Wave 2 issued sample.

For more details about response rates at Wave 1, and for details of the supplementary independent school sample, please refer to the Wave 1 User Guide.

Table 3. Summary of response rates for the NPD and boost sample combined at Wave 1 (excluding independent school sample)

	Issued sample	Achieved sample	Response rate	Response rate – (design weighted)
Total NPD sample (main and boost)				
Young people	35,719	13,112	36.7%	36.9%
Parents	35,719	11,368	31.8%	32.3%
Complete household	35,719	9,845	27.6%	27.8%

4.2 Fieldwork strategy at Wave 2

The issued sample for Wave 2 was 13,786 households. This included:

- 10,050 households where we were looking to achieve repeat interviews with both a young person and parent who also took part in Wave 1;
- 3,735 households where only the young person took at Wave 1; in these households we were seeking a repeat interview with the young person and a first-time interview with a parent.

The Wave 2 fieldwork strategy followed a sequential mixed-mode design involving an initial online data collection stage, followed by interviewer-led data collection for non-responders, and a final mop-up online stage.

Whereas at Wave 1 the interviewer-led stage only involved a face-to-face data collection stage, at Wave 2 it involved a face-to-face and telephone data collection stage.

The incentive strategy varied across each fieldwork phase and details are provided in section 4.3 below.

4.2.1. Initial online data collection stage

The initial online data collection stage ran between 18 October 2022 and 15 December 2022. During this phase, respondents were contacted as follows:

- Prenotification mailing which included an A5 postcard with some engaging findings from Wave 1
- Launch invite (by letter, email and SMS)
- Three further reminders (by letter, email and SMS)

All COSMO issued sample members for Wave 2 were contacted by letter, with supporting invitations also sent by email and SMS where relevant contact email and mobile phone information had been provided by participants at Wave 1 or as part of the follow-up keeping-in touch exercise.

Where both members of the household were invited to take part, the two postal mailings were sent in separate envelopes. Emails and SMS messages were also directed to the individual.

The reminder strategy varied by mailing, with some reminder mailings also targeting specific groups of non-responders alongside the core group of non-responders. This included sample members who had broken off before completion and partial households where we wanted to encourage full household completion at Wave 2.

Messaging for each mailing was targeted based on the type of respondent:

- Young person, parent who took part at Wave 1, or new parent not taken part at Wave 1;
- (At reminders 1 and 2) Whether a standard non-respondent or a respondent who had started online but broken off without completing the survey;
- (At reminder 3) Whether we were seeking interviews at the address with both household members or just one.

A further mop-up stage of online data collection was conducted shortly before fieldwork closed – see section 4.2.4 below.

4.2.2. Face-to-face data collection stage

At the close of the initial online phase, young people and parents in Wave 2 issued households who had not already responded online were issued to face-to-face fieldwork. This included any 'break-off' cases where the participant had started the survey online but did not reach the threshold to count as a complete interview. However, cohort members who had requested to opt out of Wave 2 during the online phase were not issued to the face-to-face stage.

Before starting work in the field, all interviewers were briefed by a member of the COSMO research team at Kantar Public. Briefings were held via live online video sessions conducted on Microsoft Teams and included input from UCL on the aims and objectives of the research. In total, over the course of fieldwork, 188 interviewers were briefed across 15 briefing sessions. Briefings were supported by a set of written interviewer instructions.

Face-to-face fieldwork ran from 22 December 2022 to 10 April 2023.

A reference to the possibility of an interviewer visit was included in the final reminder which was mailed as part of the initial online phase (see section 4.2.1 above). However, interviewers also had copies of an advance letter which provided more specific details of the face-to-face survey visit, which interviewers were able to send out themselves, so they could time these letters to land shortly before they intended to start making household visits.

Although all young people were aged over 16 and so there was no legal requirement for this, interviewers were instructed to seek parental permission before interviewing a young person as a courtesy. Interviewers were able to make initial contact via face-to-face visits, as well as by telephone or SMS where relevant contact information had been provided at Wave 1.

If, upon contact at an issued address, an interviewer found that the household had moved, interviewers were instructed to try to obtain updated contact information from the new occupiers where possible.

Not all addresses initially allocated to face-to-face were covered in the event. In some cases, the proposed interviewer assignment was too small to generate a viable interviewer assignment. In addition, there were industry-wide issues related to interviewer capacity and productivity which meant that not all sample could be covered in the timeframe available (we had a fixed deadline of mid-April as we needed to ensure that fieldwork would be finished before the start of exam season). All eligible sample which was not covered by face-to-face was issued instead as part of the telephone and/or final mop-up online data collection stage.

4.2.3. Telephone data collection stage

The telephone stage of fieldwork ran from 17 March to 6 April 2023. For the telephone survey, all sample was issued as individual-based rather than household-based sample.

The sample for the telephone survey included sample from two sources:

- Individuals originally allocated to face-to-face but who were not in the event visited by a face-to-face interviewer (see section 4.2.2 above for more details on this).

- Individuals which were selected from cases which had been worked face-to-face but which had resulted in an unproductive outcome (for example non-contacts and proxy refusals). Cases were only transferred from face-to-face to telephone if they had a valid UK telephone number (landline or mobile).

Online 'breakoff' interviews which had not reached the threshold to count as a complete interview were also included within the eligible subsample for telephone fieldwork. However, anyone who had previously opted out of Wave 2 during the online stage, or who had directly refused to be interviewed face-to-face, was not issued to the telephone stage.

Where a mobile number was available, household members allocated to the telephone stage were sent an advance SMS to let them know that a telephone interviewer would be making contact.

All telephone interviewers were briefed in advance by a member of the COSMO research team at Kantar Public, via a live video briefing conducted on Microsoft Teams. In total 10 telephone interviewers were briefed. Briefings were supported by a set of written interviewer instructions.

As noted in section 3, the telephone interview was a slightly reduced version of the online and face-to-face survey which excluded the more sensitive questions which could be fielded more privately in the other two modes.

4.2.4. 'Mop-up' online data collection stage

A final mop-up online phase was carried out in the final weeks of fieldwork from 3 April[8] to 16 April 2023, comprising one further reminder mailing only. This was also timed to coincide with school Easter holidays when we expected many young people to be at home preparing for A-levels or other exams. As noted above, the end fieldwork date was fixed to avoid any overlap with exam season.

The eligible sample for this phase was all sample members who had not taken part in Wave 2 to date but excluding anyone who had opted out or who had refused directly as part of either the face-to-face or telephone phase.

At the mop-up online stage, where two letters were being sent to the household, these were paired in the same envelope to help push up the rate of whole household completions at this final stage. Follow-up emails and SMS invites were sent to individuals where relevant contact information was available.

4.3 Incentives

Young people and their parents were offered a voucher conditional on survey completion, with unconditional vouchers also introduced for some sample members at the final mop-up online stage. The incentive strategy was adapted throughout fieldwork on an agile basis in response to patterns of fieldwork progress and to help boost survey groups of highest priority. The value of the voucher offered was therefore dependent on a) school-level information and b) the stage of fieldwork.

Depending on the above factors, this meant that individual respondents received either £10, £15, £20, £25 or £30.

4.3.1. Incentivisation by school-level information

At Wave 1, respondents were offered a voucher on completion of the survey to the value of either £10 or £20. The value varied depending on the nature of the school the young person attended, with those attending a school with the highest rates of pupils eligible for free school meals receiving the higher amount.[9] Parents received an incentive to the same value as the young person in their household. Differential incentivisation was used

---

[8] There was a small overlap between start of the online mop-up and close of face-to-face fieldwork (10 April) and telephone fieldwork (6 April). During this overlap period, a small number of participants completed the survey on more than one mode. In these cases, we removed the duplicate record. Where there was an overlap with telephone, we prioritised the online or face-to-face interview (given that the telephone interview was a truncated version of the full questionnaire). Where there was an overlap between online and face-to-face, we accepted the earlier interview in the sequence.
[9] Based on school attended when sampled for Wave 1

to help ensure a good representation of students and their parents from more disadvantaged backgrounds, who are typically less likely to respond to surveys.

At Wave 2, the same incentivisation allocation was used, so the same households were either offered a standard or higher incentive. However, the actual value of this standard and higher amount varied dependent on fieldwork stage (see section 4.3.2 below).

4.3.2. Incentivisation by fieldwork stage

The incentive strategy also varied depending on the stage of fieldwork.

Initial online phase

During this phase, all respondents were offered an 'early bird' bonus to encourage as many respondents as possible to complete the survey online. Respondents were told that they needed to complete the survey by 11 December 2023[10] in order to qualify for the higher incentive. This meant that respondents who completed online by this date and who were allocated to the standard incentive sample group were offered £20 instead of £10, while early completers allocated to the higher incentive sample group were offered £30 instead of £20.

Face-to-face and telephone stage

Once the initial online survey had closed, respondents were offered a voucher without the 'early bird' bonus: £10 (standard) or £20 (higher).

Mop-up online phase

At this final stage, we were keen to top up the sample with as many 'new' parents (those who had not already taken part at Wave 1) as possible. This was to help ensure we had parental socio-demographic information for as many households as possible.

Therefore, at this stage, we provided all new parents who had not yet taken part at Wave 2 with an **unconditional £5 incentive** as well as the conditional incentive (either £10 or £20) which they could claim after completing the survey. Any young person in the same household who was also a non-responder was offered the same amount to provide parity within the household. This meant that young people and parents in these households completing at the mop-up stage were eligible to receive a total of either £15 or £25 depending on their sample allocation.

All Wave 1 parents, and young people in a household with either a Wave 1 parent or a new parent who had already completed, were not provided with the extra unconditional gift card. Non-responders in this group were either offered £10 or £20, i.e., the original value of their incentive without either an early bird or 'new parent' bonus.

---

[10] In practice the early bird bonus continued until 15 December 2022 after which the online survey was closed, to be re-opened again later for the mop-up stage.

Unconditional vouchers were sent with postal invite letters and were provided in the form of a physical 'Love2shop' giftcard.

4.4 Keeping in touch (KIT) exercises

4.4.1 Keeping in touch between Wave 1 and Wave 2

Two 'Keeping in touch' (KIT) exercises were implemented between the period from the end of Wave 1 fieldwork (April 2022) to before the start of Wave 2 (October 2022). Full details for these are covered in detailed in the [Wave 1 User Guide](#), and a summary is provided below.

The first was a KIT exercise sent between waves to collect updated contact details as well as some missing Wave 1 data that affected a portion of the sample. This exercise was conducted in the form of a short web survey fielded between July and August 2022. A £5 voucher was provided upon completion to help boost response rates. This exercise targeted the achieved NPD and independent school sample of size 10,858 and yielded a 48.2% response rate.

The second was a prenotification mailing sent in October 2022 to all young people and parents who were in the Wave 2 issued sample. This comprised a cover letter and A5 postcard with some engaging findings from Wave 1. The aim of this mailing was to act as an engagement tool and to remind families of the study in advance of the main Wave 2 launch which followed a week or so later.

4.4.2 Keeping in touch between Wave 2 and Wave 3

A keeping in touch exercise is planned for late 2023 targeting young people, the aim being to update contact information, and collect some brief updates on young people's current activities, including information on whether they are now in higher education, and their living arrangements. At the time of writing, this exercise is currently under development and will be deposited upon data collection and processing.

