Understanding Society COVID-19

USER GUIDE

Version 10.0, December 2021

Contents

1. Acknowledgements

data collection, and Hannah Carpenter and her team at Kantar for implementing the COVID-19 telephone surveys and the November 2020 youth survey.

2. Overview

The *Understanding Society* COVID-19 study is a panel study on the experiences and reactions of the UK population to the COVID-19 pandemic. The survey is an integral part of *Understanding Society*: the UK Household Longitudinal Study. Researchers can link the data from this survey to answers respondents have given in previous (and future) waves of the annual *Understanding Society* survey (see Section 21 for more details). The first wave of the COVID-19 survey was fielded in April 2020, with monthly waves until July 2020. From September 2020 onwards the survey was fielded every two months until March 2021. A final wave was fielded in September 2021. This study was funded by the Economic and Social Research Council and the Health Foundation, and the data are available to researchers from the UK Data Service (SN8644).

The 20 minute questionnaire includes core content repeated at each data collection to track changes through the pandemic, as well as rotating content. For the May 2020 survey onwards, researchers and analysts were invited to suggest questionnaire content through an open call. In selecting content for the questionnaires, priority was given to content that:

- Focuses on *Understanding Society* core priority areas: employment, income, health, family, education, civic engagement;
- Builds on the longitudinal strengths of *Understanding Society*, and links to data we have collected before or will collect in the future;
- Needs to be collected "in the moment" rather than being asked retrospectively in a future annual survey;
- Is enhanced by the household structure of the Study;
- Is of significant research and/or policy interest;
- Complements existing *Understanding Society* content.

The questionnaire was implemented as a web survey, fielded by Ipsos MORI. In some months there was an additional telephone survey for households without internet access, fielded by Kantar; in some months there was a paper self-completion survey sent to 10-15 year olds.

This User Guide was updated every wave for the duration of the COVID-19 study, to include information about the most recent data collected.

3. How to cite the data and User Guide

The bibliographic citation for this user guide is:

> Institute for Social and Economic Research (2021) *Understanding Society COVID-19 User Guide*. Version 10.0, October 2021. Colchester: University of Essex.

The bibliographic citation for the main *Understanding Society* COVID-19 data is the following. This includes the data from the 2019 annual *Understanding Society* interviews:

> University of Essex, Institute for Social and Economic Research. (2021). Understanding Society: COVID-19 Study, 2020-2021. [data collection]. 11th Edition. UK Data Service. SN: 8644, 10.5255/UKDA-SN-8644-11.

The bibliographic citation for the LSOA11 *Understanding Society* COVID-19 data is:

> University of Essex, Institute for Social and Economic Research. (2021). Understanding Society: COVID-19 Study, 2020-2021: Special Licence Access, Census 2011 Lower Layer Super Output Areas. [data collection]. 7th Edition. UK Data Service. SN: 8663, 10.5255/UKDA-SN-8663-7.

The bibliographic citation for the Local Authority District *Understanding Society* COVID-19 data is:

> University of Essex, Institute for Social and Economic Research. (2021). Understanding Society: COVID-19 Study, 2020-2021: Special Licence Access, Local Authority District. [data collection]. 7th Edition. UK Data Service. SN: 8664, 10.5255/UKDA-SN-8664-7.

The bibliographic citation for the COVID-19 School Codes data is:

> University of Essex, Institute for Social and Economic Research. (2021). Understanding Society: COVID-19 Study, 2020: Special Licence Access, School Codes. [data collection]. 2nd Edition. UK Data Service. SN: 8730, 10.5255/UKDA-SN- 8730-2.

4. Data Access

The data from the *Understanding Society* COVID-19 study are available from the UK Data Service. The End User Licence (EUL) version, SN 8644, can be found here: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8644. The EUL version does not include some variables such as month of birth, and only includes the top-coded versions for some variables such as income, earnings and household composition (see Section 11.5).

Datasets containing the *Understanding Society* COVID-19 geographical identifiers LSOA 2011 (SN 8663), Local Authority District (SN 8664), and the School Codes (SN 8730), are also available from the UK Data Service. They can be found here:
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8663
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8664
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8730

These datasets are categorized as Special Licence and therefore have more restrictive access conditions, details of which can be found in the "Access data" tab at the above links. In addition, access to these Special Licence datasets requires users to complete forms justifying why the data is needed. Further details on the application process can be found here: https://ukdataservice.ac.uk/get-data/how-to-access/conditions/special-licence.aspx.

5. Sample and following rules

The COVID-19 study includes all members of the main *Understanding Society* samples: the *Understanding Society* General Population Sample (GPS), the Ethnic Minority Boost Sample (EMBS), the Immigrant and Ethnic Minority Boost Sample (IEMBS), and the former British Household Panel Study (BHPS). The eligible sample includes everyone in households who participated in waves 8 or 9.

These samples are probability samples of postal addresses. In England, Wales and Scotland they are clustered and stratified, in Northern Ireland unclustered systematic random samples. Northern Ireland and areas in Great Britain with high immigrant and ethnic minority populations were oversampled. For more details on the sample design see the *Understanding Society* User Guides (Institute for Social and Economic Research 2020), Lynn (2009), Berthoud, Fumagalli et al. (2009) and Lynn, Nandi et al. (2017).

In eligible sample households, all household members who were aged 16+ in April 2020 were invited to the COVID-19 study, except those who were adamant refusals or mentally or physically unable to make an informed decision to take part, and those with unknown postal addresses or addresses abroad.[1]

In waves 2 to 4, everyone eligible in wave 1 continued to be eligible irrespective of whether they had participated in any of the previous surveys. From the September 2020 (wave 5) survey onwards, only sample members who had completed at least one partial interview in any of the first four web surveys were invited to participate. From the November 2020 (wave 6) survey onwards, we stopped inviting those who had only completed the initial survey in April 2020 and none since.

---

[1] All individuals in the sample households are the core sample members (referred to as Original Sample Members, OSMs). Anyone who joins these households after the initial wave are referred to as Temporary Sample Members (TSMs) and are only eligible for interviews if they are living with at least one OSMs. Children of women OSMs are also OSMs. TSMs who are fathers of OSM children become Permanent Sample Members, PSMs and from that point onwards have the same eligibility conditions as OSMs. The variable w_SAMPST is updated each wave to reflect this sample status. This variable will not be updated during the COVID-19 study, but will be updated at the next annual interview.

6. Fieldwork

At the start of the COVID-19 study, 42,330 sample members were sent the prenotification letter inviting them to the study. Of those invited, 124 were later identified as deceased and therefore ineligible.

The prenotification letters, explaining the purpose of the COVID-19 study, what respondents were asked to do, and how they would be rewarded for participating, were sent out by post on 17 April 2020. The letter included a short FAQ section and a link to more detailed FAQs on the participant web page.[2] Respondents were offered £2 for every survey they completed and could exchange their rewards for a range of gift-cards and electronic vouchers. From the May 2020 web survey onwards respondent could also donate their incentive to NHS Charities Together. Respondents were told that the invitations to the survey would be sent out by email and/or SMS text message, depending on which contact information we held about them. They were shown a scrambled version of the mobile number and email address we hold and asked to update their contact details on the participant website if necessary. A significant number of respondents did this. Some sample members (N=479) opted out of the COVID-19 survey on receipt of the prenotification letter and were therefore not sent invitations to the surveys.

Table 6.1 summarizes the fieldwork dates and number of respondents who were sent invitations and reminders by email, SMS, postal letters, or combinations thereof, for each of the web surveys. Respondents for whom no mobile number or email address were known were sent an invitation letter by post. The email and SMS invitations included a personalized link to the survey: respondents could click the link and start their survey, without the need for a username and password. The postal invitations included the link to the survey, username and password, as well as an information leaflet including screenshots to support sample members with how to access and complete the survey. For the web surveys, respondents were sent an invitation to the survey and three subsequent reminders (on days 2, 3 and 6 of the 7-day fieldwork period).

Respondents who dropped out of the COVID-19 surveys before reaching the end are classified as non-respondents, partial, or full respondents, depending on where they dropped out. In all waves, respondents who dropped out before the end of the coronavirus illness module are classified as non-respondents and their data are not released. The cut-off point for classifying respondents as partial or full respondents varies depending on the content of the questionnaire and is summarized in Table 6.2. In the April 2020 survey for example, respondents who got past the coronavirus illness module but dropped out before the end of the finance module were classified as partial respondents; those who dropped out after this point were classified as full respondents.

---

[2] https://www.understandingsociety.ac.uk/participants/coronastudy.

6.1 Wave 1 web survey

In the April 2020 web survey, there were 433 full and 49 partial respondents, who entered a date of birth or sex that did not match the information we hold on the sample file. Some of these mismatches were due to household members sharing mobile phones: they received one SMS message including the personalized link to the survey for every person associated with that number in our sample file. Due to the character limit, the SMS messages do not include the person's name – unlike the emails which are personally addressed to the recipient using their first and last name. Respondents whose identity we were able to confirm (n=309) have been added back into the April 2020 data, from the 4th edition of the COVID-19 data onwards. These respondents are identified by the variable "ca_pidpcorrected" in the ca_indresp_w file and the xsample file. In the May 2020 survey these sample members were treated as though they had not completed the April survey. That is, they will have been asked the 'baseline questions', asked only the first time a respondent completes the COVID-19 survey, a second time.

Including the mismatched cases whose identity we were able to confirm, 42.1% of the full sample completed the COVID-19 survey (Table 6.3).

6.2 Wave 2 web survey

For the May 2020 web survey fewer invitations were sent out than in April (see Table 6.1). This is because respondents who opted out of the COVID-19 study before the wave 2 fieldwork were no longer sent an invitation, and neither were those who had been identified as ineligible (moved abroad or deceased), or where the contact details turned out to be invalid. Further, sample members who share a mobile phone number were no longer sent invitations by SMS, but by email or post only. In addition, 3,432 sample members who had not completed the April web survey and who lived in households where no-one is a regular internet user (according to prior interview data), were invited to a telephone survey and not issued to the web survey. Some of those issued to the telephone survey were later identified as deceased. Tables 6.3 and 6.4. are based on eligible sample members and for each wave exclude those who have become ineligible.

6.3 Wave 2 telephone survey

The first COVID-19 telephone survey ran in parallel to the May 2020 (wave 2) web survey, although it was based on the April questionnaire. Letters were sent to 3,432 sample members who had not completed the April web survey and who lived in households where no-one is a regular internet user on the 22 May 2020. These sample members would have already received the initial pre-notification letter sent in mid-April, and so the May letter contained less information about the study. The letter informed sample members that we were aware that launching the online survey excluded people who did not have access to the internet at home, or were not regular web users. Since this was an important group in

society, we were inviting them to take part in a telephone survey. The letter was also translated into Welsh for those who lived in Wales, and included an unconditional £2 Love2Shop gift card. Fieldwork for the telephone survey started on 28 May 2020, and finished on 7 June 2020. The questionnaire for the telephone survey was slightly different to the April web survey: many of the questions required interviewer instructions, and the wording was slightly altered so that it made sense as a telephone survey, rather than an online interview where the respondent completed the survey themselves.

6.4 Wave 3 web survey

The June 2020 web survey excluded the sample members who had been issued to the May telephone survey – except for 130 telephone respondents who indicated that they would like to do the web surveys. The decision was also made to stop sending invitation letters by post, to sample members for whom no email or mobile number were known and who had not taken part in the April or May web surveys. The letters were useful initially. In April, 164 of the 5,847 respondents who were invited by postal letter completed the survey and some provided an email address or mobile number within the questionnaire, that was used to send out invitations to later surveys. In the May survey, however, only 8 of the 4,855 sample members invited by post completed the survey. The decision was therefore made to stop sending invitation letters by post – except for 133 sample members who had completed at least one previous COVID-19 web survey.

6.5 Wave 4 web survey

The July 2020 web survey included an experiment with different respondent incentives: the control group was offered the standard COVID-19 study incentive payment of £2 for completion of the interview; the treated group was offered £2 plus a one-off £10 bonus for participating in the study. The experiment was implemented on a subset of the wave 4 issued sample. Participants excluded from the experiment were offered the standard £2 incentive payment.

Of individuals issued to wave 4, 32,533 were selected to be included in the experiment. Inclusion was restricted to those who had completed the wave 9 annual interview and had a non-zero cross-sectional weight, as this was initially used as the base for the COVID-19 study weights.

80 percent of individuals included in the experiment were allocated to the control group and so were offered the standard incentive payment; the remaining 20 percent were offered the larger incentive payment. The randomisation occurred at the address, rather than the individual, level.

The variables pertaining to the experiment are "cd_ff_incentbonus" and "cd_incentive", which are included on the xsample file. The variable "cd_ff_incentbonus" contains the

experimental allocations and "cd_incentive" contains the incentives respondents received, irrespective of whether they formed part of the randomisation set.

Due to the incentive experiment, postal invitation letters were sent to all sample members for whom we did not have a valid email address. This included sample members for whom we had neither an email nor a mobile number, and sample members for whom we had only a mobile number (Table 6.1).

6.6 Wave 4 youth survey

At the end of July, a paper youth survey – for children aged 10-15 years in households - was issued with the 25 item Strengths and Difficulties Questionnaire (SDQ). It was sent to the subsample of those issued in the July COVID-19 web survey (N=4,139) who were the responsible parent for children of this age. Being the 'responsible adult' for the child is derived from the household grid in the last annual interview. Hence only one parent was sent the children's questionnaire for each child, but parents may have been sent it for multiple children. The first mailing to the 'responsible adult' was sent out on 29 July 2020, with a reminder sent to those from whom we had had no reply by 17 August; fieldwork closed on 1 September 2020. A self-addressed envelope was enclosed for them to return the completed questionnaires and a pen as a small token of thanks. We received 1,435 responses in total, but 24 of these were duplicates and hence there were 1,411 unique responses.

6.7 Wave 5 web survey

For the September 2020 survey onwards we only invited sample members who had completed at least one partial interview in any of the preceding web surveys. This excluded sample members who had since become ineligible (died, moved abroad) or opted out of the study.

6.8 Wave 6 web survey

The November 2020 web survey invitation went to all sample members who had been invited to the study in the previous wave. This excluded sample members who had become ineligible (died, moved abroad), those who had opted out of the study since the previous wave, and those who had completed only the April wave of the study and not taken part since.

6.9 Wave 6 telephone survey

All sample members who completed the May 2020 telephone survey were invited to the second telephone survey in November 2020. This included respondents who dropped out of the May survey before the end of the coronavirus illness module, and were therefore classified as non-respondents and their data are not released. The invited sample excluded sample members who had transferred to the web survey after the first telephone survey.

Fieldwork for the telephone survey started on 23 November and finished on 5 December 2020.