4.5 Key fieldwork dates

A summary of key fieldwork dates leading up to and during Wave 2 is provided below:

Table 4. Key fieldwork dates

Fieldwork phase	Sample subgroup	Dates
Wave 1 fieldwork		
Fieldwork – all modes	All W1 issued sample	22 September – 18 April 2022
Keeping in Touch (KIT) exercise between Wave 1 and Wave 2		
KIT mailing	All young people in original issue W1 sample	4 June – 9 September 2022
Wave 2 initial online data collection stage		
Pre-notification mailing	All W2 issued sample	18 October 2022
Web survey launch letter (follow-up emails and SMS)	All W2 issued sample	26 October 2022 (29 October 2022)
Web survey reminder 1 letter (follow-up emails and SMS)	All remaining non-responders	4 November 2022 (10 November 2022)
Web survey reminder 2 letter (follow-up emails and SMS)	All remaining non-responders	18 November (23 November 2022)
Web survey reminder 3 letter (follow-up emails and SMS)	All remaining non-responders	29 November 2022 (7 December 2022)
Initial web survey close and end of 'early bird' voucher eligibility		15 December 2022
Face-to-face data collection stage		
In-home face-to-face fieldwork	All individuals who had not taken part online	22 December 2022 – 10 April 2023
Telephone data collection stage		
Telephone fieldwork	Individuals at addresses which had not been worked face-to-face, together with a selection of individuals with an unproductive outcome code from the face-to-face stage	17 March 2023 – 6 April 2023
Mop-up online data collection stage		

Fieldwork phase	Sample subgroup	Dates
Mop-up fieldwork final reminder letter + follow-up email and SMS	All remaining non-responders at all earlier modes	3 April to 16 April 2023
All samples		
Wave 2 fieldwork close	All samples	16 April 2022
Keeping in touch (KIT) exercise between Wave 2 and Wave 3		
Planned for December 2023		

4.6. Survey response at Wave 2

4.6.1 Wave 2 achieved sample sizes – main and boost

At Wave 2, the supplementary sample of independent school students was fielded in the same way as the main sample of students from the NPD at Wave 1. Therefore the Wave 1 independent school sample is now included as part of the main (that is non-boost) sample.

The deposited dataset for Wave 2 included the following numbers of young people, parents and households.

Table 5. Achieved sample sizes at Wave 2 (main and boost)

Sample group	Achieved sample (n)
Young people	11,523
Parents, of which:	10,678
– Wave 1 parent	8,760
– New parent	1,918
Complete households, of which:	10,204
– Household was also a complete household at Wave 1	8,399
– Additional complete household including a new parent	1,805

The achieved totals are provided below for the main and boost sample combined (Tables 6 to 7), and further split by survey mode.

The original achieved field numbers were a little higher than this, but some cases were removed as part of quality assurance checks (see section 5.2 for details).

Achieved samples by mode: main and boost sample

Table 6 displays the breakdown for individuals in the main sample and boost sample combined, by survey mode.

The large majority of interviews (93%) were completed online, while 7% were completed face-to-face and 0.7% by telephone. Face-to-face interviews were particularly productive for data collection with new parents: 8% of all parent interviews and 25% of new parent interviews were collected face-to-face, compared with only 5% of young people.

Table 6. Achieved sample sizes at Wave 2 by survey mode: individuals in the main sample (including those sampled as part of the independent school sample at Wave 1) and boost sample combined

	Online	F2F	Tel	Total individuals in main and boost sample combined
Young people	10,843	574	106	11,523
All parents, of which:	9,748	884	46	10,678
– W1 parent	8,317	397	46	8,760
– New parent	1,431	487	0	1,918

Table 7 displays the breakdown for complete households. In the table below, an 'online (both)' complete household is a household where **both** the young person **and** parent completed online, and similarly for face-to-face and telephone. Where a household was completed across a mixture of modes this is included in the in the 'mix of modes' column.

Table 7. Achieved sample sizes at Wave 2 by survey mode: households in the main sample (including those sampled as part of the independent school sample at Wave 1) and boost sample combined

	Online (both)	F2F (both)	Tel (both)	Mix of modes	Total households in the main and boost sample
Complete households, of which:	9,196	312	4	692	10,204
– Household was also a complete household at Wave 1	7,865	165	4	365	8,399
– Additional complete household including a new parent	1,331	147	0	327	1,805

4.6.2 Response rates at Wave 2

The issued sample for Wave 2 was all households where a young person had responded at Wave 1 (n=13,786 households).

The issued sample for young people was also therefore n=13,786.

For parents there were two sample issues:

- Households where the parent had also taken part at Wave 1 (n=10,050 issued)
- Households where no parent had taken part at Wave 1 – new parents (n=3,735 issued)

For young people, response was defined as having completed up to the end of the survey before the self-completion module (question ZWHLI in Section I of the questionnaire). For parents, response was defined as having completed the first question in the self-completion module (question XBEENVAC in Section I of the questionnaire).[11] More information on questionnaire content is provided in section 3.

Response rates have been calculated as an overall response rate, based on all modes of data collection.

---

[11] For parents, the questions at the end of the preceding self-completion section were not 'ask all' so the threshold was set to be the first 'ask all' question after the end of the preceding section.

Response rates for the main and boost sample combined

The response rates for the main and boost samples combined are documented in the table below.

Table 8. Response rates at Wave 2: all sample (main – including those sampled as part of the independent school sample at Wave 1 – and boost)

Issued sample description	Issued sample (n)	Achieved individuals at Wave 2 (n)	Response rate
Individuals			
Young people who took part at Wave 1	13,786	11,523	83.6%
Parents who took part at Wave 1	10,050	8,760	87.2%
New parents (did not take part at Wave 1)	3,735	1,918	51.4%
All parents (Wave 1 and new)	13,786	10,678	77.5%
Issued sample description	**Issued sample (n)**	**Achieved complete households at Wave 2 (n)**	**Response rate**
Households			
Complete household at Wave 1[12]	10,050	8,399	83.6%
New parent household at Wave 1[13]	3,735	1,805	48.3%
All households[14]	13,786	10,204	74.0%

---

[12] Response rates in this row are based on all households which were complete at Wave 1 and remained complete at Wave 2. The achieved totals in this row do not include complete households achieved via new parents interviewed for the first time at Wave 2.

[13] Response rates in this row are based on all households which lacked a parent interview at Wave 1 but which included both a young person and a new parent at Wave 2.

[14] Response rates in this row are based on all households issued, whether or not the household was complete at Wave 1. The achieved totals in this row are based on all complete households including either a Wave 1 or new parent responding in Wave 2.

4.7 Data linkage consent rates

As discussed in section 3.6, young people were asked for their consent to link various administrative data to their survey data at Wave 1. This included separate requests to link data from the Department for Education (DfE), Department for Work and Pensions (DWP), and HM Revenue and Customs (HMRC) to their survey data. As mentioned in section 3.6, consent to link records from DfE (including NPD and ILR), DWP and HMRC at the same time allows for linkage to the UK Government's combined Longitudinal Educational Outcomes (LEO) dataset. Therefore, in this section, consent to link information from all these three organizations are also covered to give an indication of potential LEO data linkage (Table 9).

Where consent to the above had not already been obtained at Wave 1, consent was re-asked at Wave 2. In addition, at Wave 2 three additional linkage requests were collected for the first time: the Universities and Colleges Admissions Service (UCAS), the Higher Education Statistics Agency (HESA), and the Student Loans Company (SLC).

Consent rates and numbers of young people in the main and boost sample combined, who consented to linkage to these six requests are provided in the tables below.

Based on the Wave 2 sample, for the three consents which built on Wave 1, additional consents provided at Wave 2 means that we now have consent from between 80% and 87% for each of these.

For any analysis based on the Wave 1 sample we can also now build in the additional consents from Wave 2 which means that we now have consent from 84% for DfE (compared with 74% at Wave 1), 78% for DWP (compared with 66% at Wave 1) and 76% for HMRC (compared with 65% at Wave 1).

For the consents collected at Wave 2 only, consent rates are lower. We might also expect lower consent rates for linkage to administrative datasets related to universities, as survey participants not planning to go onto higher education may not consider this relevant to their circumstances.

Table 9. Linkage consent rates for DfE, DWP and HMC based on i) all responding at Wave 2 and ii) all responding at Wave 1 (main and boost sample combined)

	Department for Education (DfE)	Department for Work and Pensions (DWP)	HM Revenue and Customs (HMRC)	Consented to all three of these
i) Consent rates based on Wave 2 sample				
Total no. young people in Wave 2 survey dataset	11,523	11,523	11,523	11,523
Total no. at wave 2 who had already consented to linkage at Wave 1	8,617	7,809	7,629	7,024
Additional YP who consented at Wave 2	1,438	1,675	1,555	742
Total no. within Wave 2 sample who consented at either wave	10,055	9,484	9,184	7,766
Overall consent rate among Wave 2 sample (from W1 +W2) based on all Wave 2 sample	87.3%	82.3%	79.7%	67.4%
ii) Consent rates based on Wave 1 sample				
Total no. young people in Wave 1 survey dataset	13,786	13,786	13,786	13,786
Total no. who consented at Wave 1 (including YP who did not take part at Wave 2)	10,137	9,137	8,946	8,221
Additional YP who consented at Wave 2	1,438	1,675	1,555	742
Total no. within Wave 1 sample who consented at either wave	11,576	10,813	10,502	8,963
Overall consent rate among Wave 2 sample (from W1 +W2) based on all Wave 1 sample	84.0%	78.4%	76.2%	65.0%

Table 10. Linkage consent rates for UCAS, HESA and SLC – main and boost sample combined

	Total no. young people in wave 2 survey dataset	Total no. who consented at Wave 2	Consent rate
Consent requested at Wave 2 only			
Universities and Colleges Admissions Service (UCAS)	11,523	8,009	69.5%
Higher Education Statistics Agency (HESA)	11,523	7,931	68.8%
Student Loans Company (SLC)	11,523	7,518	65.2%

As expected, consent rates for linkages were higher when interviews were conducted face-to-face or by telephone. For example, at Wave 2, for UCAS linkage, the consent rate was 83.3% when conducted face-to-face and 87.7% when conducted by telephone, compared with 68.6% online.

Table 11. Linkage consent rates for UCAS, HESA and SLC – main and boost sample combined- by mode

	Consent rate at Wave 2		
	Based on all who completed online (n=10,843)	Based on all who completed face-to-face (n=574)	Based on all who completed by telephone (n=106)
Consent requested at Wave 2 only			
Universities and Colleges Admissions Service (UCAS)	68.6%	83.3%	87.7%
Higher Education Statistics Agency (HESA)	67.8%	84.1%	88.7%
Student Loans Company (SLC)	64.3%	80.8%	78.3%

5 Survey Data

5.1 Summary of data

The survey data is available in two datafiles:

Young person data:

Interviews are classed as complete if all sections of the questionnaire are completed (including the incentive voucher admin questions ZMERIT1–3/ZMERITCAPICATI) and as usable partial interviews if the questionnaire is completed up to the start of CASI section (including ZWHLI). The W2_COMP_FLAG variable in the data set denotes completion status.

Table 12. Breakdown of young person interviews by type of sample and completion status

	Fully completed	Partial Complete - useable	Total
Total	11,483	125	12,104

Parent data:

 Interviews are classed as complete if all sections of the questionnaire are completed (including the incentive voucher admin questions ZMERIT1–3/ZMERITCAPICATI), and as usable partial interviews if the questionnaire is completed until the start of the CASI section (XBEENVAC for CAWI/CATI or accepted CASI or XCASISTART if CASI was refused for CAPI). The W2_COMP_FLAG variable in the data set denotes completion status.

Table 13. Breakdown of parent interviews by type of sample and completion status

	Fully completed	Partial Complete - useable	Total
Total	10,660	68	11,218

The young person is the primary cohort member so any parent interviews with no matching young person interview are not treated as part of the analytical sample. As such in the parent data file only those with a matching young person interview are weighted.

5.2 Quality checks

The quality checks were done in 2 stages.

Stage 1: Removal of non-valid cases

The files were first cleaned to remove non-valid cases as follows:

	Description	Exclusion criteria
Unusable partials	Did not reach the completion threshold	Exclude all
Duplicates	All were cases that completed on 2 or more different modes of CAWI/CAPI/CATI.	In these situations, we removed the least completed interview, or if the same completion status, we removed the later interview.

After these invalid cases were removed were left with 10,681 parent interviews and 11,523 young person interviews.

Stage 2: Quality assurance of cases to identify those which indicate poor quality data

Based on the remaining valid cases we then assessed the data across a range of quality flags including interview length, straight-lined across attitude batteries, and repeatedly picking only one option across multi-coded questions.

Based on examining the distribution of interview lengths we decided to flag all cases where the interview length was < 0.25 * median interview length.

Because the parent survey has noticeably different routing depending on if they answered the survey at Wave 1, their timing calculation and flags were done as separate groups. The timing points used were:

Parent, Wave 1 participant: Median time = 29 minutes 47 seconds.

Parent, Wave 1 non-participant: Median time = 32 minutes 41 seconds.

Young People: Median time = 29 minutes 02 seconds.

	Description	Flag variable
Short interview length	Flag cases where the length of interview is shorter than ¼ of the median length	Flagged in W2_Flag_TIMEquarter
Grid answers in a straight line	Many grid questions in each survey are structured so that selecting the same responses within a grid is fairly reasonable, e.g. having the same opinion of all 4 categories in ZAppExp	Flagged in W2_Flag_All_Straightline

	would be a legitimate response. So we focused on the grids where straight lining seemed more suspicious, for parents these were XYPJobAtt2 and XMGLPSC, for young people these were ZSchoolAtt2 and ZSOCPROV. If the respondent answered the same response in these grids they were flagged.	
Multi-response question response count	If every multi-response question was answered with just one response the interview was flagged.	Flagged in W2_Flag3oneanswer100
Email used for incentives check	The email used for incentives for parent and youth were compared and if they were the same both interviews were flagged in case this indicates the same person completed both surveys. Given that a household using the same email for a survey is understandable and not rare rather than look at every household where the same email was used we looked at cases where the interviews were started/ended on the same day within 5min of each other, and using the same device model, operating system and browser version.	Flagged in W2_same_email_consecutive

If a respondent was flagged for all 4 of the quality flags they were deleted. Only 2 parent interviews were removed this way, and no young people interviews were removed.

This left 10,678 parent interviews and 11,523 young person interviews, and a total of 10,204 complete households. Out of 10,678 parent interviews 474 of them did not receive a sampling weight because there was no matching young person interview (please see Section 6).

5.3 Licencing

The parent and young person datasets are available from the UK Data Service (UKDS). All users of the data need to be registered with the UKDS. Details of how to do this are available at https://www.ukdataservice.ac.uk/get-data/how-toaccess/registration

The datasets can be downloaded once the End User Licence access conditions have been accepted by the user. COSMO Wave 2 data available under End User Licence exclude detailed data that present a potential risk for disclosure. This applies to:

1. Verbatim responses to open-ended questions
2. Exact date of birth for young people
3. Exact age for parents
4. Detailed information on ethnicity, country of birth, and religion
5. Detailed information on sexual orientation
6. Full SOC employment codes

7. Detailed geographic information
8. School identifiers
9. Full working history (since the beginning of the pandemic in the parents' data set for new parents interviewed in Wave 2, and since the last interview for parents who also took part in Wave 1)
10. Full household grid in the young people data set
11. Variable levels with few cases (typically less than 10 individuals), or variables that could increase risk of disclosure when cross-tabulated with other variables

Some of these data may be made available to users within the ONS Secure Research Service, providing additional safeguards on disclosure risk, in due course. In addition to disclosivity, sensitivity of the data has also been assessed. For variables which cause concerns over consequences of re-identification, we have taken measures ranging from exclusion to reducing detail.

Please refer to section 5.9 for information on how these data have been deidentified for inclusion under End User Licence.

5.4 Identifiers

Household identifiers

The parent and young person interviews are in separate data sets and a household serial is included so interviews from the same household can be matched across the 2 datasets. This is the variable "W2_HHserial" which is a 6-digit serial.

This serial is consistent with the "W1_HHserial" in Wave 1 data and can be used to identify households across waves.

Individual identifiers

Each interview was assigned an individual serial, this is the "W2_HHserial" with "1" appended for young person interviews and "2" appended for parent interviews. This is the variable "W2_INDserial" which is a 7-digit serial.

This serial is consistent with the "W1_INDserial" in Wave 1 data and can be used to identify individuals across waves.

5.5 Variable names

Questionnaire variables in the data files were named to match the questionnaire question name whenever possible.

The standard convention used here for the naming of multi-responses and grid variables was to add a numeric suffix to the variable name in form of "VARNAME_01". For these suffixes we consistently used _96 for "Other", _97 for "None of These"/"None", _98 for "Don't Know" and _99 for "Prefer not to say".

For Wave 2 a prefix of "W2_" was added to variable names.

5.6 Variable description

For questionnaire variables the variable labels used in the data files are based on the wording from the survey questionnaire, shortened and kept comprehendible.

For multi-response and grid variables the variable labels were based on the wording of the question and response text from the questionnaire. For grids the value labels used were also taken from the wording from the survey questionnaire, for multi-response variables the value labels used were No/Yes to indicate if that response was selected by the respondent.

5.7 Missing values

The missing values used in the data files are used to identify questions with no valid answer, for these there are 2 types:

1) The codes –8 and –9 are used by respondents to denote the following:

   –8: Don't know

   –9: Refused/Prefer not to say

These codes above, whenever they exist, were explicitly selected by respondents in the questionnaire (or communicated as such to an interviewer if CAPI or CATI).

2) The codes –1 and –2 are used for where no respondent answer was recorded:

   –1: Not applicable

   –2: Question not asked due to respondent answers or script

The –1 "Not Applicable" code is used if the question was intentionally not asked due to script routing. The –2 "Question not asked due to respondent answers or script" is used if questions should have been asked but wasn't asked/no data recorded. These would be cases where responses based on "Other" verbatim coding meant the script did not move down the right route, or possible script issues caused an answer to not be recorded.

There is an exception in the data sets to the use of "–1" and "–2" for useable partial interviews after the cut off points (which were XBEENVAC for CAWI/CATI or accepted

CASI or XCASISTART if CASI was refused for CAPI for parents and ZWHLI for young people): If the case was a usable partial interview and the codes "-1" or "-2" were required for questions after the cut off, they were set to system missing instead. As shown at the beginning of the section, this applies to a small number of useable partial interviews and the variables that exist after the cut off points.

5.8 Variable order

The order of variables in the data files follow the questionnaire order as below:

- Identifier variables
- Sample information variables, including feedforward data from wave 1
- Whether interview was conducted as CAWI, CAPI or CATI
- Questionnaire variables in the same order
- NSSEC variables were added in the position of the work questions
- Para-data variables for interview device, interview time, number of interview sessions
- Completion flag
- Flags to denote inconsistencies in household grid data (see section 5.11)
- Geodemographic variables
- Schools level information variables
- Quality check flags (see section 5.2)
- Survey design variables
- Weighting variables

The para-data variables included are:

- W2_DeviceDetails_kantarDevice – Device used for interviews, if multiple devices used the last used is recorded. All CAPI interviews were done on laptops.
- W2_SURVEY_SUB – The month when the interview was completed
- W2_MULTI_SESSION – Number of different sessions the interview was completed over, recorded from the number of time the survey was opened
- W2_COMP_FLAG, Completion status of the interview.

Geodemographic variables included are:

- W2_Polar4_quintile – POLAR4 Quintile
- W2_Region – Region
- W2_IMD_decile – English Index of Multiple Deprivation (LSOA Decile)
- W2_IDAC_decile – English Income Deprivation Affecting Children Index (LSOA Decile)

The school level information included are:

- W2_EstablishmentTypeGroup_Y12/Y13 – School Establishment Type Group

- W2_AdmissionsPolicycode _Y12/Y13 – School Admissions Policy
- W2_PercentageFSMQuintiles _Y12/Y13 – Percentage FSM Pupils in School (Quintiles)
- W2_TrustSchoolFlagcode _Y12/Y13 – Trust School Flag
- W2_OfstedRating_Y12/Y13 – School Ofsted rating

5.9 Coding of disclosive information

Both data sets in this deposit have been assessed for disclosivity risk and some measures were taken to minimize the risk of identification of respondents. Below we summarize these measures.

5.9.1 Verbatim responses

The questionnaire collects some information as full verbatim answers. These have all been removed from the data files, the responses were used to either back code into existing responses or some new responses were made if there were enough verbatim answers of the same type.

Questions where new responses were added to the data based on verbatim answers were:

Parent file:

- W2_XRELATPAR
- W2_XChildPlan1
- W2_XAsuxApp
- W2_XASUX
- W2_ParPay
- W2_XHOMQUAL
- W2_XCOSTLIVCUT

Young People file:

- W2_ZALEVSUB
- W2_ZASLEVSUB
- W2_ZGCSESUB
- W2_ZBTECSUB
- W2_ZVCQC
- W2_ZSCHOOLMISS
- W2_ZFutHESu (fully created from open answer)
- W2_ZASUX
- W2_ZCARADVINF

Responses added from coding have the note "(created from coding)" in their labels.

Employment details given in the parent survey are used to derive SIC 2020, SOC 2020 and NSSEC for either respondent or their partner. Detailed SIC and SOC codes are excluded from this deposit, but NSSEC variables were added:

- W2_XNSSEC
- W2_XPNSSEC

The NSSEC coding is based on SOC 2020 using the ONS derivation tables linked here: https://www.ons.gov.uk/file?uri=/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020/soc2020volume3thenationalstatisticssocioeconomicclassificationnssecrebasedonthesoc2020/tables912v3.xlsx

User should note that information on occupation was not asked to parents if they have stated that it is the same as Wave 1. In this case, users are encouraged to merge this information from the Wave 1 dataset to have a full picture of occupation in Wave 2.

5.9.2 Top coding/bottom coding

The higher or lower ends of the distributions of some questions were recoded to minimize the risk of identification through extreme values.

In the young person questionnaire these include:

- W2_ZDOBY, Year of birth of youth
- W2_ZHHNUM, Number of people in household
- W2_ZHHAGE_01 – W2_ZHHAGE_06, Age band of persons 1 to 6 in household

In the parent questionnaire these include:

- W2_XDOBY, Year of birth of child
- W2_XAGEPAR, Age of parent
- W2_XNUMPEOPLE, Number of people living at parent's address
- W2_XWORK4AY, Year in which parent started current main working status
- W2_XPWORK4AY, Year in which parent's partner started current main working status
- W2_XINCBANDW, Weekly income bands for parent and partner
- W2_XBEDROOM, Number of bedrooms in home

5.9.3 Sensitive information

The young person questionnaire included questions on self-harm which are deemed highly sensitive, and are left out from the dataset.

- W2_ZHOSS, Youth still homeless
- W2_ZSelfHarm1, Whether youth has purposely hurt themselves in any way over the past 12 months
- W2_ZSelfHarm2, Whether youth has purposely hurt themselves in an attempt to end their life over the past 12 months

5.9.4 Combining response categories

Some response options were combined to reduce detail.