6.10 Wave 6 youth survey

In November 2020 a paper youth survey was sent out to children aged 10-15. It was sent to the subsample of those issued in the July COVID-19 web survey (N= 4,295) who were the responsible parent for children of this age. Being the 'responsible adult' for the child is derived from the household grid in the last annual interview. Hence only one parent was sent the children's questionnaire for each child, but parents may have been sent it for multiple children. The first mailing to the 'responsible adult' was sent out on 16th November 2020, with a reminder sent to those from whom we had had no reply by 2nd December; fieldwork closed on 18th December, although a few further questionnaires were received up until 31st December. A self-addressed envelope was enclosed for them to return the completed questionnaires and a pen as a small token of thanks. We received 1,432 unique responses.

6.11 Wave 7 web survey

The January 2021 web survey invitation again went to all sample members who had been invited to the study in the previous wave. This excluded sample members who had become ineligible (died, moved abroad) and those who had opted out of the study since the previous wave.

6.12 Wave 8 web survey

The rules for issuing the March 2021 web survey mirrored those for the wave 7 survey. In addition to the usual questionnaire modules, the wave 8 survey included COVID-19 antibody testing (see Section 6.14 below), consent to NHS and Registry data linkage, a second respondent incentive experiment, and an experiment with the position and order of the consent questions.

The incentive experiment included only those sample members who had been eligible for the wave 4 (July 2020) incentive experiment. The ineligible cases were offered the usual £2 incentive conditional on completing the survey. The eligible cases were split into six groups: three groups were promised an incentive if they completed the survey (values £7, £12, £17); three groups were promised an incentive if they completed the survey (values £2, £7, £12), plus a £5 incentive if they returned the serology test kit. Randomisation was performed at the household level and households allocated to the groups in the following proportions, respectively: ¼; 1/8; 1/8; ¼; 1/8; 1/8. Randomisation was implemented by block, where blocks were defined by the wave 4 (July 2020) experimental allocation. The variable "ch_ff_incentw8" in the xsample file contains the experimental allocations.

For the COVID-19 antibody testing and consent questions, respondents were randomised to one of seven equally sized groups, varying the position and order of these questions in the survey. The treatment groups are explained in the scripting notes of the March 2021 questionnaire. The variable "ch_ff_consentpos" in the xsample file contains the experimental allocations. Randomisation was performed at the household level and done by block, where blocks were defined by the wave 8 incentive experiment allocations.

Due to the incentive experiment, postal invitation letters were sent to all sample members for whom we did not have a valid email address. This included sample members for whom we had neither an email nor a mobile number, and sample members for whom we had only a mobile number (Table 6.1).

6.13 Wave 8 youth survey

In March 2021 another paper self-completion survey was sent out to children aged 10-16. The previous COVID-19 youth surveys were sent to 10-15-year-olds. For the March 2021 survey the age was extended to 16, to pick up those who had turned 16 since the first youth COVID-19 survey in July 2020. The survey was sent to the subsample of those issued in the July COVID-19 web survey (N= 3,439) who were responsible for children of this age. Being the 'responsible adult' for the child is derived from the household grid in the last annual interview. Hence only one parent was sent the children's questionnaire for each child, but parents may have been sent it for multiple children. The first mailing to the 'responsible adult' was sent out on 1st March 2021, with a reminder sent to those from whom we had had no reply by 18th March; fieldwork closed on 9th April. A self-addressed envelope was enclosed for them to return the completed questionnaires and a pen as a small token of thanks. We received 1,388 unique responses.

6.14 Wave 8 serology collection

As mentioned in Section 6.12 above, the March 2021 survey included COVID-19 antibody testing. Prior to the survey launch, the laboratory involved in the testing – Thriva – sent the pre-packaged kits to Ipsos MORI. During the web survey, participants were given information about the antibody testing process and asked whether we could send them the testing kit through the post. There was no telephone survey in March 2021, and so participants from the May 2020 telephone survey[3] were sent a letter with information about the serology collection, including a form that the sample member could return to indicate whether or not they would like to have a testing kit sent to them. After the web survey closed, Ipsos MORI scanned the Thriva bar code that was on the outer packaging of the kit and associated this with the participant identifier for those who had agreed to receive the kit during the web interview. The kits were then despatched to the participants.

---

[3] Note only those who had participated in the May 2020 telephone survey were eligible for the November 2020 telephone survey, and eligibility for this serology collection was not contingent on responding to the November survey as well.

Kits were despatched to telephone sample members when Ipsos MORI received the returned forms. The kits contained all the necessary equipment to allow the participant to take a small blood sample, including instructions. The kits had been pre-packaged and so whilst there was a cover letter in the kit, this was not personalised. A copy of the letter and instructions are in Appendix C. The instructions guided the participant on how to use a lancet to collect a blood sample in a small tube. The participant then put the tube back into the box it came in and placed the box into an envelope that was provided and posted it back to Thriva. Thriva received the blood samples and analysed them for COVID-19 antibodies. Each Monday morning, Thriva uploaded a file of results to a secure portal. Ipsos MORI downloaded the results file, and linking the Thriva identifier with the participant identifier, sent out the results letters to participants who had requested feedback. The results that were sent to participants indicated whether blood antibodies were clearly detectable, not clearly detectable, or whether the test had been invalid. See Section 16 for the contents of the serology data file.

6.15 Wave 9 web survey

The final web survey was fielded in September 2021. The invitations were again sent to all sample members who had been invited to the study in the previous wave. This excluded sample members who had become ineligible (died, moved abroad) and those who had opted out of the study since the previous wave.

As a thank you for participating in the COVID-19 study and to mark the final wave, all respondents were offered £10 for completing the survey. Invitation letters were sent to all sample members for whom we did not have an email address (see Table 6.1).

Table 6.1: Web survey modes of invitations and reminders

Wave: Fieldwork dates	Email & SMS	Email not SMS	SMS not Email	Letter only	Letter & Email/SMS	Total
W1: 24-30/04/2020	27,560	4,897	3,620	5,769	78	41,846[*]
W2: 27/05-02/06/2020	24,241	5,036	2,679	4,855	0	36,811
W3: 25/06-01/07/2020	24,271	5,088	2,497	133	0	31,989
W4: 24/07-31/07/2020	24,067	4,917	2,478	4,806	2,466	36,268
W5: 24/09-01/10/2020	15,815	2,917	419	236	0	19,387
W6: 24/11-01/12/2020	15,746	2,887	429	232	0	19,294
W7: 27/01-03/02/2021	15,683	2,856	425	233	0	19,197
W8: 24/03-31/03/2021	15,601	2,816	429	229	423	19,075
W9: 23/09-01/10/2021	15,512	2,754	413	229	411	18,908

Notes: [*] excludes 484 sample members who opted out of the COVID-19 survey after receiving the pre-notification letter.

Table 6.2 Cut-off point in questionnaire for classification of partial and full respondents

Survey	Partial interview marker – at end of
April 2020 web/May 2020 tel	Finance module
May 2020 web	Finance module
June 2020 web	Employment module
July 2020 web	Finance module
Sept 2020 web	Employment module
Nov 2020 web/tel	Finance module
Jan 2021 web	Employment module
March 2021 web	Finance module
Sept 2021 web	Finance module

Table 6.3: Fieldwork outcomes: web surveys

Wave	Survey outcome	N	%
W1 web	Full interview	16,662	39.5
	Partial interview	1,099	2.6
	Full or partial interview	*17,761*	*42.1*
	Non-respondent	24,446	57.9
	Total	42,207	100.0
W2 web	Full interview	14,607	34.6
	Partial interview	204	0.5
	Full or partial interview	*14,811*	*35.1*
	Non-respondent	21,955	52.1
	Opted out - not issued	2,024	4.8
	No contact details - not issued	6	0.0
	Transferred to tel survey - not issued	3,386	8.0
	Total	42,182	100.0
W3 web	Full interview	13,917	33.0
	Partial interview	206	0.5
	Full or partial interview	*14,123*	*33.5*
	Non-respondent	17,841	42.3
	Opted out - not issued	2,294	5.4
	No contact details - not issued	27	0.1
	Transferred to tel survey - not issued	3,247	7.7
	No email or mobile info - not issued	4,639	11.0
	Total	42,171	100.0
W4 web	Full interview	13,577	32.3
	Partial interview	177	0.4
	Full or partial interview	*13,754*	*32.6*
	Non-respondent	22,419	53.3
	Opted out - not issued	2,654	6.3
	No contact details - not issued	29	0.1
	Transferred to tel survey - not issued	3,246	7.7
	Total	42,102	100.0
W5 web	Full interview	12,696	30.2
	Partial interview	180	0.4
	Full or partial interview	*12,876*	*30.6*
	Non-respondent	6,496	15.4
	Opted out - not issued	3,016	7.2
	No contact details - not issued	29	0.1
	Transferred to tel survey - not issued	3,242	7.7
	No email or mobile info - not issued	4,575	10.9
	Non-respondent in all waves – not issued	11,846	28.2
	Total	42,080	100.0

Wave	Survey outcome	N	%
W6 web	Full interview	11,802	28.1
	Partial interview	233	0.6
	Full or partial interview	*12,035*	*28.6*
	Non-respondent	7,245	17.2
	Opted out - not issued	3,224	7.7
	No contact details - not issued	39	0.1
	Transferred to tel survey - not issued	3,227	7.7
	No email or mobile info - not issued	4,503	10.7
	Non-respondent in all waves - not issued	11,757	28.0
	Total	42,030	100.0
W7 web	Full interview	11,797	28.1
	Partial interview	171	0.4
	Full or partial interview	*11,968*	*28.5*
	Non-respondent	7,137	17.0
	Opted out - not issued	3,432	8.2
	No contact details - not issued	39	0.1
	Transferred to tel survey - not issued	3,216	7.7
	No email or mobile info - not issued	4,488	10.7
	Non-respondent in all waves - not issued	11,724	27.9
	Total	42,004	100.0
W8 web	Full interview	12,509	29.8
	Partial interview	171	0.4
	Full or partial interview	*12,680*	*30.2*
	Non-respondent	6,394	15.2
	Opted out - not issued	3,630	8.7
	No contact details - not issued	60	0.1
	Transferred to tel survey - not issued	3,206	7.6
	No email or mobile info - not issued	4,395	10.5
	Non-respondent in all waves - not issued	11,602	27.7
	Total	41,967	100.0
W9 web	Full interview	12,572	30.1
	Partial interview	246	0.6
	Full or partial interview	*12,818*	*30.6*
	Non-respondent	6,078	14.5
	Opted out - not issued	4,047	9.7
	No contact details - not issued	90	0.2
	Transferred to tel survey - not issued	3,172	7.6
	No email or mobile info - not issued	4,236	10.1
	Non-respondent in all waves - not issued	11,390	27.2
	Total	41,831	100.0

Notes: 'Total' number of eligible cases decreases across waves due to deaths and moving abroad.

Table 6.4: Fieldwork outcomes: telephone survey

Wave	Survey outcome	N	%
W2 tel	Full interview	650	19.2
	Partial interview	68	2.0
	Full or partial interview	*718*	*21.1*
	No contact made	1,143	33.8
	Contact made but no further contact	900	26.6
	Dead number	98	2.9
	Refusal	317	9.4
	Unknown at number	108	3.2
	Moved	20	0.6
	Incapable of Interview	82	2.4
	Total	3,386	100.0
W6 tel	Full interview	366	51.1
	Partial interview	25	3.5
	Full or partial interview	*391*	*54.4*
	No contact made	42	5.9
	Contact made but no further contact	156	21.8
	Dead number	30	4.2
	Refusal	93	13.0
	Unknown at number	5	0.7
	Total	717	100.0

7. Questionnaire content

The COVID-19 questionnaires include core content, repeated to track changes, rotating content carried intermittently, and some content carried only once. Table 7.1 summarizes which questionnaire modules were carried in which surveys.

Note that core modules have evolved over the first few waves, with additional questions or response options being added to some. For each variable, the online data documentation (Variable Search) documents in which waves of the COVID-19 survey a question was asked. Viewing the response distributions of a variable for each wave in which the question was asked will illustrate any differences in the response options. See Section 9 on the online data documentation.

From May 2020 onwards, sample members living in Wales were given the choice whether to complete the survey in Welsh or English. Only the core modules, marked with an asterisk in Table 7.1, are translated into Welsh.

Table 7.1: Questionnaire content by month fielded

Topic	04/20w 05/20t	05/20w	06/20w	07/20w	09/20w	11/20w 11/20t	01/21w	03/21w	09/21w
ID check and HH composition*	x	x	x	x	x	x	x	x	x
Coronavirus illness*	x	x	x	x	x	x	x	x	x
Long-term health conditions*	x	x	x	x	x	x	x	x	x
Loneliness*	x	x	x	x	x	x	x	x	x
Employment*	x	x	x	x	x	x	x	x	x
Finance*	x	x		x		x		x	x
Financial security*	x	x		x		x		x	x
Mental health (GHQ)*	x	x	x	x	x	x	x	x	x
Caring outside household	x					x	x	x	x
Time use	x								
(Home) schooling	x				x	x	x		
Food	x								
Alcohol consumption	x				x		x		
Smoking	x				x		x		
Exercise	x				x		x		
HH relationships		x	x	x	x	x	x	x	x
Housing		x		x	x	x	x	x	x
Partnership changes		x	x						
Couples living apart together		x							
Time use		x							
Parents and children		x			x		x		
Life satisfaction		x		x	x	x	x	x	x
Household coronavirus illness			x			x			
Work conditions			x		x		x		x
Travel to work			x			x		x	x
Job search			x		x		x		x

Topic	04/20w 05/20t	05/20w	06/20w	07/20w	09/20w	11/20w 11/20t	01/21w	03/21w	09/21w
Training			x		x				x
Partner			x		x		x		
Domestic division of labour			x		x		x		
Non-resident children			x						
Non-resident parents			x						
Grandparenting			x						
Contact with friends/family outside HH			x			x			
Neighbourhood cohesion			x			x		x	
Transport			x			x		x	
Caring within household				x		x	x	x	x
Strengths and difficulties				x	x			x	
Special educational needs				x					
Religion				x					
Volunteering				x				x	
Sleep				x					
Nutrition				x			x		
Diet and food security				x	x		x		
Fertility intentions					x				x
Young adults					x				x
Internet use					x				
Self-assessed health						x	x	x	x
Flu and coronavirus vaccine						x			
Coronavirus vaccine							x	x	x
Smartphone							x		
Attitudes to immigration							x		
Health linkage and serology consent								x	
Survey device and incentives								x	
NHS app									x

Topic	04/20w 05/20t	05/20w	06/20w	07/20w	09/20w	11/20w 11/20t	01/21w	03/21w	09/21w
Work expectations									x

Notes: w = web survey, t = telephone survey. ***** = modules translated into Welsh from May 2020 onwards.

8. How to read the questionnaire

For each question the COVID-19 questionnaires document the variable name, who was asked the question ("Universe"), the source of the question, and the wording of the question text and response options. Figure 8.1 provides an example to illustrate the questionnaire specification and how this relates to the variables in the data.