In the young person questionnaire these include:

- W2_FF_ZYPETHNIC, Ethnic group of youth Wave 1 answers
- W2_ZHHREL1_01 – W2_ZHHREL1_06, Relationship to youth of persons 1 to 6 in household into broader categories
- W2_ZMainStatW1, youth main status at Wave 1
- ZMainStatW2, Which youth classifies as current main activity
- ZWAVE2ACT, Youth main activity at W2 (Derived)
- W2_ZJHChoose1main, Main reason youth left job if they chose to leave
- W2_ZY13StateEng, Where current school attended by youth is
- W2_ZSOrient, Sexual orientation of youth
- W2_ZYPETHNIC, Ethnic group of youth
- W2_ZBornUK, Birth country youth
- W2_Zcountry, Birth country youth, outside of UK
- W2_Zreligion, Religion of youth
- A set of variables were also combined:W2_ZHHAGEN_01 – W2_ZHHAGEN_06/ W2_ZHHAGE_01 – W2_ZHHAGE_06, Age of persons 1 to 6 in household. The exact age is combined with banded age under variables W2_ZHHAGE_01 – W2_ZHHAGE_06, and W2_ZHHAGEN_01 – W2_ZHHAGEN_06 is deleted.
- W2_ZJHLeav1_1 – W2_ZJHLeav1_7/ W2_ZJHLeav1_8, Reason youth left job if they did not choose to leave. Individual reasons combined with "Another reason" under variable W2_ZJHLeav1_8 and W2_ZJHLeav1_1 – W2_ZJHLeav1_7 are deleted.
- W2_ZJHChoose107 – W2_ZJHChoose109/ W2_ZJHChoose110, Reason youth left job if they chose to leave. Combined under variable W2_ZJHChoose110, W2_ZJHChoose107 – W2_ZJHChoose109 are deleted.
- W2_ZQUAL_04/ W2_ZQUAL_05, Qualifications currently studying for – GCSE/ IGCSE. Combined under variable W2_ZQUAL_04, and W2_ZQUAL_05 is deleted.
- W2_ZGCSENUM/ W2_ZIGCSENUM, Number of GCSEs/ IGCSEs currently studying for. Combined under variable W2_ZGCSENUM, and W2_ZIGCSENUM is deleted.
- W2_ZGCSESUB_01 – W2_ZGCSESUB_96/ W2_ZIGCSESUB_01 – W2_ZIGCSESUB_96, GCSEs/IGCSEs currently studying for. Combined under variables W2_ZGCSESUB_01 – W2_ZGCSESUB_96, and W2_ZIGCSESUB_01 – W2_ZIGCSESUB_96 are deleted.
- W2_ZWHLI_06, W2_ZWHLI_07/ W2_ZWHLI_96, Youth lived in these places for at least three months. Combined under variable W2_ZWHLI_96, and W2_ZWHLI_06, W2_ZWHLI_07 are deleted.

In the parent questionnaire these include:

- W2_FF_XETHNIC, Ethnic group of parent wave 1 answers
- W2_FF_XGenderPar, Gender of parent wave 1 answers

- W2_FF_XRELIGION, religion of parent wave 1 answers
- W2_XGenderYP, Gender of child
- W2_XGenderPar, Gender of parent
- W2_XMarStat, Marital status of parent
- W2_XEconAcBefore, Parent's main status prior to lockdown (start of Mar 2020)
- W2_XECONACNEXT_01, Parent economic activity 1
- W2_XECONACNEXT_02, Parent economic activity 2
- W2_XECONACNEXT_03, Parent economic activity 3
- W2_XWork1, Parent's current main work status
- W2_XWorkDer, Parent's derived current work status
- W2_XPWork1, Partner's current main working status
- W2_XPWorkDer, Partner's derived working status
- W2_XTENURE, House tenure
- W2_XGPCountryA, Country respondent's mother was born in
- W2_XGPCountryB, Country respondent's father was born in
- W2_XGPCountryC, Country child's other grandmother was born in
- W2_XGPCountryD, Country child's other grandfather was born in
- W2_XETHNIC, Ethnic group of parent
- W2_XCOUNTRY, Country outside UK parent was born in
- W2_XPETHNICP2, Ethnic group of other parent
- W2_XPCOUNTRYP2, Country outside UK other parent was born in
- W2_XRELIGION, Religion of parent

A set of variables were also combined in the parents' data:

- W2_XAGEPARn/ W2_XAGEPAR Age of parent. The exact age is combined with banded age under variables W2_XAGEPAR, and XAGEPARn is deleted.

5.9.5 Other measures to reduce detail

In addition to the changes described above other variables were removed to reduce potentially identifiable detail such as exact birthday or detailed work history, or to compliment the capping of number of people in household questions at 6 or more.

In the young person questionnaire additional deleted variables were:

- W2_ZDOBD, Day of birth of youth
- W2_ZHHAGEN_01 to W2_ZHHAGEN_15, Exact age of all persons (person 1 to person 15) in household
- W2_ZHHGENDER_07 to W2_ZHHGENDER_15, Gender of person 7 in household to Gender of person 15 in household
- W2_ZHHAGE_07 to W2_ZHHAGE_15, Age of person 7 in household to Age of person 15 in household

- W2_ZHHREL1_07 to W2_ZHHREL1_15, Relationship to youth of person 7 in household to Relationship to youth of person 15 in household
- W2_ZJHLeav1main, Main reason youth left job if they did not choose to leave
- W2_ZMoveWhen_Month, W2_ZMoveWhen_Year, Approximate moving date.

In the parent questionnaire additional deleted variables were:

- W2_XDOBD, Day of birth of child
- W2_XECONACNEXT_04 to W2_XECONACNEXT_09, Parent economic activity 4 to Parent economic activity 9
- W2_XECONACSTOP2M_04 to W2_XECONACSTOP2M_09, Month when parent stopped economic activity 4 to Month when parent stopped economic activity 9
- W2_XECONACSTOP2Y_04 to W2_XECONACSTOP2Y_09, Year when parent stopped economic activity 4 to Year when parent stopped economic activity 9

Some non-questionnaire variables were also edited to reduce the amount of detail, for both the young person and parent files these were:

- W2_SURVEY_SUB, Date of survey submission
- W2_Region, Region
- W2_EstablishmentTypeGroup_Y12, W2_EstablishmentTypeGroup_Y13, School Establishment Type Group
- W2_FurtherEducationType_Y12, W2_FurtherEducationType_Y13, Further Education Type

5.10 Data errors and inconsistencies

Users should be aware of the following error and inconsistencies in the data:

Young person questionnaire:

- ZVCQC: Those who only have DV_APPRENTICE = 1 were not asked this question. This affected 270 interviews where this should have been asked but were not.

- ZFinCour, ZJHStop, ZStudyIss: These all used FF_ZCURSTAT in the filter logic, and due to a setup issue only respondents who ONLY had a particular response was counted by the filter. Particularly for the FF_ZCURSTAT = 1 condition only respondents who had ONLY that response were asked these questions. For ZFinCour 180 interviews who should have been asked were not, for ZJHStop this was 1,079 interviews and for ZStudyIss this was 2,347.

Parent questionnaire:

- XETHNIC, XBORNUK: For CAWI, respondents who those who refused or answered "Don't Know" at W1 were not re-asked these questions as intended. This affected 358 interviews where these should have been asked but were not.

The cases affected were all coded as -2 "Question not asked due to respondent answers or script" in the data.

5.11 Variables denoting data inconsistencies

The young person data includes questions on household members (gender, relationship to the young person, age group). In rare cases there are inconsistencies between these variables. After checking for these for each of the household members separately, we created the below variables to show if any inconsistencies existed or not in each young person's household.

- W2_HHFlag1: A variable to denote if more than 2 people have been selected as parents in the young person's household (a value of 1 shows this was observed, and a value of 0 shows otherwise).

- W2_HHFlag2: A variable to denote if, in the young person's household, any individual's relationship to the young person was reported as a parent or a grandparent while their age being reported as 18 years old or younger (a value of 1 shows this was observed, and a value of 0 shows otherwise).

- W2_HHFlag3: A variable to denote if anyone in the young person's household was reported to be their child while their age being 6 years old or older (a value of 1 shows this was observed, and a value of 0 shows otherwise).

- W2_HHgridFlag: A variable to denote if any of the above three inconsistencies are present in the young person's household (a value of 1 shows at least one inconsistency was observed, and a value of 0 shows no inconsistencies were observed).

Because of the efforts spent to ensure no persons or households could be identified (detailed in Section 5.9), not all above inconsistencies may be possible for data users to observe (some categories might have been combined to reduced detail, masking some inconsistencies listed above). The above flag variables were calculated on a data version that precedes disclosivity checks and carried over to the current version to ensure users can be aware of inconsistencies.

6 Weighting

6.1 Introduction

Weights need to be applied when conducting analysis to ensure that the sample is representative of the population and that the findings are generalisable. For this study, weights were needed for two reasons: (1) to compensate for the disproportionate sample design, and (2) to compensate for systematic non-response and attrition.

The archived Wave 2 dataset includes five different weight variables. The correct weight to use depends on the analysis that is being conducted.

For overall analysis there are two different weights:

- W2_AllFamilyFull_weight – should be used when analysing survey data for all complete households (from the main study and from the boost). There are weights for 10,204 households.

- W2_AllYPFull_weight – should be used when analysing only young people (from the main study and from the boost). There are weights for 11,523 respondents.

For analysis limited to cases eligible for the Sutton Trust Boost (both boost sample cases, and NPD sample cases which met the eligibility criteria for the boost sample) there are two different weights:

- W2_BoostFamilyFull_weight – should be used when analysing survey data only for complete households eligible for the Sutton Trust boost (where both the pupil and a parent in the household responded). There are weights for 1,705 complete households.

- W2_BoostYPFull_weight – should be used when analysing only young people eligible for the Sutton Trust boost (i.e., this includes data from some households where just the pupil responded to the survey). There are weights for 1,963 respondents.

In addition, a weight has been generated to allow for complete household analysis using Wave 1 and Wave 2 data:

- W2_AllW1W2Family_weight – should be used for complete household analysis where Wave 2 young person survey data is linked to parental data from Wave 1 or Wave 2. This analysis would be limited to using parent variables that were included in both Wave 1 and Wave 2. There are weights for 10,787 complete households.

- There were 475 households where only the parent was successfully interviewed at Wave 2. These cases have been included in the archived parents dataset but have

not been given a weight (the value is missing). This means that these cases will be excluded from analysis when any of the survey weights are applied.

Five further weights have been produced to analyse survey data linked to administrative education records (from the National Pupil Database (NPD)). Separate weights are required for this analysis, as not all respondents consented to having their survey responses linked to the administrative data. These weights compensate for systematic differences in agreement rates to the linkage. These weight variables are not included in the UK Data Archive datasets, and are only included in the datasets that are planned to be made available in the ONS Secure Research Service (SRS) where analysis with linked data will be possible.

- W2_AllFamily_NPD_weight – should be used when analysing survey data for complete households linked to NPD education records. There are weights for 8,912 households.

- W2_AllYP_NPD_weight – should be used when analysing survey data for young people only, linked to the NPD education records. There are weights for 10,055 respondents.

- W2_BoostFamily_NPD_weight – should be used when analysing survey data for complete households eligible for the Sutton Trust boost linked to NPD education records. There are weights for 1,499 complete households.

- W2_BoostYP_NPD_weight – should be used when analysing survey data for young people eligible for the Sutton Trust boost linked to NPD education records. There are weights for 1,734 respondents.

- W2_AllW1W2FamilyNPD_weight – should be used when analysing wave 2 young person survey data with parental data from Wave 1 or Wave 2 linked to NPD education records. There are weights for 9,408 complete households.

6.2 Approach used to derive Wave 1 weights

A four-stage process was used to derive the Wave 1 weights. This process is briefly summarised below and the full process is outlined in Section 6 of the Wave 1 user guide.

First, all respondents were given a design weight equal to one divided by their sampling probability.

Next, all respondents were given a non-response weight equal to one divided by their estimated response probability (based on regression modelling). Every respondent was then given a 'base' weight equal to one divided by the product of the sampling and estimated response probabilities.

The base-weighted respondent sample was then calibrated so that its distribution with respect to some critical variables was an *exact* match for the estimation population, *so far as this is known*.