Figure 8.1 Example question specification

testresult [Result of coronavirus test]
Universe: IF tested = 1 // Ask if tested for coronavirus.
Source: UKHLS covid-19 survey
Text: What was the result of your coronavirus test?
1. Positive - it showed I had coronavirus
2. Negative - it showed I did not have coronavirus
3. Inconclusive
4. Waiting for results

In the data, the variable corresponding to the question in Figure 8.1 is "testresult". The label for that variable is "Result of coronavirus test", and its values (1 to 4) are labelled according to the response options in the questionnaire specification.

The Universe specifies who was eligible for this question. In this case, all respondents for whom the value of a previous question ("tested") is 1. The Source indicates that this is a question that was written for the COVID-19 study. For questions that were taken from the annual *Understanding Society* interviews or from other surveys, the Source documents the name of that survey and whether the question is a direct copy or adapted.

For the telephone survey the questions have been adapted so that they are suitable for telephone interviewing. For example the wording has changed slightly, to reflect the fact that the interviewer is asking the question, and interviewer instructions have been added where necessary.

9. Online data documentation

The online data documentation for the COVID-19 study is available at
https://www.understandingsociety.ac.uk/documentation/covid-19. The documentation
includes the most recent versions of this User Guide, the questionnaires, the fieldwork
documents (invitations, reminders, etc. sent to sample members) and a variable search
facility covering the content from the COVID-19 surveys. You can search by variable name,
by data file or by question module. For each variable, the documentation shows in which
data file the variable can be found, in which waves it was asked, in which questionnaire
module, the question wording, and the response distributions.

10. Data structure

10.1 Data files

The COVID-19 data include three types of files released in every wave:

- Sample file: a cumulative data set that includes observations about all sample
  members who were eligible for the study in April 2020. The file was updated with
  each data release to include information about the outcomes of all surveys to date.
  This is an individual level file with each row identified uniquely by "pidp", the unique
  cross-wave individual identifier in the annual survey.
- Baseline file: a cumulative data set that includes the answers to questions that were
  asked only the first time a sample member completed a full COVID-19 survey. This is
  an individual level file with each row identified uniquely by "pidp".
- Survey files: one file for each month's survey, with separate files for the web and
  telephone surveys. These are individual level files with each row identified uniquely
  by "pidp".

For some months there are additional files, depending on the content of the questionnaire:

- School child-level files: the April, July, September, November 2020, January 2021 and
  March 2021 questionnaires included modules that asked about (home) schooling,
  special education needs, or the Strengths and Difficulties Questionnaire (SDQ). These
  questions were asked about each child in turn. For each of these waves, there is a
  child-level file, which includes one row for each child, identified uniquely by the
  child's "pidp", as well as the "pidp"s of the child's parents/guardians.
- Youth file: in July 2020, November 2020 and March 2021 we sent out paper self-
  completion questionnaires to our 11-15/16 year-old sample members. The youth
  files include one row for each child, identified uniquely by the child's "pidp".
- Serology file: In March 2021 participants were asked to send a blood sample so it
  could be tested for COVID-19 antibodies (see Section 6.14 above). The Serology file

includes the results of these antibody tests and has a row for each respondent who was invited to the serology testing, identified uniquely by the respondent's "pidp".

10.2 File naming conventions

The file names for the survey data have a prefix that includes the letter "c" to reflect this is the COVID-19 survey rather than the annual survey, and a second letter to denote the wave of the survey (ca_, cb_, etc). The COVID-19 study included nine waves (a=April 2020, b=May 2020, c=June 2020, d= July 2020, e=September 2020 , f=November 2020, g=January 2021, h=March 2021, i=September 2021). The file names also include a suffix that denotes whether the data are from the web (_w), the telephone survey (_t), or the paper youth self-completion questionnaire (_p).

The file containing the questionnaire data from the April 2020 web survey is therefore called "ca_indresp_w". The file containing the questionnaire data from the May 2020 telephone survey is called "cb_indresp_t" – although the content of the May 2020 telephone survey is the same as the April 2020 web survey. See Table 7.1 above.

As the sample file is a cross-wave file the name does not include a wave prefix. Instead it includes the letter "x" to reflect that it contains data from across the waves: "xsample". Similarly the file "xbaseline" includes data about a pre-pandemic baseline in January/February 2020 collected in the first four waves.

10.3 Summary of data files

Table 10.1 is a summary of the different waves of data collection for the COVID-19 survey. For each wave the table lists the calendar month of the survey, the wave number, the prefix used to denote that wave, and the data files available for that wave.

Table 10.1: Summary of survey waves and data files

Date of survey	Wave	Wave prefix	INDRESP_W	INDRESP_T	SCHILD_W	SCHILD_T	YOUTH_P
April 2020	1	a	X		X		
May 2020	2	b	X	X		X	
June 2020	3	c	X				
July 2020	4	d	X		X		X
Sept 2020	5	e	X		X		
Nov 2020	6	f	X	X	X	X	X
Jan 2021	7	g	X		X		
March 2021	8	h	X		X		X
Sept 2021	9	i	X				

10.4 Variable naming conventions

Variables are named according to the questionnaire specification described in Section 8. For questions that are taken from the annual *Understanding Society* questionnaires, the original *Understanding Society* variable names are used. If the *Understanding Society* question has been modified, then the original variable name is used, with an additional "_cv" suffix to flag up that it is not identical to the original question. Modifications of *Understanding Society* questions include: changing the reference period to ask about a month rather than a year, adding response categories, or changing the question wording so that the question can be presented as an individual question rather than a grid of questions. Where questions have been modified, the Source information in the questionnaire documents what was changed.

Derived variables include the suffix "_dv" to indicate that they are not directly related to questions in the questionnaire, but have been derived from answers given.

As with data files, the variable names include a wave prefix to reflect the wave in which they were collected. For example, the variable "testresult" collected in the first month's survey is named "ca_testresult".

Variables with prefixes "i_" to "k_" are from the annual interview data from waves 9 to 11.

Variables without prefixes are stable characteristics copied across from the annual interview data.

10.5 Naming of variables from multicode questions

For some questions respondents are asked to *"Please select all that apply"* from a list of response options. For such multicode questions, the data files include one variable for each response option, with the values 0 "not mentioned" and 1 "mentioned". These binary indicators are named according to the question name documented in the questionnaire, followed by the number that corresponds to the response option. As an example, for the question showcased in Figure 10.1, the responses are recorded in the variables "carewho1", "carewho2", …, "carewho8".

Figure 10.2: Example multicode question

---

carewho [Who caring for outside the household]

Universe: IF caring = 1 // Ask if gives care to someone outside household.

Source: UKHLS covid-19 survey

Who did you help?

Please select all that apply.

1. Adult children, including in-laws
2. Parents or grandparents, including in-laws
3. Siblings
4. Spouse or partner
5. Former spouse or partner
6. Friends
7. Neighbours
8. Someone else

10.6 Missing values

Missing observations are recorded in the same way as in the annual *Understanding Society* interview data, using negative values rather than system missings. In the web survey, respondents are initially only shown the substantive response options. If they click "Next" without selecting a response option, they are shown response options for "Don't know" and "Prefer not to say". In the telephone survey interviewers record if respondents spontaneously say "Don't know" or "Prefer not to say". Table 10.2 documents the values used to record the reason why the answer to a question is missing. For respondents who broke off the survey before reaching the end, all questions that were not asked for this reason are coded as "Ineligible". The variable "cW_lastq" documents the last question the respondent answered.

Table 10.3 Missing value codes

Value	Label	Description
-1	Don't know	Respondent reports they "Don't know"
-2	Refusal	Respondent reports they "Prefer not to say"
-8	Inapplicable	Respondent is not asked the question due to routing or because they broke off the survey prior to this question
-9	Missing	Respondent still does not provide an answer after being shown "Don't know"/ "Prefer not to say"

10.7 How to link data across months and to the annual interview data from the main survey

The personal identifier ("pidp") is included in all data files and can be used to link data from within one month (for example, linking the sample and questionnaire data), link data across months (for example, linking the questionnaire data from multiple months).

The variable "pidp" can also be used to link individual level data from the COVID-19 study to individual level data from the annual *Understanding Society* interviews. The household identifier variables, "i_hidp", "j_hipd" and "k_hidp", which uniquely identify households within waves 9, 10 and 11 of the annual survey, are also made available in the COVID-19 survey data files and can be used to link to households in the wave 9 and 10 annual survey data. See Section 21 for more information on linking to the main annual survey datasets.

10.8 How to use the telephone survey data

As described in Section 6.3, a subset of sample members who had not completed the April web survey and lived in households where no-one regularly uses the internet were invited to a telephone survey. The first telephone survey was based on the April 2020 web questionnaire, although it was implemented in May, roughly coinciding with the May 2020 web survey. The data from the first telephone survey are in the file cb_indresp_t and have the same structure as the corresponding data in the web survey files (cW_indresp_w).

These telephone survey responses can be used in conjunction with either the April 2020 or May 2020 web survey responses. Since the telephone survey instrument was based on the April 2020 web survey questionnaire, combining the telephone respondents with the April 2020 web survey respondents offers the greatest common content. On the other hand, combining the telephone respondents with the May 2020 web survey respondents offers less common content, but with the advantage that the telephone responses and web survey responses refer to the same period of time.

A second telephone survey was carried out in November 2020 (see Section 6.9). This was based on the November web questionnaire and fielded at the same time. The data (in the file cf_indresp_t) have the same structure as the corresponding data in the web survey files.

11. Contents of the INDRESP files

The INDRESP files from the COVID-19 study include responses to the web surveys and telephone surveys.

11.1 Additional information

These files also include some additional variables that are not documented in the corresponding month's questionnaire. These include the sample design variables ("psu", "strata"), from when the respondent was first selected into the *Understanding Society* sample, and socio-demographic information that is only asked the first time a respondents completes an *Understanding Society* interview (birthy, racel_dv, bornuk_dv). Information for these variables are collected the first time a sample members is interviewed.

Some information from waves 9-11, such as survey outcomes and household identifiers in those waves are also included. The household identifiers are "i_hidp" to "k_hidp" and the survey interview outcome variables are "i_ioutcome" to "k_ioutcome". The interview outcomes will allow identification of the last wave a respondent was interviewed in the annual survey.

A unique address identifier ("cW_aid_dv") is also provided which identifies individuals living at the same address.

Using information on addresses, the region (formerly Government Office Region) in the UK where the respondent lives is identified and provided in the variable "cW_gor_dv".

Table 11.1 Additional variables in the COVID-19 INDRESP files

Variable	Description
pidp	Cross-wave Person Identifier (Public Release) [4]
psu	Primary sampling unit[3]
strata	Sampling strata[3]
birthy	Year of birth[3]
racel_dv	Ethnic group[3]
bornuk_dv	Born in UK (derived) [3]
i_hidp	Household Identifier (Public Release) – wave 9 annual interview[3]
j_hidp	Household Identifier (Public Release) – wave 10 annual interview[3]
k_hidp	Household Identifier (Public Release) – wave 11 annual interview[3]
i_ioutcome	Individual outcome – wave 9 annual interview[3]
j_ioutcome	Individual outcome – wave 10 annual interview[3]
k_ioutcome	Individual outcome – wave 11 annual interview[3]
cW_gor_dv	Government Office Region[3]
cW_aid_dv	Address identifier – identifies respondents living at the same address
cW_outcome	Survey outcome – whether full or partial respondent. Values: 11 Full interview, 12 Partial interview
cW_lastq	Last question respondent answered. String variable. Example: "EndComp"
cW_lastmodule	Last module the respondent started (from May 2020 data onwards)
cW_scghq1_dv	Subjective wellbeing (GHQ): Likert
cW_scghq2_dv	Subjective wellbeing (GHQ): Caseness
cW_clinvuln_dv	At risk of serious illness from COVID-19 (see Section 11.7 below)
cW_betaindin_xw	Cross-sectional individual web survey weight, beta version
cW_betaindin_xw_t	Cross-sectional individual tel + web survey weight, beta version
cW_betaindin_lw	Longitudinal individual web survey weight, beta version
cf_betaindin_lw_t1	Longitudinal individual tel + wave 1 web survey weight, beta version (November 2020 data only)
cf_betaindin_lw_t2	Longitudinal individual tel + wave 2 web survey weight, beta version (November 2020 data only)
ca_pidpcorrected	Identifies respondents who reported a date of birth or sex in the April 2020 web survey that did not match the sample data and whose identity has since been confirmed. Values: -8 Inapplicable, 1 Strong dob match, sex match, 2 Strong dob match, sex missing, 3 Medium dob match, sex match, 4 Medium dob match, sex missing
cW_link	Which link respondent used to access the survey. Values: 1 Email, 2 SMS, 3 Postal (by typing survey url given in the letter into a browser)
cW_surveystart	Date and time respondent started the survey
cW_surveyend	Date and time respondent ended the survey
cW_surveytime	Total time spent in the survey, seconds
cW_screenres	Screen resolution String. Example: "360x747"

---

[4] See *Understanding Society* data documentation at
https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation

Variable	Description
cW_browserres	Browser resolution. String. Example: "1366x657"
cW_useragentstring	User agent string of browser: identifies the browser the respondent used to complete the survey, the version of the browser, and the operating system of the respondent's device. String variable. Example: ""Mozilla/5.0 (Linux; Android 9; FIG-LX1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.117 Mobile Safari/537.36"
ch_sentkit	Whether respondent was sent a serology testing kit. Values: 0 No, they did not consent 1 Yes, they were sent a kit 2 No, they consented but have moved abroad 3 No, they consented but we do not have an address for them

The variable "cW_outcome" documents whether the respondent completed the full questionnaire or dropped out part-way through (as shown in Table 6.2). The variable "cW_lastq" is a string variable that documents the name of the last question the respondent answered.

The variables "cW_scghq1_dv" and "cW_scghq2_dv" are derived from the COVID-19 survey data. The definitions and Stata code used to create these variables are available in Appendix B.

The survey weights "cW_betaindin_xw", "cW_betaindin_xw_t", "cW_betaindin_lw", "cf_betaindin_lw_t1", and "cf_betaindin_lw_t2" are documented in Section 17 below.

Variables "cW_link" to "cW_useragentstring" listed in Table 11.1 are paradata collected automatically by the survey software.

The variable "ch_sentkit" (only in the file ch_indresp_w) indicates whether respondents who consented to be sent a COVID-19 antibody test kit where actually sent one, and if not whether because they had moved abroad or because we did not have a valid address.

11.2 Feed-forward variables

The data include a number of feed-forward variables that contain information from previous waves of the COVID-19 survey. These variables are part of the sample file and are read into the questionnaire script, to determine routing through the questionnaire. For each survey wave, the feed-forward variables used in that wave are copied in to the cW_indresp_w file. All feed-forward variables have the prefix "cW_ff_".

The variable cW_ff_prevsurv indicates whether the respondent was a full respondent in at least one previous COVID-19 survey (0 "No", 1 "Yes"). The variables cW_ff_intd, cW_ff_intm, cW_ff_inty record the day, month and year when the respondent last completed a COVID-19 survey.

All other feed-forward variables are derived from responses to particular questions. Note these are only derived for respondents who have completed at least one full survey, that is, reached at least the cut-off point in the questionnaire beyond which they are treated as full respondents (see Table 6.2). The feed-forward variables are updated prior to each wave, but only for those who completed a full survey in the previous wave. For partial respondents and non-respondents, the previous values of the feed-forward variables are carried forward (except for an error in wave 6, see Section 22.5). How the feed-forward variables are derived is documented in the questionnaire: at the start of each module, any feed-forward variables used for routing are specified. There are three types of feed-forward variables: baseline, last interview, and cumulative.

The 'baseline' feed-forward variables relate to survey questions about the respondent's situation before the start of the pandemic. These questions are only asked the first time a respondent completes the COVID-19 survey. The feed-forward variable takes on the value of the answer and is never over-written in future waves. For example:

---

ff_blwork [Baseline: worked in Jan Feb 2020]
Notes: values of blwork from the first COVID-19 survey the respondent completed.
Don't over-write.
1. Yes, employed
2. Yes, self-employed
3. Yes, both employed and self-employed
4. No

The 'last interview' feed-forward variables take on the value of answers from the last COVID-19 survey the respondent has completed to date. For example:

---

ff_pregnow [Whether pregnant]
Notes: values of pregnow from the last COVID-19 survey completed.
1. Yes
2. No
3. Don't know

Please note that in a given wave, the last interview will be different for different respondents. So, when responding to questions which refer to the last interview, respondents will be referring to different time periods depending on when their last interview was. The "cW_outcome" variables in the xsample file can be used to identify which wave the respondent last completed.

The 'cumulative' feed-forward variables take on the value "1" if the respondent has reported the relevant status in any of the previous COVID-19 surveys. Note that in these

variables the value "0" includes all respondents not classified as "1": respondents who have not been asked the relevant question and respondents who have given answers that do not fall in category "1". In the example below, category "0" includes respondents not in employment and employees who have not been furloughed.

---

ff_furlough [Furloughed under the Coronavirus Job Retention Scheme]
    0. Not furloughed: IF furlough is not 1 in any of the previous COVID-19 surveys
    1. Furloughed: IF furlough = 1 in any of the previous COVID-19 surveys

---

11.3 Income and earnings variables

The survey collects two types of earnings information, individual and household. Unlike in the annual interviews for the main *Understanding Society* survey, household earnings are reported by the respondent rather than computed from individual reports of all adult respondents. The first time a respondent completes the COVID-19 survey, they are also asked their net individual and household earnings in January or February 2020, to reflect their 'baseline' economic status before the COVID-19 pandemic hit the UK. From May 2020 overall household income information is also collected.

Each type of income or earnings information is provided in two variables: the variable with the suffix "_amount" is the amount reported and the variable with the suffix "_period" shows the period over which the amount is reported. See Table 11.2 for a complete list of these variables and their descriptions.

Table 11.2 Income and earnings variables

Variable name	Variable description
cW_blpay_amount	Net individual earnings in Jan/Feb 2020
cW_blpay_period	Period (week, two weeks, month, year)
cW_blhhearn_amount	Net total household earnings in Jan/Feb 2020
cW_blhhearn_period	Period (week, two weeks, month, year)
cW_netpay_amount	Net individual earnings at time of interview
cW_netpay_period	Period (week, two weeks, month, year)
cW_grosspay_amount	Gross individual earnings at time of interview (w6 onwards)
cW_grosspay_period	Period (week, two weeks, month, year) (w6 onwards)
cW_hhearn_amount	Net total household earnings at time of interview
cW_hhearn_period	Period (week, two weeks, month, year)
cW_hhincome_amount	Net household income at time of interview (w2 onwards)
cW_hhincome_period	Period (week, two weeks, month, year) (w2 onwards)
cW_ghhincome_amount	Gross household income at time of interview (w5 onwards)
cW_ghhincome_period	Period (week, two weeks, month, year) (w5 onwards)

11.4 Responses to open-ended questions

Answers to open-ended questions are not released, due to confidentiality reasons. The COVID-19 survey includes three types of open-ended questions: 1) follow-ups to questions that include an open-ended "other" response category, 2) an "anything else you would like to tell us?" question at the end of the survey, and 3) open-ended questions about occupation.

Table 11.3 documents the questions that include open-ended "other" response options. For each of these, the open-ended responses were coded and where answers mapped onto existing responses options, the answers to those were updated. All remaining responses, i.e. new response categories that had not been included in the original question, are made available in the corresponding variable with the suffix "_oth". For questions that are repeated in later waves, the response options to the original questions have been expanded to include the additional categories and the open-ended follow-up questions have been dropped. The response categories coded from the open-ended questions can be found by searching for the corresponding "oth" variables in the Variable Search in the online documentation at https://www.understandingsociety.ac.uk/documentation/covid-19/dataset-documentation.

The April 2020 questionnaire also included an open-ended debrief question asking *"Is there anything else you would like to tell us about how the coronavirus is affecting your life, that we have not asked you about?"*. From May 2020 onwards this question has been amended to *"…that you have not previously told us about?"* When resources allow, we will code the responses to these questions and include them in a later release of the data.

Open-ended questions about occupations in the job search module are coded using Standard Occupational Classification (SOC) codes. The classification is coded after fieldwork, using the respondent's verbal description of the occupation, whether manager or foreman/supervisor, and industry. The SOC 2010 code frame is documented in Appendix A. The EUL version of the data includes the condensed 3-digit version of the SOC codes. The variable names include the suffix "_cc" to indicate the condensed codes.

The coding of the open-ended occupation questions in the employment module is still ongoing, see Section 22.2.

Table 11.3 Open-ended "other" response categories

Wave	Module	Question including open-ended "other" response category	Coded open-ended responses
W1	Employment	hrschange1	ca_hrschange1_oth
		hrschange2	ca_hrschange2_oth
		hrschange3	ca_hrschange3_oth
	Finance	inoutflows	ca_inoutflows_oth
W2	Household relationships	cvwhymove cvwhymovein cvwhymoveout	cb_cvwhymove_oth
			cb_cvwhymovein_oth
			cb_cvwhymoveout_oth
W3	Non-resident parents	cmrecmore	cc_cmrecmore_oth
	Non-resident children	kidsptmore	cc_kidsptmore_oth
	Working conditions	prodfall	cc_prodfall_oth
W4	Financial security	mpc3	cd_mpc3_oth
W5	Return to school	sumnogo	ce_sumnogooth_child*
		autnogo	ce_autnogooth_child*
W5	Job search	julk4whynot	ce_julk4whynot_oth
W6	Coronavirus illness	lgcvsymp	cf_lgcvsymp_oth
		whynotapp	cf_whynotapp_oth
		covidappnot	cf_covidappnot_oth
W7	Coronavirus illness	lgcvsymp	cg_lgcvsymp_oth
	Coronavirus vaccine	vaxno	cg_vaxno_oth
		vaxwhy	cg_vaxwhy_oth
	Job search	julk4whynot	cg_julk4whynot_oth
	Diet and food security	fdbkwhy	cg_fdbkwhy_oth
W8	Coronavirus illness	lgcvsymp	ch_lgcvsymp_oth
W9	Coronavirus illness	lgcvsymp	ci_lgcvsymp_oth
	NHS app	whynotapp2	ci_whynotapp_oth
		whyuninst	ci_whyuninst_oth
		covidappnot2	ci_covidappnot_oth
		nhsappno	ci_nhsappno_oth
		nhsappyes	ci_nhsappyes_oth
	Job search	julk4whynot	ci_julk4whynot_oth

11.5 Top coding of variables

In the EUL version of the data, the individual and household earnings amounts and the household size and composition variables are top-coded in order to prevent disclosure.

Individual earnings are top-coded at £4,000 net per week. Household earnings amounts and period codes are top-coded where the difference between household and individual earnings exceeds £4,000 net per week. Household income amounts and period codes are top-coded where the difference between household income and individual earnings exceeds £4,000 per week. Top-coded amounts are identified with a value of 5 "Weekly (top-coded)" in the corresponding period code variable.

The variables related to household size and composition (cW_hhcompa – cW_hhcompe) are also top-coded, to prevent potential identification of unusually large households. These variables are counts of the number of people in different age groups living in the household. For each variable and wave, the highest category with more than 100 observations becomes the top-coded category, that is, all higher categories are recoded to this. For example, if ca_hhcompa includes 2,900 observations for the value 0, 4,500 for the value 1, 350 for value 2, 65 for value 3, and 20 for value 4, then all values above 2 are recoded to 2 and this variable now has 3 values: 0, 1, 2, where the label for 2 is "2 and above". As the frequencies for these variables will be different across the variables and waves, the value at which the top-coding occurs will be different. The household composition variables asked only of Welsh speakers (cW_hhcompwsha – cW_hhcompwshe) are dropped, since the values reported are included in the derived variables cW_hhcompa – cW_hhcompe.

11.6 Recoding of rare health conditions

Rare health conditions reported in response to the question hcond_cv and hcondew_cv in the "long term health condition management" module are combined with the "other" category, to reduce risks of disclosure. The rare conditions recoded in this way are H.I.V. (response option 20), problems with spleen or removed spleen (25), cystic fibrosis (28), and sickle cell disease (26). The same recoding is applied to the feed-forward variable cW_ff_hhcond18, with feed-forward variables cW_ff_hcond26, cW_ff_hcond28, cW_ff_hcond25, and cW_ff_hcond20 dropped.

11.7 At risk of serious illness from COVID-19

The NHS created a two-level list of health conditions which made people at risk from serious illness if they are infected by COVID-19 (https://www.nhs.uk/conditions/coronavirus-covid-19/people-at-higher-risk/whos-at-higher-risk-from-coronavirus/). Based on this list, in April, we added extra health conditions to our standard conditions list, as well as capturing different kinds of treatments which were included in the definitions of the at-risk groups in April 2020. We then matched our data to the NHS definitions from May 2020, as set out in

the table below, and created a single variable ("cW_clinvuln_dv") for being clinically vulnerable to serious illness from COVID-19 with three mutually exclusive categories:

- cW_clinvuln_dv = 0 no risk (not clinically vulnerable)
- cW_clinvuln_dv = 1 moderate risk (clinically vulnerable)
- cW_clinvuln_dv = 2 high risk (clinically extremely vulnerable)

Respondents are allocated to the highest risk category if they are in both at risk groups e.g. a person over 70 and who has leukaemia would be in cW_clinvuln_dv = 1 extremely clinically vulnerable not cW_clinvuln_dv = 2.

Over time the NHS has refined its definitions, and our questions are broader than more precise NHS records will be. This variable therefore is likely to be an overestimation of those at risk of serious illness if they are infected with COVID-19.

Table 11.4 NHS categorisation of 'At risk of serious illness from COVID-19' and variables used to define this in the INDRESP files

At risk of serious illness from COVID-19 (NHS definition May 2020)	How capture in Understanding Society data* Variable = cW_clinvuln_dv
high risk (clinically extremely vulnerable) cW_clinvuln_dv =2	
have had an organ transplant	cW_hcond_treat1=1 Medication following an organ transplant
having chemotherapy or antibody treatment for cancer, including immunotherapy	cW_hcond_treat3=1 Targeted therapy or chemotherapy for cancer treatment
having an intense course of radiotherapy (radical radiotherapy) for lung cancer	cW_hcond_treat4 =1 and ca_cancertyp_cv = 2 cW_hcond_treat4 =1 and (cb_cancertyp_cv = 2 or cb_cancertypn_cv = 2) *Wave c onwards* cW_hcond_treat4 =1 and cW_cancertyp_cv2 = 1 cW_hcond_treat4 =1 and cW_cancertypn_cv2 = 1 Radiotherapy treatment and lung cancer
having targeted cancer treatments that can affect the immune system (such as protein kinase inhibitors or PARP inhibitors)	cW_hcond_treat3 =1 Targeted therapy or chemotherapy for cancer treatment

have blood or bone marrow cancer (such as leukaemia, lymphoma or myeloma)	ca_ cancertyp_cv= 8 cb_cancertyp_cv = 8 or cb_cancertypn_cv = 8 *Wave c onwards* cW_cancertyp_cv8=1 or cW_cancertypn_cv8 = 1 Blood or bone marrow cancer, such as leukaemia
have had a bone marrow or stem cell transplant in the past 6 months, or are still taking immunosuppressant medicine	cW_hcond_treat5=1 Other treatment or medication that may affect immune system
have been told by a doctor they have a severe lung condition (such as cystic fibrosis, severe asthma or severe COPD)	cW_hcond_cv28=1 or cW_hncondnew_cv28=1 Cystic fibrosis Note given variables available we have defined 'severe' as taking treatment such as steroids, or other medications that weaken the immune system. cW_hcond_treat2 = 1 or cW_hcond_treat5 =1 While technically for this category this should be limited to those with respiratory conditions listed below, in fact everyone with these treatments is defined as clinically extremely vulnerable so the respiratory variables are not included in the extremely vulnerable syntax.
have a condition that means they have a very high risk of getting infections (such as SCID or sickle cell)	cW_hcond_cv26=1 or cW_hcondnew _cv26=1 Sickle cell disease cW_hcond_cv25=1 or cW_hcondnew_cv25=1 Spleen removed
taking medicine that makes them much more likely to get infections (such as high doses of steroids or immunosuppressant medicine)	cW_hcond_treat5=1 Other treatment or medication that may affect immune system cW_hcond_treat2=1 Medicines such as steroid tablets that weaken the immune system
have a serious heart condition and are pregnant	cW_pregnow =1 or cW_stillpreg =1 AND (cW_hcond_cv3=1 or cW_hcondnew_cv3 =1 Congestive heart failure

| | cW_hcond_cv4=1 or cW_hcondnew_cv4 =1 Coronary heart disease |
| | cW_hcond_cv6=1 or cW_hcondnew_cv6 =1 Heart attack or myocardial infarction) |

moderate risk (clinically vulnerable) cW_clinvuln_dv =1

70 or older	Age >=70
have a lung condition that's not severe (such as asthma, COPD, emphysema or bronchitis)	cW_hcond_cv1=1 OR cW_hcondnew _cv1=1 Asthma cW_hcond_cv8=1 OR cW_hcondnew _cv8=1 Emphysema cW_hcond_cv21=1 OR cW_hcondnew _cv21=1 COPD cW_hcond_cv11=1. OR cW_hcondnew _cv11=1 Chronic bronchitis Note given variables available we have defined 'not severe' as not taking treatment such as steroids, or other mediations that weaken the immune system. AND cW_hcond_treat2 ne 1 or cW_hcond_treat5 ne 1
have heart disease (such as heart failure)	cW_hcond_cv3=1 or cW_hcondnew_cv3 =1 Congestive heart failure cW_hcond_cv4=1 or cW_hcondnew_cv4 =1 Coronary heart disease cW_hcond_cv6=1 or cW_hcondnew_cv6 =1 Heart attack or myocardial infarction
have diabetes	cW_hcond_cv14=1 or cW_hcondnew_cv14=1 Diabetes
have chronic kidney disease	cW_hcond_cv23=1 or cW_hcondnew_cv23=1 Chronic kidney disease
have liver disease (such as hepatitis)	cW_hcond_cv12=1 or cW_hcondnew_cv12=1 Any kind of liver condition
have a condition affecting the brain or nerves (such as Parkinson's disease, motor neurone disease, multiple sclerosis or cerebral palsy)	cW_hcond_cv24=1 or cW_hcondnew_cv24=1 Conditions affecting the brain and nerves, such as Parkinson's disease, motor neurone disease, a learning disability or cerebral palsy cW_hcond_cv19=1 or cW_hcondnew_cv19=1 Multiple Sclerosis
have a condition that means they have a high risk of getting infections	

taking medicine that can affect the immune system (such as low doses of steroids)	Given variables in the Study, these two categories are subsumed into extremely clinical vulnerable group above, as we cannot distinguish high and low dose steroids.
very obese (a BMI of 40 or above)	cW_hcond_cv27=1 or cW_hncondnew_cv27=1 Are very overweight (having a BMI of 40 or above)
pregnant	cW_pregnow =1 or cW_stillpreg=1

*Note cW_hcond_treat_cv1 to cW_hcond_treat_cv16 was asked of everyone in April and from July onwards. Unfortunately, in May and June, a routing error meant this question was only asked of new entrants in May and June, and hence people in May and June who newly started any of these treatments will have been missed. We do not believe this will be a significant number of respondents.