Finally, a stage of constrained calibration weighting was used to reduce the variance of the weights. The calibration weight was divided by the design weight to calculate a non-inclusion weight. The non-inclusion weights were trimmed, and the respondent sample was re-calibrated using the trimmed weights as base weights.

6.3 Approach used to derive Wave 2 weights

The wave two weighting scheme was designed to: (a) take into account the weights assigned to respondents at Wave 1, (b) up-weight respondents who had a relatively lower probability of taking part in Wave 2, and (c) down-weight respondents who had a relatively higher probability of taking part in Wave 2.

A two-step process was used:

• Regression modelling was used to model attrition from Wave 1 and to derive weights that compensated for this

• A final raking stage was then used to ensure the sample exactly matched marginal population totals (using the same benchmarks as at Wave 1)

This process was used to generate all Wave 2 weights. These steps are outlined in more detail below.

6.3.1 Modelling and compensating for attrition

At this stage, a weight was calculated to compensate for the original sample design, non-response at Wave 1, and attrition between Wave 2 and Wave 1. The broad approach can be summarised as follows:

$$\text{W2\_weight}_i = \text{W1\_YPweight}_i * [1/ \text{Pr}(\text{W2}|\text{W1})_i]$$

Where:

W1_YPweight_i is the weight that was assigned to the Young Person respondent *i* at Wave 1; and

$\text{Pr}(\text{W2}|\text{W1})_i$ is the estimated probability that the Wave 1 respondent *i* will also have responded at Wave 2.

Below we outline which Wave 1 weights were used in this derivation and the approach used to estimate Pr(W2|W1).

The Wave 1 weight used

The W1_YP weight used to derive each Wave 2 weight is listed in the table below.

New Wave 2 weight	Wave 1 weight used in derivation
W2_AllYPFull_weight	W1_AllYPFull_weight
W2_AllFamilyFull_weight	W1_AllYPFull_weight
W2_AllYP_NPD_weight	W1_AllYPFull_weight
W2_AllFamily_NPD_weight	W1_AllYPFull_weight
W2_BoostYPFull_weight	W1_BoostYPFull_weight
W2_BoostFamilyFull_weight	W1_BoostYPFull_weight
W2_BoostYP_NPD_weight	W1_BoostYPFull_weight
W2_BoostFamily_NPD_weight	W1_BoostYPFull_weight
W2_AllW1W2Family_weight	W1_AllYPFull_weight
W2_AllW1W2FamilyNPD_weight	W1_AllYPFull_weight

It should be noted that:

- For all Wave 2 "Family" weights we used the "YP" weight from Wave 1 in the derivation. This is because at wave 2 we were attempting to interview parents that did not take part at Wave 1 (and therefore some complete households at Wave 2 were households that were not complete at Wave 1).

- For all Wave 2 "NPD" weights we used the "Full" weight from Wave 1. This is because at Wave 2 we obtained consent for NPD linkage from some individuals that had not consented at Wave 1.

Regression modelling

A logistic regression was used to estimate the probability that a Wave 1 respondent also responded at Wave 2 (referred to as Pr(W2|W1)). The regression models used Wave 1 ask all questions from the young person questionnaire as predictors. These variables were limited to those that were found to be the most predictive of whether an interview was obtained at Wave 2. This selection used the process described below.

The first step in selecting predictors was to inspect the Wave 1 frequency distributions of variables (based on all Wave 1 young people interviews, n=13,786). In total, 95 variables were identified as being potentially suitable for the modelling.

Some derived variables were then produced to (i) combine variables that were highly correlated with one another, or (ii) to combine low frequency categories together. Following this review, we were left with 61 variables.

We then carried out variable-by-variable chi-square tests of independence with respect to Wave 2 young person interview status. Any variables with a p-value > 0.1 were then eliminated because they were unlikely to provide any additional predictive value to the model. Following this analysis, we were left with a list of 36 candidate predictors to include in the regression models.

These 36 predictor variables were then fed into a backstep logistic regression. A backstep algorithm was used which eliminated candidate predictors one by one if they did not improve the model fit.[15] The use of a backstep regression meant that the final variables included in the model for each weight varied slightly.

The base for the regression model used to estimate Pr(W2|W1) for each Wave 2 weight is provided in the table below. It should be noted that these are also the bases for the Wave 1 weights used in the derivation.

Wave 2 weight	Base for regression modelling
W2_AllYPFull_weight	
W2_AllFamilyFull_weight	
W2_AllYP_NPD_weight	All young people interviewed at Wave 1 (13,786)
W2_AllFamily_NPD_weight	
W2_AllW1W2Family_weight	
W2_AllW1W2FamilyNPD_weight	
W2_BoostYPFull_weight	
W2_BoostFamilyFull_weight	All Sutton Boost eligible young people interviewed at Wave 1 (2,249)
W2_BoostYP_NPD_weight	
W2_BoostFamily_NPD_weight	

Appendix 1 includes the model outputs for W2_AllYPFull_weight and W2_AllFamilyFull_weight, as well as the list of final variables used in the modelling for each weight.

---

[15] Model fit was assessed based on the AIC (Akaike Information Criterion).

6.3.2 Raking to population totals

The final stage of weighting was to calibrate the respondent sample so that its distribution with respect to some critical variables is an exact match for the population.[16]

The weight generated at step one was input as the base weight and we then used the iterative proportional fitting algorithm (also known as 'raking' or 'rim weighting') to generate the final weight.

We used the same variables and benchmarks as at Wave 1:

- Size of school's Year 11:
    - Under 150 pupils
    - 150–249
    - 250+ pupils
    - Independent in Y11 and Y12*
- Type of school provision:
    - Special
    - Alternative
    - Selective Other
    - Other
    - Independent in Y11 and Y12*
- Region:
    - East Midlands
    - East of England
    - London
    - North East
    - North West
    - South East
    - South West
    - West Midlands
    - Yorkshire and the Humber
    - Independent in Y11 and Y12*
- FSM eligibility by SEND status:
    - FSM last 6 years & EHC plan
    - FSM last 6 years & other SEND status
    - FSM last 6 years & no SEND status
    - No FSM last 6 years & EHC plan*
    - No FSM last 6 years & other SEND status*
    - No FSM last 6 years & no SEND status*
    - Independent in Y11 and Y12*

---

[16] Different benchmarks were used for the overall sample weights and for the Sutton Boost eligible weights. The targets were based on population counts from the NPD database used to source the state school sample. Further information on these benchmarks can be found in the Wave 1 user guide.

- Language
  - English is primary language / not recorded
  - English is an additional language
  - Independent in Y11 and Y12*
- Sex:
  - Male
  - Female
  - Independent in Y11 and Y12*
- Ethnic group:
  - Indian
  - Bangladeshi
  - Pakistani
  - Black African
  - Black Caribbean
  - White British / no data
  - White non–British
  - Mixed / Other
  - Independent in Y11 and Y12*
- Sex by broad ethnic group:
  - Male White British
  - Male Other
  - Female White British
  - Female Other
  - Independent in Y11 and Y12*
- KS2 scores (maths / reading / GPS)
  - Upper tertile in all three
  - Upper tertile in two, middle tertile in one
  - Upper tertile in one, middle tertile in two
  - Others with at least one in upper tertile or at least two in middle tertile
  - Lower tertile in two, middle tertile in one*
  - Lower tertile in all three*
  - Missing data*
  - Independent in Y11 and Y12*

*These categories were not used for any of the Sutton Boost eligible weights (as no members of the boost eligible population were included in these cells).

6.4 Sample representativity and effectiveness of weights

To examine the effectiveness of the weights in restoring sample representativity we have compared the final weighted survey sample profiles to the benchmark population statistics (which were used when calibrating the data[17]).

These comparisons are presented in Appendix 2.

6.5 Estimation of standard errors

To ensure that standard errors are estimated correctly it is important to take into account the impact of the weighting, clustering and pre-stratification. If this is not done, the confidence intervals estimated are likely to be too narrow and there is an increased risk of Type I errors (false positives).

The variables that need to be used:

- Weight variable – as outlined in the Weighting section of this user guide (section 6), the correct weight needs to be selected for each analysis. The four weights currently available are:
  – W1_AllFamilyFull_weight
  – W2_AllYPFull_weight
  – W2_BoostFamilyFull_weight
  – W2_BoostYPFull_weight
  – W2_AllW1W2Family_weight
  – W2_AllW1W2FamilyNPD_weight
  – W2_AllFamily_NPD_weight – note this weight variable will be available in the ONS SRS deposit only
  – W2_AllYP_NPD_weight – note this weight variable will be available in the ONS SRS deposit only
  – W2_BoostFamily_NPD_weight – note this weight variable will be available in the ONS SRS deposit only

---

[17] The benchmark population statistics are sourced from the NPD database (at the time the original sample was drawn). These benchmarks do not cover independent schools and there are no comprehensive population statistics for this population. As such, "independent schools" have been included as a standalone category in the variables used for final stage of raking used in the weighting. This ensures that independent school pupils are at the correct level in the sample.

- W2_BoostYP_NPD_weight – note this weight variable will be available in the ONS SRS deposit only
- Cluster variable: W2_PSU_all
- Stratification variable*: W2_AnalysisStratum_v2

*If users run into issues when conducting sub-group analysis because of there not being two clusters in each stratum, we would suggest conducting the analysis with W2_SchoolStratum_v2. If there are further singleton stratum problems when using W2_SchoolStratum_v2, we would recommend omitting the stratification variable entirely from the survey design. While these adjustments may be necessary for standard errors to be estimated, it should be noted that they are likely to lead to slightly inflated standard error estimates.

Below we have provided exemplar code for specifying the survey design correctly in different analysis programs.

6.5.1 Stata – using the svy[18] commands

In Stata robust standard errors can be estimated using the survey commands.

Before conducting any analysis, the survey design needs to be declared for the dataset. Note that the survey design declared will need to be changed each time a different weight needs to be used (changing the text highlighted in yellow below).

```
svyset W2_PSU_all [pweight= W2_AllFamilyFull_weight],
strata(W2_AnalysisStratum_v2)
```

Subsequent commands should then be conducted using the svy prefix – e.g.,

```
svy: proportion
```

6.5.2 R – using the "survey"[19] package

First, an object specifying the survey design needs to be created. The survey design needs to reference the object in which the dataset is stored (text highlighted in green below). A different survey design object will need to be created for each weight (changing the text highlighted in yellow to reference the correct weight, and the text highlighted in grey to change the name of the object that will store each survey design).

---

[18] https://www.stata.com/manuals/svy.pdf
[19] https://cran.r-project.org/web/packages/survey/survey.pdf

```
        library(survey)

design1 <- svydesign(id=~ W2_PSU_all,

                strata = ~ W2_AnalysisStratum_v2,

                weights=~ W2_AllFamilyFull_weight,

                data=DataObject)
```

This survey design object should then be referenced in later analysis which is conducted using the "survey" package – e.g., svymean, svyglm, etc.

6.5.3 SPSS – using the Complex Samples module[20]

A complex sample plan file needs to be saved (the file name and location need to be specified – see text highlighted in grey). Note that a separate plan file needs to be created for each weight – changing the weight variable name (highlighted in yellow) and the *.csaplan file name (highlighted in grey).

```
CSPLAN ANALYSIS
  /PLAN FILE='\\location file should be saved\file name.csaplan'
  /PLANVARS ANALYSISWEIGHT= W2_AllFamilyFull_weight
  /SRSESTIMATOR TYPE=WR
  /PRINT PLAN
  /DESIGN STRATA=W2_AnalysisStratum_v2 CLUSTER=W2_PSU_all
  /ESTIMATOR TYPE=WR.
```

This sample plan should then be referenced when doing analysis using the Complex Samples module of SPSS – e.g. CSDESCRIPTIVES, CSTABULATE, etc.

---

[20]   https://www.ibm.com/downloads/cas/5RWERDKG

6.6 Weighting variables in datafiles

The weighting variables in the datafiles are:

Table 14. List of weights by data deposit

		Included in UKDS data	To be included in SRS data
W2_AnalysisStratum_v2	Analysis stratum – scrambled	Y	Y
W2_SchoolStratum_v2	School stratum – scrambled	Y	Y
W2_PupilStratum_v2	Pupil stratum – scrambled	Y	Y
W2_PSU_all	PSU	Y	Y
W2_BoostFamilyFull_weight	Final weight: Eligible for Sutton Trust Boost – All Complete households	Y	Y
W2_BoostYPFull_weight	Final weight: Eligible for Sutton Trust Boost – All Young People	Y	Y
W2_BoostFamily_NPD_weight	Final weight: Eligible for Sutton Trust Boost – Complete households that consented to NPD linkage		Y
W2_BoostYP_NPD_weight	Final weight: Eligible for Sutton Trust Boost – Young People that consented to NPD linkage		Y
W2_AllFamilyFull_weight	Final weight: All (Main & Sutton Trust Boost) – All Complete households	Y	Y
W2_AllYPFull_weight	Final weight: All (Main & Sutton Trust Boost) – All Young People	Y	Y
W2_AllFamily_NPD_weight	Final weight: All (Main & Sutton Trust Boost) – Complete households that consented to NPD linkage		Y
W2_AllYP_NPD_weight	Final weight: All (Main & Sutton Trust Boost) – Young People that consented to NPD linkage		Y
W2_AllW1W2Family_weight	Final weight all (main + boost) Full HH either wave parents (YP + W1/W2 parent)	Y	Y
W2_AllW1W2FamilyNPD_weight	Final weight all (main + boost) Full HH either wave parent (YP + W1/W2 parent) linked to NPD in W1/W2		Y

7 Mode Effects

As described in sections 1 and 4, after the initial online fieldwork stage, the survey involved two interviewer-led stages, a face-to-face phase and a telephone phase, which were used to offer non-responders to the web survey an alternative method of completing the survey.

The mode by which each respondent completed the survey is recorded in the following variables:

- For young people – W2_ZMODE
- For parents – W2_XMODE

The following table shows the number of respondents that participated using each mode:

Table 15. Distribution of interviews at Wave 2 by mode (main and boost combined)

	Total	Online	Face-to-face	Telephone
Young People	11,523	10,843	574	106
All parents, of which:	10,678	9,748	884	46
– Wave 1 parent	8,760	8,761	397	46
– New parent	1,918	1,431	487	0

When using survey data collected using multiple modes, it is important to consider how this may affect analyses. "Mode effects" are generally taken to mean differences in observed responses to survey items which are due solely to the mode of data collection.

The survey was initially designed as an online questionnaire. However, in translating to a mixed-mode questionnaire attempts were made to ensure that the face-to-face and telephone versions were as equivalent as possible to the online approach.

No rotation or randomisation of responses was used so the response lists were presented in the same order for all modes. For the face-to-face survey, most questions (except for simple yes/no style questions) used showcards and all "Don't Know" and "Prefer not to say" options were made explicit in both modes; this meant that these options were included on showcards in the face-to-face version. This allowed the questions to be presented in an equivalent way for face-to-face and online respondents. It is more difficult to make a telephone survey equivalent to online and face-to-face approaches as there are no visual cues in a telephone survey. Therefore, to make the telephone survey as equivalent as possible, all lists in CATI were prompted and read out by the interviewer, except in the case of very short questions such as yes/no questions or

factual questions such as tenure where a 'Prompt as necessary' approach was used. 'Don't know' options were also read out by the interviewer to make these explicit as in other modes.

Additionally, the face-to-face interviews for both parents and young people included a self-completion (CASI) section which respondents completed on their own and which included some of the more sensitive items (please see section 3, Tables 1 and 2 to see content covered in the CASI sections). This meant that respondents could answer the most sensitive questions in private which encourages greater honesty.

On telephone, it wasn't possible to collect data via self-completion for the more sensitive questions (we considered but rejected the idea of doing a CASI follow-up due to the high rates of expected attrition). Therefore, most questions in the CASI section were read out by interviewers. For the parent survey, all CASI questions were included in the telephone survey. For the young person survey, all CASI questions were included except for the following questions which were deemed exceptionally sensitive and unsuitable for interviewer administration (for further detail of these exclusions, please refer to Chapter 3, Table 1):

- Peer relationships, loneliness, bullying and harassment

- Seeking help for mental health

- Self harm and suicide

- Sexual orientation

Despite the efforts described above, mode effects are unavoidable as the three approaches can never be truly identical, and mode effects are particularly likely to be apparent when comparing the telephone mode with other modes. However, as < 1% of all interviews were conducted by telephone, the impact of this will be negligible. Some examples of why measurement may still vary between modes:

- Face-to-face and telephone interviewers can provide motivation or clarification when required; this cannot truly be replicated online.

- People who would not disclose sensitive personal information or socially undesirable opinions/behaviours to an interviewer may be more willing to provide this information online. In particular, for the questions drawn from the CASI module which were read out by telephone interviewers, we would expect to see mode effects as it is likely that some respondents were less inclined to give honest respondents in the presence of an interviewer. The questions especially likely to be affected by this are questions related to mental health.

- On the telephone, there is the possibility of so-called 'recency effects' when reading out long lists, whereby respondents are more likely to recall and select the later items in a read-out list.

In addition, it should be noted that respondents were not randomly allocated to mode. As respondents self-selected into each mode, they are likely to differ in potentially important ways.

Without appropriate control for these (possibly unobserved) characteristics, it is not necessarily possible to determine whether an observed between-mode difference in a given variable is due to selection or truly a mode effect (or a combination of both).

8 Appendices

APPENDIX 1 – Attrition modelling outputs

All full households (10,204 respondents)

Table A1. Binary logistic regression predicting whether a full household was interviewed at Wave 2 (based on households where a young person was recruited at Wave 1)

Parameter	B	Std. Error	95% Confidence Interval		P–value
			Lower	Upper	
(Intercept)	1.07	0.39	0.32	1.85	0.01
Sample type & incentive (Boost & £20 vs Boost & £10)	0.74	0.62	–0.36	2.12	0.23
Sample type & incentive (Core & £10 vs Boost & £10)	–0.05	0.23	–0.52	0.38	0.83
Sample type & incentive (Core & £20 vs Boost & £10)	0.62	0.25	0.12	1.09	0.01
Sample type & incentive (Independent school & no incentive vs Boost & £10)	–1.59	0.25	–2.10	–1.12	0.00
Mode of interview (CAPI vs CAWI)	–1.40	0.13	–1.65	–1.15	0.00
Number of data linkage consents given in wave 1 (1 vs 0)	–0.04	0.12	–0.27	0.20	0.75
Number of data linkage consents given in wave 1 (2 vs 0)	0.14	0.13	–0.11	0.39	0.27
Number of data linkage consents given in wave 1 (3 vs 0)	0.08	0.12	–0.14	0.31	0.47
Number of data linkage consents given in wave 1 (4 vs 0)	0.67	0.13	0.42	0.93	0.00
Number of data linkage consents given in wave 1 (5 vs 0)	0.51	0.07	0.38	0.64	0.00
Likelihood of going to university (Fairly likely vs Very likely)	0.07	0.07	–0.07	0.21	0.30
Likelihood of going to university (Not very likely vs Very likely)	–0.32	0.08	–0.48	–0.17	0.00
Likelihood of going to university (Not at all likely vs Very likely)	–0.20	0.09	–0.37	–0.02	0.03
Likelihood of going to university (Missing vs Very likely)	0.06	0.12	–0.17	0.29	0.64
Main activity (Working vs Studying)	–0.28	0.11	–0.50	–0.06	0.01
Main activity (Other vs Studying)	–0.42	0.14	–0.70	–0.14	0.00
Number of people living in household (2 vs 1)	–0.13	0.25	–0.64	0.34	0.59
Number of people living in household (3 vs 1)	–0.06	0.23	–0.54	0.38	0.80
Number of people living in household (4 vs 1)	0.17	0.23	–0.31	0.61	0.47
Number of people living in household (5 vs 1)	0.29	0.24	–0.19	0.74	0.22
Number of people living in household (6 vs 1)	0.31	0.25	–0.19	0.78	0.20
Number of people living in household (7+ vs 1)	0.38	0.26	–0.14	0.87	0.14
Feeling unmotivated (Only one lockdown vs Both lockdowns)	–0.21	0.08	–0.36	–0.06	0.01
Feeling unmotivated (Neither lockdown vs Both lockdowns)	–0.12	0.06	–0.25	0.00	0.05

Attended school in person during Year 11 (No vs Yes)	−0.21	0.13	−0.47	0.05	0.11
Attended school in person during the first lockdown due to school re-opening (Yes vs No)	0.18	0.06	0.05	0.30	0.01
Contact with teachers outside of lessons during the first lockdown (About once a week vs More than once a week)	−0.05	0.08	−0.21	0.11	0.55
Contact with teachers outside of lessons during the first lockdown (About once every two weeks vs More than once a week)	0.15	0.11	−0.06	0.36	0.17
Contact with teachers outside of lessons during the first lockdown (About once a month vs More than once a week)	0.11	0.12	−0.12	0.34	0.37
Contact with teachers outside of lessons during the first lockdown (Less often vs More than once a week)	0.27	0.10	0.08	0.46	0.01
Contact with teachers outside of lessons during the first lockdown (Not at all during this period vs More than once a week)	0.18	0.09	0.00	0.36	0.05
Contact with teachers outside of lessons during the first lockdown (Missing vs More than once a week)	0.06	0.12	−0.17	0.29	0.63
Year of birth (2004 and earlier vs 2005 and later)	−0.20	0.05	−0.30	−0.09	0.00
If someone is not a success in life, it is usually their own fault (Agree vs Strongly agree)	−0.01	0.12	−0.24	0.22	0.93
If someone is not a success in life, it is usually their own fault (Disagree vs Strongly agree)	0.15	0.12	−0.08	0.37	0.20
If someone is not a success in life, it is usually their own fault (Strongly disagree vs Strongly agree)	0.05	0.13	−0.21	0.32	0.69
If someone is not a success in life, it is usually their own fault (Missing vs Strongly agree)	−0.04	0.14	−0.31	0.22	0.75
Days per week spent on school work during first lockdown (1 vs 0)	0.01	0.13	−0.24	0.27	0.91
Days per week spent on school work during first lockdown (2 vs 0)	−0.09	0.12	−0.32	0.15	0.46
Days per week spent on school work during first lockdown (3 vs 0)	−0.09	0.11	−0.31	0.13	0.43
Days per week spent on school work during first lockdown (4 vs 0)	0.06	0.13	−0.19	0.32	0.64
Days per week spent on school work during first lockdown (5 vs 0)	0.30	0.10	0.11	0.49	0.00
Days per week spent on school work during first lockdown (6+ vs 0)	0.42	0.13	0.17	0.68	0.00
Days per week spent on school work during first lockdown (Missing vs 0)	−0.10	0.13	−0.34	0.15	0.45
Spoke to parent(s)/guardian(s) about decisions for future (Yes vs No)	0.19	0.07	0.05	0.32	0.01
Attended school in person during the first lockdown due to special education needs (Yes vs No)	−0.37	0.16	−0.69	−0.05	0.02
Attended school in person during the first lockdown due to boarding school/in care/other reason (Yes vs No)	−0.19	0.12	−0.43	0.05	0.11
Had to share devices with other family members (Only one lockdown vs Both lockdowns)	0.31	0.15	0.02	0.59	0.04