11.8 How to identify respondents in the same household in the COVID-19 surveys

The COVID-19 surveys do not include household enumeration grids and so there is no direct way to identify respondents in the same household. The unique address identifier ("cW_aid_dv") provided in the "cW_indresp_w" files can be used to identify individuals living at the same address. As those living at the same address may, in a minority of cases, not be living in the same household, we suggest using this information along with the household identifier at the last interview to identify individual respondents belonging to the same household. The last interview wave can be identified using the interview outcome variables "i_ioutcome", "j_ioutcome" and "k_ioutcome" available in the "cW_indresp_w" files. The household identifier for that wave ("i_hidp", "j_hidp" and "k_hidp" also available in the "cW_indresp_w" files) can then be used to identify whether the respondents were in the same household in that wave.

11.9 How to identify spouses and partners in the COVID-19 surveys

In the COVID-19 survey all adults in the households in the "active sample" are eligible. Since the survey does not include a household enumeration grid, there is no direct way to identify a partner/spouse. There is an indirect way to do this. The variable "w_ppid" is the pidp of the co-resident partner or spouse in the main annual survey wave w. If you identify the pidp of the spouse/partner from the last known main annual survey wave when that respondent was enumerated, and that pidp is the pidp of someone who currently lives at the same address as the respondent (use "cW_aid_dv") and the respondent also reports living as a couple ("cW_couple), then that person is most likely the partner/spouse.

12. Contents of the sample file XSAMPLE

The XSAMPLE file includes one observation on each of the 44,067 sample members who were eligible for the COVID-19 study in April 2020. So, unlike the INDRESP files, it also includes information on those who were eligible for the study but not invited (because they had attrited from the annual survey or no contact details were known in April 2020) and on non-respondents. This file was updated after each wave to include information about the survey implementation and outcomes from the latest survey. These variables are documented in Table 12.1.

The XSAMPLE file also includes the feed-forward variables for all waves. See Section 11.2 for a description of these variables.

Table 12.1: Survey implementation and outcome variables in the COVID-19 XSAMPLE file

Variable	Description and values
emailknown	Whether email known at start of study. Values: 0 No, 1 Yes
mobknown	Whether mobile number known at start of study. Values: 0 No, 1 Yes
preletter	Prenotification letter version: depending on what contact information was known for the sample member in April 2020. Values: 1 Email and mobile, 2 Email but no mobile, 3 Mobile but no email, 4 No email or mobile
batch	Mailing batch: randomized allocation to batches used to send out invitations and reminders in hourly slots, used for all mailings in all months. Range: 1-8
emailoptout	Opted out of emails. Values: 0 No, 1 Yes
smsoptout	Opted out of SMS. Values: 0 No, 1 Yes
cW_outcome	Outcome of web survey. Values: 11 Full interview, 12 Partial interview, 21 Non-respondent, 22 Opted out (not issued), 31 Ineligible (deceased), 32 Ineligible (moved abroad), 33 No contact details (not issued), 41 Transferred to telephone survey (not issued), 42 No email or mobile info (not issued), 43 Non-respondent in waves 1-4 – (not issued), 51 Attrited from mainstage before April 2020 (not issued), 52 Unknown contact details in April 2020 (not issued)
cW_outcome_t	Outcome of telephone survey. Values: 1 Full interview, 2 Partial interview, 3 No contact made, 4 Contact made but no further contact, 5 Dead number, 6 Refusal / Hard Refusal / Respondent Refusal, 7 Unknown at number, 8 Moved, 9 Incapable of Interview, 10 Died
cW_letteri	Invitation to survey sent by post: if no email or mobile phone number known. Values: 0 No, 1 Yes
cW_emaili	Invitation sent by email. Values: 0 No, 1 Yes
cW_emailr2	1st reminder email: if survey not yet completed. Values: 0 No, 1 Yes
cW_emailr2	2nd reminder email: if survey not yet completed. Values: 0 No, 1 Yes

Variable	Description and values
cW_emailr3	3rd reminder email: if survey not yet completed. Values: 0 No, 1 Yes
cW_smsi	Invitation sent by SMS. Values: 0 No, 1 Yes
cW_smsr1	1st reminder SMS: if survey not yet completed. Values: 0 No, 1 Yes
cW_smsr2	2nd reminder SMS: if survey not yet completed. Values: 0 No, 1 Yes
cW_smsr3	3rd reminder SMS: if survey not yet completed. Values: 0 No, 1 Yes
cW_letteri_d	Date invitation letter sent. Missing if n/a
cW_emaili_d	Date and time stamp invitation email sent. Missing if n/a
cW_emailr1_d	Date and time stamp 1st reminder email sent. Missing if n/a
cW_emailr2_d	Date and time stamp 2nd reminder email sent. Missing if n/a
cW_emailr3_d	Date and time stamp 3rd reminder email sent. Missing if n/a
cW_smsi_d	Date and time stamp invitation SMS sent. Missing if n/a
cW_smsr1_d	Date and time stamp 1st reminder SMS sent. Missing if n/a
cW_smsr2_d	Date and time stamp 2nd reminder SMS sent. Missing if n/a
cW_smsr3_d	Date and time stamp 3rd reminder SMS sent. Missing if n/a
cd_ff_incentbonus	Randomised allocation to the July 2020 web survey incentive experiment. Values: -8 Inapplicable (£2), 1 £2 + £10 bonus, 2 £2
cd_incentive	Incentive respondent received. Values: 1 £12, 2 £2
ca_pidpcorrected	Identifies respondents who reported a date of birth or sex in the April 2020 web survey that did not match the sample data and whose identity has since been confirmed. Values: -8 Inapplicable, 1 Strong dob match, sex match, 2 Strong dob match, sex missing, 3 Medium dob match, sex match, 4 Medium dob match, sex missing
ch_ff_incentw8	Randomised allocation to the March 2021 web survey incentive experiment. Values: -8 Inapplicable (£2), 1 £2 for survey + £5 for blood sample, 2 £7 for survey + £5 for blood sample, 3 £12 for survey + £5 for blood sample, 4 £7 for survey, 5 £12 for survey, 6 £17 for survey
ch_ff_consentpos	Randomised allocation for the March 2021 position and order of consent questions. Values: 1 Early, health then serology, 2 Early, serology then health, 3 Early in context, health then serology, 4 Early in context, serology then health, 5 Late, health then serology, 6 Late, serology then health, 7 Early in context serology, later in context health

The XSAMPLE file also includes the sample design variables, stable characteristics and outcomes from previous annual interviews that are included in the INDRESP files: "pidp", "psu", "strata", "birthy", "racel_dv", "bornuk_dv", "i_hidp" to "l_hidp", "i_ioutcome" to "l_ioutcome", plus the address identifier cW_aid_dv. For descriptions of these variables see Section 11.1. It also includes the variable "sex_dv" copied from the previous annual interview data. Note that the response options for the question asking about respondent's

sex in the survey changed in wave 3 and so the variable in the cW_indresp files was changed from "cW_sex" in waves 1 and 2, to cW_sex_cv in wave 3 onwards.

Finally, the XSAMPLE file contains the treatment allocation and outcome variables for the incentive experiment implemented in the July 2020 web survey: "cd_ff_incentbonus" and "cd_incentive" (see Section 6.5 for a description of the experiment); and for the incentive experiment and consent position and order experiment in the March 2021 web survey: "ch_ff_incentw8" and "ch_ff_consentpos" (See Section 6.12).

13. Contents of baseline question file XBASELINE

The COVID-19 surveys asked a series of questions about a January-February 2020 pre-pandemic "baseline". These questions, which cover aspects of health, employment and finances were only asked of respondents once: the first time they completed a full COVID-19 survey. This means that baseline information for the sample of respondents is spread across different data files, depending on when the respondent first completed the survey. To simplify analyses, we have collected the responses to these questions in a single file, "xbaseline". See Table 13.1 for the items included in this data file.

Not all baseline questions were however asked of everyone, and some baseline questions were changed across waves. Where item missingness occurs because of this, the respondent value is set to "Inapplicable" (-8). Such instances are explained next.

The finance module was included in waves 1, 2, 4 and 6 of the web survey and in the May and November telephone survey. This means that respondents who first answered the COVID-19 survey in wave 3 were never asked the baseline questions in the finance module. The answers to these items are therefore not available for 1,169 of the 20,462 COVID-19 survey respondents in the "xbaseline" data file.

Changes in question wording and routing over waves also led to some item missingness and/or multiplication of variables. "blnonwork" was not asked at wave 1 or in the May telephone survey, but was asked of all respondents from wave 2 onwards. Hence, answers are not available for respondents who only completed wave 1 or the May telephone survey. In addition, benefit questions (in the finance module) were only asked of those aged 65 and under in wave 4, which is again reflected by separate sets of variables.

The file additionally includes the unique respondent identifier, "pidp", a "firstinterview" variable indicating the wave at which they were first interviewed, and a "bl_outcome" variable detailing whether the interview was a full or partial one. It also includes the cross-sectional respondent weight "bl_betaindin_xw" (see Section 17.4 for a description).

Note that the xbaseline file does not include baseline questions from modules that were asked only once in April, May, June or July 2020. For these questions the data from all respondents are in the cW_indresp_w file for the corresponding wave.

A small number of respondents were asked baseline questions twice. These are the 'mismatched' cases in the April web survey, who reported a date of birth and/or sex that did not match the information on the sample file (see Section 6.1). For the May web fieldwork these cases were treated as non-respondents to the April survey, since we did not have reliable feed-forward data for them. They were therefore asked the baseline questions again. For these respondents the variables in the xbaseline file are from the April survey, since that is closest in time to the period about which the baseline questions asked (January/February 2020).

Table 13.1: Baseline questions in the XBASELINE file. 'T' indicates May telephone survey.

Variable	Waves	Description and values
pidp	1,2,3,4,T	Cross-wave Person Identifier (Public Release)
nhsshield_dv	1,2,3,4,T	Have you received a letter from the NHS or Chief Medical Officer saying that you have been identified as someone at risk of severe illness if you catch coronavirus, because you have an underlying disease or health conditions? Values 0 No, 1 Yes
hcond_cv1_dv	1,2,3,4,T	Baseline: health condition - Asthma. Values: 0 Not mentioned, 1 Mentioned
hcond_cv2_dv	1,2,3,4,T	Baseline: health condition - Arthritis. Values: 0 Not mentioned, 1 Mentioned
hcond_cv3_dv	1,2,3,4,T	Baseline: health condition - Congestive heart failure. Values: 0 Not mentioned, 1 Mentioned
hcond_cv4_dv	1,2,3,4,T	Baseline: health condition - Coronary heart disease. Values: 0 Not mentioned, 1 Mentioned
hcond_cv5_dv	1,2,3,4,T	Baseline: health condition - Angina. Values: 0 Not mentioned, 1 Mentioned
hcond_cv6_dv	1,2,3,4,T	Baseline: health condition - Heart attack or myocardial infarction. Values: 0 Not mentioned, 1 Mentioned
hcond_cv7_dv	1,2,3,4,T	Baseline: health condition - Stroke. Values: 0 Not mentioned, 1 Mentioned
hcond_cv8_dv	1,2,3,4,T	Baseline: health condition - Emphysema. Values: 0 Not mentioned, 1 Mentioned
hcond_cv10_dv	1,2,3,4,T	Baseline: health condition - Hypothyroidism or an under-active thyroid. Values: 0 Not mentioned, 1 Mentioned
hcond_cv11_dv	1,2,3,4,T	Baseline: health condition - Chronic bronchitis. Values: 0 Not mentioned, 1 Mentioned
hcond_cv12_dv	1,2,3,4,T	Baseline: health condition - Any kind of liver condition. Values: 0 Not mentioned, 1 Mentioned
hcond_cv13_dv	1,2,3,4,T	Baseline: health condition - Cancer or malignancy. Values: 0 Not mentioned, 1 Mentioned
hcond_cv14_dv	1,2,3,4,T	Baseline: health condition - Diabetes. Values: 0 Not mentioned, 1 Mentioned