Had to share devices with other family members (Neither lockdown vs Both lockdowns)	0.15	0.10	−0.05	0.35	0.13
Having a job/career in future is important (Agree a little vs Agree strongly)	0.24	0.09	0.07	0.41	0.01
Having a job/career in future is important (Disagree a little vs Agree strongly)	−0.14	0.21	−0.54	0.29	0.52
Having a job/career in future is important (Disagree strongly vs Agree strongly)	−0.46	0.30	−1.04	0.15	0.13
Having a job/career in future is important (Missing vs Agree strongly)	0.30	0.17	−0.02	0.63	0.07
Falling behind classmates due to COVID-19 (Agree slightly vs Agree strongly)	0.01	0.09	−0.16	0.18	0.89
Falling behind classmates due to COVID-19 (Neither agree nor disagree vs Agree strongly)	0.15	0.09	−0.02	0.32	0.08
Falling behind classmates due to COVID-19 (Disagree slightly vs Agree strongly)	0.27	0.10	0.08	0.47	0.01
Falling behind classmates due to COVID-19 (Disagree strongly vs Agree strongly)	0.28	0.10	0.09	0.47	0.00
Falling behind classmates due to COVID-19 (Not applicable vs Agree strongly)	−0.04	0.16	−0.35	0.27	0.79
Spoke to brother/sister about decisions for future (Yes vs No)	−0.08	0.06	−0.20	0.03	0.15
Had regular problems with internet connection (Only one lockdown vs Both lockdowns)	0.04	0.11	−0.16	0.25	0.70
Had regular problems with internet connection (Neither lockdown vs Both lockdowns)	−0.15	0.07	−0.28	−0.02	0.03
School provided real-time online learning for other reasons (Only one lockdown vs Neither lockdown)	−0.16	0.11	−0.38	0.06	0.16
School provided real-time online learning for other reasons (Both lockdowns vs Neither lockdown)	−0.20	0.09	−0.38	−0.02	0.03

All Young People (11,523 respondents)

Table A2. Binary logistic regression predicting whether young people recruited at Wave 1 also responded to Wave 2

Parameter	B	Std. Error	95% Confidence Interval		P-value
			Lower	Upper	
(Intercept)	1.09	0.37	0.37	1.84	0.00
Sample type & incentive (Boost & £20 vs Boost & £10)	0.71	0.61	–0.38	2.08	0.25
Sample type & incentive (Core & £10 vs Boost & £10)	–0.06	0.23	–0.53	0.36	0.78
Sample type & incentive (Core & £20 vs Boost & £10)	0.54	0.24	0.05	1.00	0.03
Sample type & incentive (Independent school & no incentive vs Boost & £10)	–1.42	0.24	–1.92	–0.97	0.00
Mode of interview (CAPI vs CAWI)	–1.33	0.12	–1.58	–1.09	0.00
Number of data linkage consents given in wave 1 (1 vs 0)	–0.02	0.12	–0.24	0.21	0.88
Number of data linkage consents given in wave 1 (2 vs 0)	0.13	0.13	–0.11	0.38	0.28
Number of data linkage consents given in wave 1 (3 vs 0)	0.06	0.11	–0.16	0.29	0.58
Number of data linkage consents given in wave 1 (4 vs 0)	0.66	0.13	0.41	0.92	0.00
Number of data linkage consents given in wave 1 (5 vs 0)	0.51	0.06	0.39	0.64	0.00
Likelihood of going to university (Fairly likely vs Very likely)	0.06	0.07	–0.08	0.19	0.44
Likelihood of going to university (Not very likely vs Very likely)	–0.33	0.08	–0.48	–0.17	0.00
Likelihood of going to university (Not at all likely vs Very likely)	–0.21	0.09	–0.39	–0.04	0.01
Likelihood of going to university (Missing vs Very likely)	0.08	0.12	–0.15	0.31	0.51
Main activity (Working vs Studying)	–0.28	0.11	–0.49	–0.05	0.01
Main activity (Other vs Studying)	–0.38	0.14	–0.65	–0.11	0.01
Number of people living in household (2 vs 1)	–0.16	0.24	–0.65	0.30	0.50
Number of people living in household (3 vs 1)	–0.21	0.23	–0.67	0.22	0.36
Number of people living in household (4 vs 1)	0.02	0.22	–0.44	0.45	0.93
Number of people living in household (5 vs 1)	0.17	0.23	–0.30	0.60	0.45
Number of people living in household (6 vs 1)	0.17	0.24	–0.32	0.62	0.49
Number of people living in household (7+ vs 1)	0.23	0.25	–0.28	0.70	0.36
Feeling unmotivated (Only one lockdown vs Both lockdowns)	–0.23	0.07	–0.37	–0.08	0.00

Feeling unmotivated (Neither lockdown vs Both lockdowns)	–0.12	0.06	–0.24	0.00	0.06
Attended school in person during Year 11 (No vs Yes)	–0.25	0.13	–0.51	0.01	0.06
Attended school in person during the first lockdown due to school re-opening (Yes vs No)	0.16	0.06	0.04	0.28	0.01
Contact with teachers outside of lessons during the first lockdown (About once a week vs More than once a week)	–0.10	0.08	–0.26	0.06	0.21
Contact with teachers outside of lessons during the first lockdown (About once every two weeks vs More than once a week)	0.06	0.11	–0.15	0.26	0.59
Contact with teachers outside of lessons during the first lockdown (About once a month vs More than once a week)	0.04	0.12	–0.19	0.27	0.76
Contact with teachers outside of lessons during the first lockdown (Less often vs More than once a week)	0.20	0.10	0.01	0.39	0.04
Contact with teachers outside of lessons during the first lockdown (Not at all during this period vs More than once a week)	0.09	0.09	–0.09	0.27	0.33
Contact with teachers outside of lessons during the first lockdown (Missing vs More than once a week)	–0.04	0.12	–0.27	0.19	0.71
Year of birth (2004 and earlier vs 2005 and later)	–0.18	0.05	–0.29	–0.08	0.00
My current school provided additional support to help me catch up on learning lost due to COVID (Agree slightly vs Agree strongly)	0.20	0.10	0.00	0.38	0.04
My current school provided additional support to help me catch up on learning lost due to COVID (Neither agree nor disagree vs Agree strongly)	0.26	0.09	0.08	0.45	0.00
My current school provided additional support to help me catch up on learning lost due to COVID (Disagree slightly vs Agree strongly)	0.26	0.11	0.05	0.47	0.01
My current school provided additional support to help me catch up on learning lost due to COVID (Disagree strongly vs Agree strongly)	0.29	0.10	0.09	0.49	0.00
My current school provided additional support to help me catch up on learning lost due to COVID (Not applicable vs Agree strongly)	0.26	0.10	0.06	0.46	0.01
If someone is not a success in life, it is usually their own fault (Agree vs Strongly agree)	–0.06	0.12	–0.29	0.17	0.62
If someone is not a success in life, it is usually their own fault (Disagree vs Strongly agree)	0.11	0.11	–0.12	0.33	0.33
If someone is not a success in life, it is usually their own fault (Strongly disagree vs Strongly agree)	0.06	0.13	–0.21	0.31	0.68
If someone is not a success in life, it is usually their own fault (Missing vs Strongly agree)	–0.07	0.13	–0.33	0.19	0.61
Days per week spent on school work during first lockdown (1 vs 0)	0.01	0.13	–0.24	0.26	0.94
Days per week spent on school work during first lockdown (2 vs 0)	–0.11	0.12	–0.34	0.12	0.35

Days per week spent on school work during first lockdown (3 vs 0)	−0.10	0.11	−0.32	0.12	0.36
Days per week spent on school work during first lockdown (4 vs 0)	0.09	0.13	−0.16	0.34	0.50
Days per week spent on school work during first lockdown (5 vs 0)	0.26	0.10	0.07	0.44	0.01
Days per week spent on school work during first lockdown (6+ vs 0)	0.40	0.13	0.16	0.65	0.00
Days per week spent on school work during first lockdown (Missing vs 0)	−0.07	0.12	−0.31	0.18	0.59
Spoke to parent(s)/guardian(s) about decisions for future (Yes vs No)	0.16	0.07	0.03	0.30	0.02
Attended school in person during the first lockdown due to special education needs (Yes vs No)	−0.29	0.16	−0.59	0.03	0.07
Attended school in person during the first lockdown due to boarding school/in care/other reason (Yes vs No)	−0.15	0.12	−0.38	0.08	0.20
Had to share devices with other family members (Only one lockdown vs Both lockdowns)	0.30	0.14	0.02	0.58	0.04
Had to share devices with other family members (Neither lockdown vs Both lockdowns)	0.16	0.10	−0.04	0.35	0.11
Having a job/career in future is important (Agree a little vs Agree strongly)	0.22	0.08	0.05	0.38	0.01
Having a job/career in future is important (Disagree a little vs Agree strongly)	−0.02	0.21	−0.41	0.41	0.94
Having a job/career in future is important (Disagree strongly vs Agree strongly)	−0.52	0.29	−1.08	0.08	0.08
Having a job/career in future is important (Missing vs Agree strongly)	0.21	0.16	−0.11	0.54	0.20
Falling behind classmates due to COVID–19 (Agree slightly vs Agree strongly)	−0.02	0.09	−0.19	0.16	0.86
Falling behind classmates due to COVID–19 (Neither agree nor disagree vs Agree strongly)	0.10	0.09	−0.08	0.27	0.27
Falling behind classmates due to COVID–19 (Disagree slightly vs Agree strongly)	0.22	0.10	0.03	0.42	0.02
Falling behind classmates due to COVID–19 (Disagree strongly vs Agree strongly)	0.24	0.10	0.05	0.43	0.01
Falling behind classmates due to COVID–19 (Not applicable vs Agree strongly)	−0.06	0.16	−0.38	0.25	0.69
Had regular problems with internet connection (Only one lockdown vs Both lockdowns)	0.05	0.10	−0.15	0.26	0.62
Had regular problems with internet connection (Neither lockdown vs Both lockdowns)	−0.17	0.07	−0.30	−0.04	0.01

Table A3. Wave 1 variables included in attrition models used as part of the Wave 2 weight generation

W2 AllYP NPD weight	W2 AllFamily NPD weight	W2 AllYP Full weight	W2 AllFamily Full weight	W2 BoostYP NPD weight	W2 BoostFamily NPD weight	W2 BoostYP Full weight	W2 BoostFamily Full weight	W2 YP with Parent W1orW2 weight	W2 YP with Parent W1orW2 NPD weight
Attended school in person during Year 11 [derived from *ZAtSchool*]									
✓	✓	✓	✓				✓	✓	✓
Attended school in person during the first lockdown due to boarding school/in care/other reason [*ZSCHPERSON1_3, ZSCHPERSON1_4, ZSCHPERSON1_96*]									
✓	✓	✓	✓	✓	✓			✓	✓
Attended school in person during the first lockdown due to school re-opening [*ZSCHPERSON1_5*]									
		✓	✓	✓	✓			✓	✓
Attended school in person during the first lockdown due to special education needs [*ZSCHPERSON1_2*]									
	✓	✓	✓				✓	✓	✓
Attended school in person during the second lockdown due to boarding school/in care/other reason [ZSCHPERSON2_3, ZSCHPERSON2_4, ZSCHPERSON2_5]									
							✓	✓	✓
Attended school in person during the second lockdown due to special education needs [*ZSCHPERSON2_2*]									
							✓	✓	✓
Attending the same school as the one in Year 11 [ZY12School]									
								✓	
Being able to catch up on learning lost due to COVID-19 [*ZCATCHUPCONCERN_03*]									
				✓	✓	✓	✓		
Contact with teachers outside of lessons during the first lockdown [*ZPastcare1*]									
✓	✓	✓	✓					✓	✓
Days per week spent on school work during first lockdown [*ZSchWorkdays1*]									
✓	✓	✓	✓					✓	✓
Effect of COVID-19 on overall motivation to learn [*ZMotivation*]									
						✓	✓		
Falling behind classmates due to COVID-19 [*ZCATCHUPCONCERN_04*]									
✓	✓	✓	✓					✓	✓
Feeling unmotivated [*ZSTUDYISS1_4, ZSTUDYISS2_4*]									
✓	✓	✓	✓	✓	✓			✓	✓
Had regular problems with internet connection [*ZSTUDYISS1_2, ZSTUDYISS2_2*]									
✓	✓	✓	✓					✓	✓
Had to share devices with other family members [*ZSTUDYISS1_5, ZSTUDYISS2_5*]									
✓	✓	✓	✓					✓	✓
Having a job/career in future is important [*ZJOBATT_01*]									
✓	✓	✓	✓			✓	✓	✓	✓
I'll just wait and see where I end up about the future [*ZJOBATT_05*]									
							✓	✓	✓