Variable	Waves	Description and values
hcond_cv15_dv	1,2,3,4,T	Baseline: health condition - Epilepsy. Values: 0 Not mentioned, 1 Mentioned
hcond_cv16_dv	1,2,3,4,T	Baseline: health condition - High blood pressure/hypertension. Values: 0 Not mentioned, 1 Mentioned
hcond_cv18_dv	1,2,3,4,T	Baseline: health condition - Other long standing/chronic condition. Values: 0 Not mentioned, 1 Mentioned
hcond_cv19_dv	1,2,3,4,T	Baseline: health condition - Multiple Sclerosis. Values: 0 Not mentioned, 1 Mentioned
hcond_cv21_dv	1,2,3,4,T	Baseline: health condition - COPD (Chronic Obstructive Pulmonary Disease). Values: 0 Not mentioned, 1 Mentioned
hcond_cv22_dv	1,2,3,4,T	Baseline: health condition - An emotional, nervous or psychiatric problem. Values: 0 Not mentioned, 1 Mentioned
hcond_cv23_dv	1,2,3,4,T	Baseline: health condition - Chronic kidney disease. Values: 0 Not mentioned, 1 Mentioned
hcond_cv24_dv	1,2,3,4,T	Baseline: health condition - Conditions affecting the brain and nerves. Values: 0 Not mentioned, 1 Mentioned
hcond_cv27_dv	1,2,3,4,T	Baseline: health condition - Are very overweight (having a BMI of 40 or above). Values: 0 Not mentioned, 1 Mentioned
hcond_cv96_dv	1,2,3,4,T	Baseline: health condition - None of these. Values: 0 Not mentioned, 1 Mentioned
arthtyp_dv	1,2,3,4,T	Baseline: if has arthritis. Values: 1 Osteoarthritis, 2 Rheumatoid arthritis, 3 Other type of arthritis, 4 More than one of the above, 5 Don't know
cancertyp_cv1_dv	1,2,3,4,T	If has cancer: Bowel / colorectal. Values: 0 Not mentioned, 1 Mentioned
cancertyp_cv2_dv	1,2,3,4,T	If has cancer: Lung. Values: 0 Not mentioned, 1 Mentioned
cancertyp_cv3_dv	1,2,3,4,T	If has cancer: Breast. Values: 0 Not mentioned, 1 Mentioned
cancertyp_cv4_dv	1,2,3,4,T	If has cancer: Prostrate. Values: 0 Not mentioned, 1 Mentioned
cancertyp_cv5_dv	1,2,3,4,T	If has cancer: Liver. Values: 0 Not mentioned, 1 Mentioned
cancertyp_cv6_dv	1,2,3,4,T	If has cancer: Skin cancer or melanoma. Values: 0 Not mentioned, 1 Mentioned
cancertyp_cv7_dv	1,2,3,4,T	If has cancer: Other. Values: 0 Not mentioned, 1 Mentioned
cancertyp_cv8_dv	1,2,3,4,T	If has cancer: Blood or bone marrow cancer, such as leukaemia. Values: 0 Not mentioned, 1 Mentioned
hcond_treat1_dv	1,2,3,4,T	Baseline: existing treatments - Medication following an organ transplant. Values: 0 Not mentioned, 1 Mentioned
hcond_treat2_dv	1,2,3,4,T	Baseline: existing treatments - Medicines that weaken the immune system. Values: 0 Not mentioned, 1 Mentioned

Variable	Waves	Description and values
hcond_treat3_dv	1,2,3,4,T	Baseline: existing treatments - Targeted/chemotherapy for cancer treatment. Values: 0 Not mentioned, 1 Mentioned
hcond_treat4_dv	1,2,3,4,T	Baseline: existing treatments - Radiotherapy for cancer treatment. Values: 0 Not mentioned, 1 Mentioned
hcond_treat5_dv	1,2,3,4,T	Baseline: existing treatments - Other treatment or medication. Values: 0 Not mentioned, 1 Mentioned
hcond_treat6_dv	1,2,3,4,T	Baseline: existing treatments - None of these. Values: 0 Not mentioned, 1 Mentioned
nhs2019gp_dv	1,2,3,4,T	In the 12 months to the end of February 2020, did you see your GP or other primary care provider? Values: 1 Yes, 2 No, 3 Not required
nhs2019pm_dv	1,2,3,4,T	In the 12 months to the end of February 2020, did you use prescription medicine? Values: 1 Yes, 2 No, 3 Not required
nhs2019op_dv	1,2,3,4,T	In the 12 months to the end of February 2020, did you go to a hospital or clinic as an outpatient? Values: 1 Yes, 2 No, 3 Not required
nhs2019ip_dv	1,2,3,4,T	In the 12 months to the end of February 2020, did you go to a hospital or clinic as an inpatient? Values: 1 Yes, 2 No, 3 Not required
nhs2019111_dv	1,2,3,4,T	In the 12 months to the end of February 2020, did you use NHS 111 in England, Wales and Northern Ireland or NHS 24 in Scotland? Values: 1 Yes, 2 No, 3 Not required
chsc2019pharm_dv	1,2,3,4,T	In the 12 months to the end of February 2020, did you speak to a local pharmacist? Values: 1 Yes, 2 No, 3 Not required
chsc2019otcm_dv	1,2,3,4,T	In the 12 months to the end of February 2020, did you use over the counter medicine? Values: 1 Yes, 2 No, 3 Not required
chsc2019carer_dv	1,2,3,4,T	In the 12 months to the end of February 2020, did you have someone visiting you at home to help with personal care/ medications / shopping / cooking / cleaning / wound dressing / injections?. Values: 1 Yes, 2 No, 3 Not required
chsc2019psy_dv	1,2,3,4,T	In the 12 months to the end of February 2020, did you speak to a counsellor or therapist? Values: 1 Yes, in person, 2 Yes, by telephone or online, 3 Yes, group sessions, 4 No , 5 Not required
blwork_dv	1,2,3,4,T	Were you in paid work or self-employment at any time in January or February 2020. Values: 1 Yes, employed, 2 Yes, self-employed, 3 Yes, both employed and self-employed, 4 No
blnonwork_dv	2,3,4	Baseline: activity in Jan Feb 2020. Values: 1 Unpaid or voluntary work, 2 Unemployed, 3 Permanently sick or disabled, 4 Looking after home/family, 5 In education, 6 Retired, 7 Other

Variable	Waves	Description and values
blhours_dv	1,2,3,4,T	During January and February 2020, how many hours did you usually work per week?
blhrshow_dv	1,2,3,4,T	How were your hours set during January and February? Values: 1 Fixed weekly, 2 I chose my hours, 3 My employer, with a minimum guaranteed, 4 My employer, with no minimum guaranteed.
blpayhow_dv	1,2,3,4,T	How were you paid during January and February 2020? Values: 1 Fixed salary, 2 For each job, 3 By hours worked, 4 By tips and commission, 5 A salary or hourly wage plus tips or commission.
blpay_amount_dv	1,2,3,4,T	What was your usual take-home pay/earnings during January and February 2020? Take-home pay is after tax, National Insurance and pension contributions have been deducted. Please include all jobs and self-employment activities.
blpay_period_dv	1,2,3,4,T	Earnings period. Values: 1 Week, 2 Two weeks, 3 Month, 4 Year.
blwah_dv	1,2,3,4,T	During January and February how often did you work at home? Values: 1 Always, 2 Often, 3 Sometimes, 4 Never
blhhearn_amount_dv	1,2,3,4,T	Household take home earnings in Jan / Feb 2020
blhhearn_period_dv	1,2,3,4,T	Household earnings period. Values: 1 Weekly, 2 Two weeks, 3 Monthly, 4 yearly.
blbenefits1_dv	1,2,T	Which of the following benefits were you receiving in January/February 2020: Universal credit. Values: 0 Not mentioned, 1 Mentioned
blbenefits2_dv	1,2,T	Which of the following benefits you were receiving in January/February 2020: Working tax credit. Values: 0 Not mentioned, 1 Mentioned
blbenefits3_dv	1,2,T	Which of the following benefits you were receiving in January/February 2020: Child tax credit. Values: 0 Not mentioned, 1 Mentioned
blbenefits4_dv	1,2,T	Which of the following benefits you were receiving in January/February 2020: Jobseekers Allowance. Values: 0 Not mentioned, 1 Mentioned
blbenefits5_dv	1,2,T	Which of the following benefits you were receiving in January/February 2020: Employment and Support Allowance. Values: 0 Not mentioned, 1 Mentioned
blbenefits6_dv	1,2,T	Which of the following benefits you were receiving in January/February 2020: None of these. Values: 0 Not mentioned, 1 Mentioned
blbenchange_dv	1,2,4,T	Has the amount you receive for Universal Credit, Working Tax Credit or Child Tax. Values: 1 Yes, increased, 2 Yes, decreased, 3 No

Variable	Waves	Description and values
blbenefitsb651_dv	4	Baseline: benefits - Universal Credit. Values: 0 Not mentioned, 1 Mentioned
blbenefitsb652_dv	4	Baseline: benefits - Working Tax Credit. Values: 0 Not mentioned, 1 Mentioned
blbenefitsb653_dv	4	Baseline: benefits - Child Tax Credit. Values: 0 Not mentioned, 1 Mentioned
blbenefitsb654_dv	4	Baseline: benefits - Jobseeker's Allowance. Values: 0 Not mentioned, 1 Mentioned
blbenefitsb655_dv	4	Baseline: benefits - Employment and Support Allowance. Values: 0 Not mentioned, 1 Mentioned
blbenefitsb656_dv	4	Baseline: benefits - None of these. Values: 0 Not mentioned, 1 Mentioned
blbenchangeb65_dv	4	Amount of benefits changed. Values: 1 Yes, increased , 2 Yes, decreased , 3 No
firstinterview	1,2,3,4,T	Wave at which first interview completed. Values: 1 w1_web, 2 w2_web, 3 w3_web, 4 w4_web, 5 w5_web, 6 w6_web, 1000 w1_tel, 2000 w_tel
bl_outcome	1,2,3,4,T	Survey outcome. Values: 11 Full interview, 12 Partial interview
bl_betaindin_xw	1,2,3,4,T	Cross-sectional baseline respondent weight, beta version

14. Contents of school child file SCHILD

Data on children have been collected from parents on four occasions: in the April, July, September and November web surveys as outlined below. The April questionnaire was also used for the May telephone survey. The November questionnaire was used for both the web and phone surveys. Please note the telephone samples for children are small and should only be used in conjunction with the web survey for the same questionnaire.

14.1 April 2020 web and May 2020 telephone: Home schooling (ages 4-18)

The adult April web survey and May telephone survey both included a module on home schooling that asked a series of questions about each child aged 18 or younger in the household who was enrolled in a school or college and for whom the respondent was the parent or guardian. The variables from these questions are included in the "ca_indresp_w" (for April web survey) and "cb_indresp_t" (for May telephone survey) file in wide format for each adult respondent. For each question there are 10 variables, corresponding to up to 10 children in the household. For example, for the question "atschool", the variables in the April web survey are "ca_atschool_childa", "ca_atschool_childb", etc. Although routed on a question *"Do you have any children aged 18 or younger in your household for whom you are the parent or guardian, who are currently enrolled in a school or college?"*, a range of people answered this question, including children over 16 years answering for themselves and

siblings, as well as one or both parents, and a range of relatives. The "ca_indresp_w" and "cb_indresp_t" files includes all such respondents' data whether or not they are the child's parent/guardian.

For the April web and May telephone surveys, the data are also available in child-level files ("ca_schild_w", "cb_schild_t" respectively), which contain a single observation per child. The answers to each of the home schooling questions are provided in 11 different variables, depending on who gave the answer. For example, "ca_atschool_mnspid" is data on whether the child is still attending school from the mother and "ca_atschool_fnspid" is the same question answered by the father. Table 14.1 summarizes the suffixes used to distinguish data provided by different people.

To identify unique children and their *Understanding Society* pidps, from parents' responses, children were linked to the main *Understanding Society* datasets by the child's first name, their date of birth and parents' date of birth and pidps. Where multiple respondents reported children with the same date of birth, careful checking of the child's name and other details were undertaken to distinguish between twins and different respondents' answers about the same child. Respondents aged 16-18 who answered the module for themselves or for siblings were removed. Any children below school age, i.e. born after August 2015, were also removed since they should not have been reported in the module on home schooling.

For the April web survey, this process identified a total of 4,627 unique school children, of whom 4,311 were identified in the annual *Understanding Society* data (pidps provided in "ca_schild_w"). This left 316 children not found in the annual interview data for whom new child pidps have been generated. These children are identified by the variable "newpidp_c=1". They may be children who have joined the Study since the last annual survey either through birth or by joining *Understanding Society* sample households, or children for whom the respondent entered the wrong month or year of birth.

The May telephone sample was smaller and hence there are fewer children of school age. The cb_schild_t file therefore only contains observations on 57 children. The telephone survey data should only be analysed in combination with the web survey data.

In addition to the data from the home schooling module, the files "ca_schild_w" and "cb_schild_t" include a set of identifiers and background information derived from the annual interview data. These variables are summarized in Table 14.2. They include the child's pidp ("pidp_c") and the pidps of the respondents who provided the answers to the home schooling questions. Since some respondents reported on multiple children, the variables "ca_childno_*" number the children a given person reported on. Each file also includes child weights. To analyse the data collected in the April web survey and the May telephone survey, users should rename the variables in the "ca_schild_w" and "cb_schild_t" files by dropping the wave prefixes (ca_, cb_), then combine (append) the data files, and use

the weights "betasch_xw_t" available in both these files. Please note, data from "cb_schild_t" should not be used by itself.

The April child file "ca_schild_w" also includes two derived variables: child's education stage ("ca_eduks_dv") and child's education phase ("ca_eduph_dv"). Variable ca_eduks_dv "Age group by education key stages" indicates the academic year group that the child was in, in April 2020, recoded into 6 education key stages. The academic year group was calculated based on the child's detailed date of birth and school admission age rules, so that: 1 "Year R: reception year (age 4)" includes children born between 1st Sept 2014-31st August 2015; 2 "KS1: years 1-2 (age 5-7)" includes children born between 1st Sept 2012-31st August 2014; 3 "KS2: years 3-6 (age 7-11)" includes children born between 1st Sept 2008-31st August 2012; 4 "KS3: years 7-9 (age 11-14)" includes children born between 1st Sept 2005-31st August 2008; 5 "KS4: years 10-11 (age 14-16)" includes children born between 1st Sept 2003-31st August 2005; 6 "KS5: years 12-13 (age 16-18)" includes children born between 1st Sept 2001-31st August 2003. Variable ca_eduph_cv "Age group by education phase" indicates child's age group recoded into three education phases (primary: reception- key stage 2, secondary: key stage 3-key stage 4, post-secondary: key stage 5) based on ca_eduks_dv.

The construction of the weights is described in Section 17.2 below.

14.2 July 2020: Special Education Needs (SEN) and Strengths and Difficulties Questionnaire (SDQ) (ages 5-17)

In the July adult web survey, parents of children aged 5 to 17 years old were asked if each of their children had a health or behavioural problem, special education needs or disability (SEN), and if so if they had a local authority plan in place (or equivalent in devolved nations). Parents of children aged 5-11 were also asked to complete the 25-item Strengths and Difficulties Questionnaire (SDQ). The variables from these questions are included in the "cd_indresp_w" file in wide format for each adult respondent. For each question there are 5 variables, corresponding to up to 5 children in the household.

All the data for children collected from parents are also provided in child level data file ("cd_schild_w"), which contains a single observation per child. The answers to each of the SEN questions are provided by one or both parents for children aged 5-17 years, which results in 8 different variables, depending on who gave the answer and depending on the age group of a child (data on children 5-11 and on children 12-17 are stored in separate variables). For example, "cd_send511_m" is data on whether the child in the 5-11 age group has an Education, Health and Care (EHC) plan, statement or coordinated support plan from the local authority from the mother and "cd_send511_f" is the same question answered by the father. Similarly, the 25 item SDQ scale, the total score and the five subscales are provided for each child by one or both parents for children aged 5-11 years. This resulted in a minimum of 31 variables if just one parent provided the data or 62 if both parents did. The derived SDQ scores provided are:

- Total difficulties score
- Emotional problems score
- Conduct problems score
- Hyperactivity score
- Peer problems score
- Prosocial score

These match the derived variables in the mainstage survey and code for creating them can be found in Appendix B.

Unique children and their *Understanding Society* pidps, were identified in the same way as outlined above for the April "ca_schild_w" file: by linking parents' responses for children to the main *Understanding Society* datasets by their date of birth and parents' date of birth and pidps. Where multiple respondents reported children with the same date of birth, careful checking of the child's name and other details were undertaken to distinguish between twins and different respondent's answers about the same child.

This process identified a total of 3,340 unique children aged 5-17 years with either SEN or SDQ or both data, of whom 3,305 were identified in the annual *Understanding Society* data (pidps provided in "cd_schild_w"). This left 35 children not found in the annual interview data for whom new child pidps have been generated. These children are identified by the variable "newpidp_c=1". They may be children who have joined the Study since the last annual survey either through birth or by joining *Understanding Society* sample households, or children for whom the respondent entered the wrong month or year of birth. Where these new children are identical to those in the April file, the same new child pidps have been used.

The file also includes child cross-sectional weights ("cd_betasch_xw") which are described in Section 17.2 below.

14.3 September 2020: Return to education (age 4-18 years) and Strengths and Difficulties Questionnaire (SDQ) age 5-11 years

In September parents of children aged 5-11 years were asked to complete the SDQ for each child, and parents of children aged 4-18 years completed questions about their return to school experience. This means there were two questionnaire loops for each parent for each of their children of relevant ages. These variables are included in the ce_indresp_w file in wide format.

All the data for children collected from parents are also provided in a child level data file ("ce_schild_w"), which contains a single observation per child. Unique children and their *Understanding Society* pidps, were identified in the same way as outlined above for preceding months.

This process identified a total of 3,291 unique children aged 4-18 years, 3,228 of whom were identified in the annual *Understanding Society* data (pidps provided in "ce_schild_w"). This left 63 children not found in the annual interview data for whom new child pidps have been generated. These children are identified by the variable "newpidp_c=1". They may be children who have joined the Study since the last annual survey either through birth or by joining *Understanding Society* sample households, or children for whom the respondent entered the wrong month or year of birth. Where these new children are identical to those in the April or July files, the same new child pidps have been used.

Of the children aged 4-18 years, 3,276 had data from one or both parents on the return to school module. For the subset of 1,540 children aged 5-11 years, 1,392 had data from at least one parent in the SDQ module.

The file also includes the cross-sectional child weights ("ce_betasch_xw") which are described in Section 17.2 below.

14.4 November 2020: Education experiences (age 4-18 years)

The November web and telephone surveys included a module in the adult questionnaire on their children's experience of education. Parents of children aged 4-18 were asked about each child's attendance at school, the need for home schooling and how they think their children feel about the impact of the pandemic and lockdown on their education. These variables are included in the adult files: cf_indresp_w (web survey) and cf_indresp_t (telephone survey) in wide format.

All the data for children collected from parents are also provided in a child level data file ("cf_schild_w", "cf_schild_t"), which contain a single observation per child. Unique children and their *Understanding Society* pidps, were identified in the same way as outlined above for preceding months.

In the web survey this process identified 2,987 unique children aged 4-18 years, 2,915 of whom were identified in the annual *Understanding Society* data. In the telephone survey data are provided on only 13 children. The telephone survey respondents should only be analysed in combination with the web survey data.

As in previous months, pidps from the annual survey are included where matched. For those children not found in the annual interview data, new child pidps have been generated. These children are identified by the variable "newpidp_c=1".

The files also include the cross-sectional child weights. The web survey file cf_schild_w includes "cf_betasch_xw" for analyses of the web survey only and "cf_betasch_xw _t " for analyses of the combined web and telephone data. The file cf_schild_t includes "cf_bestsch_xw_t" for the telephone sample component. The construction of the weights is described in Section 17.2 below.

14.5 January 2021: Home schooling (age 4-18 years)

The January web survey included a module in the adult questionnaire on children's home schooling. Parents of children aged 4-18 were asked some of the questions from the April 2020 home schooling module, plus questions about additional resources for home schooling and free school meals. These variables are included in the adult file cg_indresp_w in wide format.

All the data for children collected from parents are also provided in a child level data file ("cg_schild_w"), which contains a single observation per child. Unique children and their *Understanding Society* pidps, were identified in the same way as outlined above for preceding months. This process identified 2,870 unique children aged 4-18 years, 2,814 of whom were identified in the annual *Understanding Society* data. As in previous months, pidps from the annual survey are included where matched. For those children not found in the annual interview data, new child pidps have been generated. These children are identified by the variable "newpidp_c=1". The file also includes the cross-sectional child weight ("cg_betasch_xw"). The construction of the weights is described in Section 17.2 below.

14.6 March 2021: Strengths and Difficulties Questionnaire (SDQ) age 5-11 years

The March 2021 web survey included the Strengths and Difficulties Questionnaire (SDQ) in the adult questionnaire, for parents of children aged 5-11. These variables are included in the adult file ch_indresp_w in wide format.

All the data for children collected from parents are also provided in a child level data file ("ch_schild_w"), which contains a single observation per child. Unique children and their *Understanding Society* pidps, were identified in the same way as outlined above for preceding months. This process identified 1,416 unique children aged 5-11 years, 1,405 of whom were identified in the annual *Understanding Society* data. As in previous months, pidps from the annual survey are included where matched. For those children not found in the annual interview data, new child pidps have been generated. These children are identified by the variable "newpidp_c=1". The file also includes the cross-sectional child weight ("ch_betasch_xw"). The construction of the weights is described in Section 17.2 below.

Table 14.1 Suffixes used in SCHILD files to identify who provided answers

Variable name suffix	Who provided answers about child
_mnspid	Natural/step/adoptive mother identified via child file in annual interview data
_grmpid	Grandmother identified via child file in annual interview data
_fnspid	Natural/step/adoptive father identified via child file in annual interview data
_grfpid	Grandfather identified via child file in annual interview data
_afm	Female parent/guardian relative identified via egoalt file in annual interview data
_am	Male relative identified via egoalt file in annual interview data
_nmm	Male adult not identified via child and egoalt to annual interview data
_nmfm	Female adult not identified via child and egoalt to annual interview data
_m*	Main female adult/guardian/'mother': takes values provided by mnspid if available; or grmpid if available and mnspid not available; or by female adult identified in egoalt if the previous two not available; or by another female adult (nmfm) who provided answers for a child if mnspid/grmpid/afm not available
_f*	Equivalent of _m but for men/'fathers'
_1pg	'1st parent/guardian': takes values provided by mnspid if available; or fnspid if mnspid not available; or grmpid if the previous two not available; or grfpid if the previous not available; or apidpfm if the previous not available; or apidpm if the previous not available; or unmatched female guardian/'mother' if the previous not available; or unmatched male guardian/'father' if the previous not available

*In the July, September, November 2020, January 2021 and March 2021 web survey data there are only two suffixes (_m and _f), as the answers for each child were provided only by parents (or adult respondents who identified themselves as main guardians). This means that there are maximum two sets of answers per child: one provided by a mother/main female guardian (_m) and one provided by a father/main male guardian (_f)

Table 14.2: Additional variables in the COVID-19 SCHILD file

Variable	Description
newpidp_c*	Origin of child's pidp. Values: 0 Pidp from annual interview data, 1 New pidp, child not identified in annual data
pidp_c*	Cross-wave Person Identifier, child
pidp_m*	Cross-wave Person Identifier, main female adult/guardian/'mother'
pidp_f*	Cross-wave Person Identifier, main male adult/guardian/'father'
pidp_1pg	Cross-wave Person Identifier, 1st parent/guardian
pidp_mnspid*	Cross-wave Person Identifier, natural/step/adoptive mother
pidp_fnspid*	Cross-wave Person Identifier, natural/step/adoptive father
pidp_grmpid	Cross-wave Person Identifier, grandmother
pidp_grfpid	Cross-wave Person Identifier, grandfather
pidp_afm	Cross-wave Person Identifier, female relative (see ca_relationship_dv)
pidp_am	Cross-wave Person Identifier, male relative (see ca_relationship_dv)
pidp_nmfm	Cross-wave Person Identifier, female adult not matched to annual data
pidp_nmm	Cross-wave Person Identifier, male adult not matched to annual data
cW_relationship_dv$	Relationship to child, other. Values: 9 "Natural parent" 10 "Adoptive parent" 11 "Foster parent" 12 "Step-parent" 14 "Natural brother/sister" 15 "Half-brother/sister" 21 "Grand-parent" 23 "Aunt/uncle" 30 "Other non-relative"
ca_childno_mnspid	Child number for mnspid if mnspid non-missing. Values: 1 Child a, 2 Child b, etc.
ca_childno_fnspid	Child number for fnspid if fnspid non-missing. Values: 1 Child a, 2 Child b, etc.
ca_childno_grmpid	Child number for grmpid if grmpid non-missing. Values: 1 Child a, 2 Child b, etc
ca_childno_grfpid	Child number for grfpid if grfpid non-missing. Values: 1 Child a, 2 Child b, etc
ca_childno_afm	Child number for awpidp if awpidp non-missing. Values: 1 Child a, 2 Child b, etc
ca_childno_am	Child number for ampidp if ampidp non-missing. Values: 1 Child a, 2 Child b, etc
ca_childno_ nmfm	Child number for nonmtchfm if nonmtchfm non-missing. Values: 1 Child a, 2 Child b, etc
ca_childno_nmm	Child number for nonmtchm if nonmtchm non-missing. Values: 1 Child a, 2 Child b, etc
cW_childno_m*	Child number for m_pidp. Values: 1 Child a, 2 Child b, etc
cW_childno_f*	Child number for f_pidp. Values: 1 Child a, 2 Child b, etc
ca_childno_1pg	Child number for pidp_1pg. Values: 1 Child a, 2 Child b, etc

cW_chsex *	Child's sex from annual interview data. Values: -9 Missing, child not found in annual data, 0 Inconsistent 1 Male, 2 Female
ca_eduks_dv	Child's age group by educational key stages. Values: 1 Year R: reception year (age 4), 2 KS1: years 1-2 (age 5-7), 3 KS2: years 3-6 (age 7-11), 4 KS3: years 7-9 (age 11-14), 5 KS4: years 10-11 (age 14-16), 6 KS5: years 12-13 (age 16-18)
ca_eduph_dv	Child's age group by education phase. Values: 1 Primary: reception-KS2, 2 Secondary: KS3-KS4, 3 Post-secondary: KS5
cW_aid_dv*	Address identifier – identifies children living at the same address
cW_betasch_xw*	Cross-sectional child weight, beta version
cW_betasch_xw_t	Cross-sectional tel + web child survey weight, beta version
cW_chsdqes_dv[+]	SDQ Subscale: Emotional Symptoms
cW_chsdqcp_dv[+]	SDQ Subscale: Conduct Problems
cW_chsdqha_dv[+]	SDQ Subscale: Hyperactivity/Inattention
cW_chsdqpp_dv[+]	SDQ Subscale: Peer Relationship Problems
cW_chsdqps_dv[+]	SDQ Subscale: Prosocial
cW_chsdqtd_dv [+]	SDQ Total Difficulties Score
cW_augustage*	Age on 15/08/2020
cW_childdob_y*	DOB: Year, derived
pidpcorrected_c	New variable that identifies children whose reported date of birth did not match previous records but for whom it was possible to match them to their actual date of birth with high or medium certainty. Values: -8 inapplicable, 1 Strong dob match, year match, month wrong, 2 Strong dob match, month match, year wrong, 3 Medium dob match, both year and month wrong but close

*Variables present in April, July September, November 2020, January 2021, and March 2021 SCHILD files. [+]Variables in July, September and November are potentially provided by both male and female parents/guardians denoted by the _m and _f suffix. [$]Variable present in April and July SCHILD files only.

15. Contents of youth file YOUTH

15.1 July 2020 youth survey

The "cd_youth_p" file contains the answers to the paper self-completion Strengths and Difficulties Questionnaire (SDQ) that was sent out to sample members aged 10-15 in July 2020 (see Section 6.6 on the implementation of the youth survey). The file includes the identifier of the child, "pidp_c", and their answers to the 25 SDQ items. In addition, the file includes the variables documented in Table 15.1: the total SDQ score and the five subscale scores described in Section 14.2, the child's sex ("cd_chsex"), year of birth ("cd_childdob_y"), age at the date of interview ("cd_augustage"), and the address identifier ("cW_aid_dv"). The file also includes a weight called "cd_betayth_xw", which is described below in Section 17.3.

15.2 November 2020 youth survey

The "cf_youth_p" file contains the answers to the paper self-completion questionnaire that was sent out to sample members aged 10-15 in November 2020 (see Section 6.10). The November survey repeated the SDQ module along with questions about technology use and screen time, friendships and socialising, family, schoolwork, caring, health, nutrition, and exercise.

In addition to the answers from the questionnaires, the file also contains the variables documented in Table 15.1, including a longitudinal weight ("cf_betayth_lw"). See Section 17.9 for a description of the longitudinal weights.

15.3 March 2021 youth survey

The "ch_youth_p" file contains the answers to the paper self-completion questionnaire that was sent out to sample members aged 10-16 in March 2021 (see Section 6.13). The file includes the identifier of the child, "pidp_c", and their answers to the 25 Strengths and Difficulties Questionnaire (SDQ) items.

In addition, the file includes the variables documented in Table 15.1: the total SDQ score and the five subscale scores described in Section 14.2, the child's sex ("ch_chsex"), year of birth ("ch_childdob_y"), age at the date of interview ("ch_augustage"), and the address identifier ("cW_aid_dv"). The file also includes a weight called "ch_betayth_xw", which is described below in Section 17.3.

Table 15.1: Additional variables in the YOUTH file

Variable	Description
pidp_c	Cross-wave Person Identifier, child
cW_chsex	Child's sex from annual interview data. Values: -9 Missing, child not found in annual data, 0 Inconsistent 1 Male, 2 Female
cW_ypsdqes_dv	SDQ Subscale: Emotional Symptoms
cW_ypsdqcp_dv	SDQ Subscale: Conduct Problems
cW_ypsdqha_dv	SDQ Subscale: Hyperactivity/Inattention
cW_ypsdqpp_dv	SDQ Subscale: Peer Relationship Problems
cW_ypsdqps_dv	SDQ Subscale: Prosocial
cW_ypsdqtd_dv	SDQ Total Difficulties Score
cW_ageon20200815	Age on 15 August 2020
cW_childdob_y	Date of birth: Year, derived
cW_aid_dv	Address identifier
cW_betayth_xw	Cross-sectional youth weight, beta version
cW_betayth_lw*	Longitudinal youth weight, beta version

Notes: * In November 2020 and March 2021 data only (cf_betayth_lw, ch_betayth_lw).