Question									
If someone is not a success in life, it is usually their own fault [ZSCHOOLATT2_1]									
✓	✓	✓	✓				✓		✓
Likelihood of going to university [ZUniLikely]									
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Main activity [ZMainStat]									
		✓	✓					✓	✓
Mode of interview [ZMODE]									
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
My current school provided additional support to help me catch up on learning lost due to COVID-19 [ZCATCHUPCONCERN_05]									
✓		✓			✓	✓			
My progress during Year 11 has suffered due to COVID-19 [ZCATCHUPCONCERN_01]									
				✓					
Number of data linkage consents given in wave 1 [ZYPCONDFE, ZYPNTP, ZYPCONHEAT, ZYPCONDWP, ZYPCONHMRC]									
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Number of people living in household [ZHHNUM]									
✓	✓	✓	✓					✓	✓
Sample type & incentive [ZBOOST, ZVOUCHER]									
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
School provided real-time online learning for other reasons [ZLIVELESSON1_2, ZLIVELESSON2_2]									
✓	✓		✓	✓	✓	✓	✓	✓	✓
School provided real-time online learning for subjects in the first lockdown [ZLIVELESSON1_1, ZLIVELESSON2_1]									
✓	✓			✓	✓	✓	✓		
Spoke to brother/sister about decisions for future [ZCARADVINF_02]									
✓	✓		✓					✓	✓
Spoke to friends about decisions for future [ZCARADVINF_05]									
						✓	✓	✓	✓
Spoke to parent(s)/guardian(s) about decisions for future [ZCARADVINF_01]									
✓	✓	✓	✓					✓	✓
Year of birth [ZDOBY]									
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Had device to access the Internet [ZDeviceNeed1, ZDeviceNeed2]									
								✓	✓
Likely to do in two years [ZStatus2Y]									
									✓
Received advice from careers advisor for decisions for future [ZCARADV_01]									
								✓	✓

APPENDIX 2 – Weight effectiveness

Table A4. All full households (10,204)

	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)[21]	Final weight (linked to NPD 8912)[22]
FSM eligibility * SEN status	Percent	Percent	Percent	Percent	Percent
FSM last 6 years & EHC plan	1.9	1.2	1.0	1.9	1.9
FSM last 6 years & other SEND status	4.3	6.3	3.5	4.3	4.3
FSM last 6 years & no SEND status	18.4	37.5	18.7	18.4	18.4
No FSM last 6 years & EHC plan	2.1	0.9	1.3	2.1	2.1
No FSM last 6 years & other SEND status	6.7	3.6	5.4	6.7	6.7
No FSM last 6 years & no SEND status	61.1	47.3	69.5	61.1	61.1
Independent in Y11 and Y12	5.6	3.2	0.5	5.6	5.6
Ethnicity					
Indian	2.8	6.1	3.4	2.8	2.8
Bangladeshi	1.7	6.6	2.2	1.7	1.7
Pakistani	4.2	6.0	4.6	4.2	4.2
Black African	3.8	6.0	3.9	3.8	3.8
Black Caribbean	1.2	3.2	0.9	1.2	1.2
White British / no data	65.0	53.8	69.3	65.0	65.0
White non-British	5.8	4.1	5.0	5.8	5.8
Mixed / Other	9.9	11.0	10.1	9.9	9.9
Independent in Y11 and Y12	5.6	3.2	0.5	5.6	5.6
Gender					
Male	48.3	45.0	46.2	48.3	48.3
Female	46.1	51.7	53.2	46.1	46.1
Independent in Y11 and Y12	5.6	3.2	0.5	5.6	5.6
Ethnicity * Gender					
Male White British	33.3	25.1	32.1	33.3	33.3
Male Other	15.0	20.0	14.2	15.0	15.0
Female White British	31.7	28.8	37.2	31.7	31.7
Female Other	14.4	23.0	16.0	14.4	14.4
Independent in Y11 and Y12	5.6	3.2	0.5	5.6	5.6

---

[21] With **W2_AllFamilyFull_weight** applied
[22] With **W2_AllFamily_NPD_weight** applied

Table A4. (continued)

	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)	Final weight (linked to NPD (8912))
KS2 – maths, reading, GPS	Percent	Percent	Percent	Percent	Percent
Upper tertile in all three	14.3	20.7	22.6	14.3	14.3
Upper tertile in two, middle tertile in one	11.8	15.8	15.7	11.8	11.8
Upper tertile in one, middle tertile in two	10.7	11.9	12.5	10.7	10.7
Others with at least one in upper tertile or at least two in middle tertile	24.5	22.8	24.1	24.5	24.5
Lower tertile in two, middle tertile in one	10.3	8.3	8.5	10.3	10.3
Lower tertile in all three	14.1	11.1	10.0	14.1	14.1
Missing data	8.7	6.1	6.0	8.7	8.7
Independent in Y11 and Y12	5.6	3.2	0.5	5.6	5.6
English as an Additional Language					
English is primary language / not recorded	78.4	73.5	83.2	78.4	78.4
English is an additional language	15.9	23.3	16.3	15.9	15.9
Independent in Y11 and Y12	5.6	3.2	0.5	5.6	5.6
School size					
Under 150	22.8	23.7	24.2	22.8	22.8
150–249	53.8	55.4	56.7	53.8	53.8
Over 249	17.8	17.7	18.5	17.8	17.8
Independent in Y11 and Y12	5.6	3.2	0.5	5.6	5.6
School provision					
Special	1.2	0.3	0.6	1.2	1.2
Alternative	0.8	0.6	0.4	0.8	0.8
Selective Other	4.2	5.2	6.2	4.2	4.2
Other	88.1	90.7	92.3	88.1	88.1
Independent in Y11 and Y12	5.6	3.2	0.5	5.6	5.6
School region					
East Midlands	8.2	7.9	8.9	8.2	8.2
East of England	10.6	9.5	11.6	10.6	10.6
London	14.1	20.4	14.5	14.1	14.1
North East	4.4	4.8	4.7	4.4	4.4
North West	13.2	12.6	13.0	13.2	13.2
South East	14.8	12.5	16.1	14.8	14.8
South West	8.8	6.9	9.0	8.8	8.8
West Midlands	10.7	12.7	11.7	10.7	10.7
Yorkshire and the Humber	9.6	9.4	9.8	9.6	9.6
Independent in Y11 and Y12	5.6	3.2	0.5	5.6	5.6

Table A5. All Young People (11,523)

	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)[23]	Final weight (linked to NPD 10,055)[24]
FSM eligibility * SEN status	Percent	Percent	Percent	Percent	Percent
FSM last 6 years & EHC plan	1.9	1.2	1.0	1.9	1.9
FSM last 6 years & other SEND status	4.3	6.1	3.5	4.3	4.3
FSM last 6 years & no SEND status	18.4	37.8	19.0	18.4	18.4
No FSM last 6 years & EHC plan	2.1	0.9	1.3	2.1	2.1
No FSM last 6 years & other SEND status	6.7	3.5	5.4	6.7	6.7
No FSM last 6 years & no SEND status	61.1	46.8	69.2	61.1	61.1
Independent in Y11 and Y12	5.6	3.6	0.6	5.6	5.6
Ethnicity					
Indian	2.8	5.9	3.3	2.8	2.8
Bangladeshi	1.7	6.4	2.2	1.7	1.7
Pakistani	4.2	6.0	4.6	4.2	4.2
Black African	3.8	6.2	4.1	3.8	3.8
Black Caribbean	1.2	3.3	0.9	1.2	1.2
White British / no data	65.0	53.1	68.6	65.0	65.0
White non-British	5.8	4.2	5.1	5.8	5.8
Mixed / Other	9.9	11.2	10.5	9.9	9.9
Independent in Y11 and Y12	5.6	3.6	0.6	5.6	5.6
Gender					
Male	48.3	44.5	45.9	48.3	48.3
Female	46.1	51.9	53.5	46.1	46.1
Independent in Y11 and Y12	5.6	3.6	0.6	5.6	5.6
Ethnicity * Gender					
Male White British	33.3	24.6	31.6	33.3	33.3
Male Other	15.0	19.9	14.3	15.0	15.0
Female White British	31.7	28.5	37.0	31.7	31.7
Female Other	14.4	23.4	16.5	14.4	14.4
Independent in Y11 and Y12	5.6	3.6	0.6	5.6	5.6

---

[23] With **W2_AllYPFull_weight** applied
[24] With **W2_AllYP_NPD_weight** applied

Table A5. (continued)

	Population	Unwtd (all cases)	Design weighted (all cases)	Final weight (all cases)	Final weight (linked to NPD (10,055))
KS2 – maths, reading, GPS	Percent	Percent	Percent	Percent	Percent
Upper tertile in all three	14.3	20.6	22.5	14.3	14.3
Upper tertile in two, middle tertile in one	11.8	16.0	15.9	11.8	11.8
Upper tertile in one, middle tertile in two	10.7	11.7	12.4	10.7	10.7
Others with at least one in upper tertile or at least two in middle tertile	24.5	22.5	23.9	24.5	24.5
Lower tertile in two, middle tertile in one	10.3	8.3	8.5	10.3	10.3
Lower tertile in all three	14.1	11.0	9.9	14.1	14.1
Missing data	8.7	6.2	6.2	8.7	8.7
Independent in Y11 and Y12	5.6	3.6	0.6	5.6	5.6
English as an Additional Language					
English is primary language / not recorded	78.4	73.2	82.9	78.4	78.4
English is an additional language	15.9	23.1	16.5	15.9	15.9
Independent in Y11 and Y12	5.6	3.6	0.6	5.6	5.6
School size					
Under 150	22.8	23.2	23.9	22.8	22.8
150–249	53.8	55.6	57.1	53.8	53.8
Over 249	17.8	17.6	18.4	17.8	17.8
Independent in Y11 and Y12	5.6	3.6	0.6	5.6	5.6
School provision					
Special	1.2	0.3	0.7	1.2	1.2
Alternative	0.8	0.5	0.4	0.8	0.8
Selective Other	4.2	5.3	6.3	4.2	4.2
Other	88.1	90.2	92.0	88.1	88.1
Independent in Y11 and Y12	5.6	3.6	0.6	5.6	5.6
School region					
East Midlands	8.2	7.9	8.8	8.2	8.2
East of England	10.6	9.4	11.5	10.6	10.6
London	14.1	20.5	14.7	14.1	14.1
North East	4.4	4.7	4.6	4.4	4.4
North West	13.2	12.5	13.0	13.2	13.2
South East	14.8	12.7	16.5	14.8	14.8
South West	8.8	7.0	9.0	8.8	8.8
West Midlands	10.7	12.5	11.6	10.7	10.7
Yorkshire and the Humber	9.6	9.2	9.6	9.6	9.6
Independent in Y11 and Y12	5.6	3.6	0.6	5.6	5.6