16. Contents of serology file SEROLOGY

The "ch_serology" data file contains information from the serology collection that was part of the March 2021 survey. The file contains observations on both web respondents and telephone respondents (identified by the variable "ch_sampletyp").

The variables "ch_serolconsent", "ch_serolconsent_t", and "ch_sentkit" document whether respondents consented to be sent the serology test kit, and whether a kit was actually sent.

The COVG results are associated with the presence of antibodies due to natural infection, and the cut-off point for a positive result is a value greater than or equal to 1.0. The COVS results are associated with the presence of antibodies due to vaccination, and the cut-off point for a positive result in a value greater than or equal to 0.8.

The weights "ch_betaserol_xw" and "ch_betaserol_xw_t" are documented in Section 17 below.

Table 16.1: Variables in the SEROLOGY file

Variable	Description
pidp	Cross-wave Person Identifier
ch_received_d	Date sample arrived at Thriva lab - Day
ch_received_m	Date sample arrived at Thriva lab - Month
ch_received_y	Date sample arrived at Thriva lab - Year
ch_processed_d	Date the sample was processed by Thriva lab - Day
ch_processed_m	Date the sample was processed by Thriva lab - Month
ch_processed_y	Date the sample was processed by Thriva lab - Year
ch_covgvalue	The value of the COVG test. Values: numeric
ch_covgresult	The result of the COVG test. Values: 1 Negative, 2 Positive, 3 Void
ch_covsvalue	The value of the COVS test. Values: numeric
ch_covsresult	The result of the COVS test: Values: 1 Negative, 2 Positive, 3 Void
ch_result	Overall result from Thriva: Values: 1 Negative, 2 Positive, 3 Void
ch_serolconsent	Serology consent (web respondents). Values: 1 Yes, 2 No, 3 I have one of the health or medical conditions that means I cannot take part
ch_serolconsent_t	Consented to send serology test kit (telephone respondents). Values: 1 Yes, 2 No, 3 Did not return consent letter
ch_sentkit	Was this respondent sent a testing kit? Values: 0 No, they did not consent, 1 Yes, they were sent a kit, 2 No, they consented but have moved abroad, 3 No, they consented but we do not have an address for them
ch_sampletyp	Sample type. Values: 1 Web sample, 2 Telephone sample
ch_betaserol_xw	Cross-sectional serology weight, beta version
ch_betaserol_xw_t	Cross-sectional tel + web serology weight, beta version

17. Weighting and adjusting standard errors for sample design

As all samples other than the Northern Ireland sample had a clustered and stratified design, and as statistical software packages assume that the data are from a simple random sample, in order to estimate standard errors correctly you will need to explicitly inform the software about the clustering variable "psu" (primary sampling unit) and the stratification variable "strata". In Stata use the SVY suite of commands, in R use the SVYDESIGN commands, in SPSS use the CSPLAN commands along with ComplexSamples, in SAS use the SURVEY procedures.

Weights are also provided with the data to facilitate population inferences. If you undertake an unweighted analysis of the data, you should be clear on the assumptions that justify an unweighted analysis. A discussion of those assumptions can be found in the main *Understanding Society* user guide (Institute for Social and Economic Research 2020).[5] We now describe the different forms of weights provided with the different datasets.

17.1 Respondent cross-sectional weights

For each wave of the COVID-19 study, data users are provided with a single cross-sectional weight ("cW_betaindin_xw") that adjusts for unequal selection probabilities and differential nonresponse. The basis of this weight is an extension of the weighting strategy used in the annual survey. The COVID-19 study is treated as an instrument of the wave 10 annual interview. Hence, a weight can be calculated using the wave 10 cross-sectional analysis weight for those who completed an adult main interview ("j_indinui_xw") and an additional adjustment for differential nonresponse to the relevant wave of the COVID-19 study, conditional on wave 10 response. Details of the *Understanding Society* weights can be found in the main *Understanding Society* user guide (Institute for Social and Economic Research 2020). In addition, to maximise weighted dataset sizes, we also utilise a weight sharing strategy that enables weights to be computed for all COVID-19 survey respondents, not just those with wave 10 responses and weights.

Response to the COVID-19 study wave in question, conditional on wave 10 individual response, is first modelled using a probit model. Variable selection is undertaken via an a priori LASSO procedure. The predictors used include basic demographics, household composition, previous survey outcomes, COVID-19 survey paradata, economic variables and health variables. Second, weights for these respondents (i.e. those with wave 10 weights) are calculated as the product of their wave 10 weights and inverse of their response propensity as predicted by the probit model. Third, these weights are shared with COVID-19 study respondents without wave 10 weights (10-20% of respondents depending on the wave). The relationship between respondent characteristics from the COVID-19 study and their computed weights is modelled, and predicted weights are generated for all COVID-19

---

[5] Available at https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/weighting-guidance

study respondents who lack wave 10 weights. The final set of weights is computed by sharing the initially computed weights with respondents for whom these are missing using the nearest neighbours of the predicted weights. This last step ensures population totals are maintained, and hence that population inference is still valid.

We also bound weights to reduce variability (the degree of bounding was determined by examination of Kish's DEFF; see Valliant et al. 2013). In addition, we scale the weights to have a mean of one. Some evidence on the performance of the weights can be found in Benzeval et al. (2020, 2021). As of December 2021, these cross-sectional weights, along with the other weights provided, are released as beta-versions. Refinement of the weighting models continues, and updates with future releases of the data are likely.

The COVID-19 web survey cross-sectional weights scale respondents to the eligible population, which is the wave 10 population, adjusted for deaths, incapacity and emigration occurring between wave 10 and the start of the relevant COVID-19 web survey. That is, using "cW_betaindin_xw" will provide estimates that are representative of the population of all adults (16+) who were resident in private households in the UK at the time of wave 10, and who did not die or emigrate before the relevant web survey. The weights correct both for attrition from *Understanding Society* between wave 10 and the relevant web survey wave, and nonresponse to that web survey.

17.2 Child cross-sectional weights

April web and May 2020 telephone survey

The *Understanding Society* COVID-19 study is a survey of individuals aged 16 and older. In the April 2020 web survey, respondents were asked to report information about each child under age 18 and attending school, whom they live with and for whom they are a parent or guardian. This generates data on a sample of children and some researchers will wish to analyse it as such. We have therefore produced a child-level data set, in which each observation is a child reported by one or more adults in the *Understanding Society* COVID-19 survey (see Section 14 on the SCHILD file). In order that these data can be used to draw inferences about the population of school age children, we have developed child cross-sectional weights "ca_betasch_xw" for analysing the web only data. The phone survey should only be analysed in conjunction with the web survey. The weights for analysing the combined web & telephone data "ca_betasch_xw_t" are included in the file "ca_schild_w" and "cb_schild_t".

The general principle used to create the child cross-sectional weights is that the probability that a given child is observed in the survey depends on the probability of one or more adult parents or guardians with whom they reside being observed in the survey. For a child living with one parent or guardian this is simply the probability that the adult is observed, and so we set the child weight equal to that adult's weight. For a child living with more than one

parent or guardian, we take the minimum of the weights of the adults that report them. The logic of this is as follows: the probability that the child is observed is the probability that at least one of those adults responds to the survey. This joint probability of course depends on the correlation between the response probabilities of parents and guardians in the same household, which, though likely positive, is not easily estimated. However, the highest response probability among those adults provides a lower bound to the joint probability. As the weights are inverse probabilities, the lowest weight among those adults therefore provides an upper bound to the correct weight.

July 2020 web survey

In the July adult COVID-19 web survey, as noted in Section 14 above, parents were asked, for each of their children, about their child(ren)'s special education needs (for those aged 5-17 years) and the strengths and difficulties (for those aged 5-11 years). As for April, we have produced a child-level data file in which each observation is a child reported by one or more adults. Using the same procedure as in April, we have developed child cross-sectional weights ("cd_ betasch_xw") that are included in the "cd_schild_w" file.

September 2020 web survey

In the September adult COVID-19 web survey, as noted in Section 14 above, parents were asked, for each of their children, about their child(ren)'s return to school (for those aged 4-18 years) and their strengths and difficulties (for those aged 5-11 years). As for April and July, we have produced a child-level data file in which each observation is a child reported by one or more adults. Using the same procedure as in April and July, we have developed child cross-sectional weights ("ce_betasch_xw") that are included in the "ce_schild_w" file.

November 2020 web and telephone surveys

Using the same procedure as described above, we have developed child cross-sectional weights ("cf_betasch_xw") that are included in the "cf_schild_w" file for analysing the November web survey data. As noted in Section 14 above, there is also a child data file based on the November telephone survey. This should only be used in conjunction with the web survey data. The weight that should be used to combine the web and telephone child datasets for November is "cf_betasch_xw_t", which can be found in cf_schild_w for the web children sample and cf_schild_t for the telephone children sample.

January 2021 web survey

In the January 2021 adult COVID-19 web survey parents were asked about home schooling, for each of their children aged 4-18. As for the previous waves we have produced a child-level data file in which each observation is a child reported by one or more adults. Using the same procedure as for the April web and May 2020 telephone survey, we have developed child cross-sectional weights ("cg_betasch_xw") that is included in the "cg_schild_w" file.

March 2021 web survey

In the March 2021 adult COVID-19 web survey parents of 5-11 year olds were asked the Strengths and Difficulties Questionnaire (SDQ). As for the previous waves we have produced a child-level data file in which each observation is a child reported by one or more adults. Using the same procedure as for the April web and May 2020 telephone survey, we have developed child cross-sectional weights ("ch_betasch_xw") that is included in the "ch_schild_w" file.

17.3 Youth cross-sectional weights

July 2020 youth survey

In July 2020 we also asked children aged 10-15 years to complete the SDQ questionnaire themselves through a paper questionnaire. These data are provided in the "cd_youth_p" file. This file also includes a youth cross-sectional weight ("cd_betayth_xw"), which supports inferences about the UK youth population in a similar manner to how the adult web-only ("cW_betaindin_xw") cross-sectional weights support inferences about the UK adult population.

We used analogous methods to those used for the adult weights to produce "cd_betayth_xw". Given youth eligible sample selection (see Section 6.6 and Section 15), as a basis for the weight we utilised the individual's cross-sectional enumeration weight "j_psnenui_xw" at wave 10. Next, to provide COVID-19 study weights for youths without enumeration weights, we shared the extant enumeration weights. We could not apply the weight sharing method used for the adults, because the youth questionnaire included only the SDQ questions that are too limited to predict weights. Instead, we modelled the relationship between the individual's characteristics measured in the annual UKHLS survey and enumeration weights and then created predicted enumeration weights for all individuals. We then computed a new set of enumeration weights by sharing the extant weights of nearest neighbours in terms of the predicted weights with individuals who did not have a wave 10 enumeration weight.

Following this procedure, we modelled response to the COVID-19 youth survey conditional on the subject having a (shared) enumeration weight. We used the predicted response probabilities to compute the final set of weights as with the adult weights. Since youth do not complete the annual individual interview, the range of potential variables to include in the weighting models is more limited and we only included a reduced selection of socio-demographic characteristics from the household grid and as well as predictors from the household questionnaire. Some individuals however lack even this information (~80-90 in each wave), meaning that we were unable to produce weights for them. These respondents are assigned zero "cd_betayth_xw" values in the data file.

November 2020 youth survey

In November 2020 we sent another paper self-completion survey to children aged 10-15 years. This included the SDQ questions plus additional modules. These data are provided in the "cf_youth_p" file. This file also includes a youth cross-sectional weight ("cf_betayth_xw"), which is produced using the same methods as for the cross-sectional weight supplied for the July 2020 survey, and similarly scales respondents up to the UK youth population.

March 2021 youth survey

In March 2021 we sent another paper self-completion survey to children aged 10-16 years, which included the SDQ questions. These data are provided in the "ch_youth_p" file. This file also includes a youth cross-sectional weight ("ch_betayth_xw"), which is produced using the same methods as for the cross-sectional weight supplied for the July 2020 survey, and similarly scales respondents up to the UK youth population.

17.4 xbaseline data file cross-sectional weights

A "xbaseline" data file is also released including COVID-19 survey respondent answers to questions about their characteristics prior to the pandemic in Jan / Feb 2020 that were presented to all first-time respondents (see Section 13). A cross-sectional weight ("bl_betaindin_xw") is included to support inferences about the same population as the adult web-only survey ("cW_betaindin_xw") weights. The same methods are used to produce this weight as for the "cW_betaindin_xw" weights, utilising the COVID-19 survey wave 1 set of eligible individuals. The only difference in the estimation of the two types of weight is the set of respondents (here, all respondents to any of the first four COVID-19 surveys, or the May telephone survey).

17.5 Cross-sectional weights for using the May 2020 telephone data with the April or May 2020 web data

To facilitate the use of the telephone responses, the April and May waves of COVID-19 web survey come with second "web plus telephone" cross-sectional weights ("cW_betaindin_xw_t"), which should be used in place of the standard "web only" weight ("cW_betaindin_xw") if the telephone responses are used as well as the web responses. These two weights are also provided on the relevant telephone data file ("cb_indresp_t") for the telephone respondents.

The web plus telephone ("cW_betaindin_xw_t") weights support inferences about the same population as the standard web-only ("cW_betaindin_xw") weights, and they are derived in the same way. The only difference in the estimation of the two types of weights is the set of respondents.

17.6 Cross-sectional weights for using the Nov 2020 telephone data with the Nov 2020 web data

We also similarly produce a cross-sectional weight for the combined Nov 2020 web and telephone responses ("cf_betaindin_xw_t"), which should be used in place of the standard "web only" weight ("cf_betaindin_xw") if the telephone responses are used as well as the web responses. Again, the weight is also provided on the relevant telephone data file, supports inferences about the same population as the standard web-only ("cW_betaindin_xw") weights, and is derived in the same way.

17.7 Serology dataset cross-sectional weights

The "ch_serology" file includes cross-sectional weights for analyses of the serology test dataset collected at the time of wave 8 in March 2021. There are two weights, one for web only respondents ("ch_betaserol_xw"), and one for web plus telephone respondents ("ch_betaserol_xw_t"). These weights scale individuals returning a valid test to the UK adult population eligible for the COVID-19 surveys. The weight "ch_betaserol_xw" is intended for analyses of the serology data in combination with the March 2021 web survey data (ch_indresp_w file); the weight "ch_betaserol_xw_t" is intended for analyses of the entire serology data set, in combination with information available for the entire sample from the web and telephone surveys, but not restricted to any particular wave. The weights are produced similarly to the other cross-sectional weights.

17.8 Respondent longitudinal weights

A key feature of the COVID-19 survey is that it is longitudinal, enabling individuals to be tracked over the course of the pandemic. Given this, for the waves 2 to 9 main adult COVID-19 survey data files (the cW_indresp_w files), longitudinal weights ("cW_betaindin_lw") are provided. Such weights are produced for individuals that respond to focal sets of all waves leading up to and including the final, target wave: for example, the wave 3 weight is computed for individuals who respond to all of waves 1, 2, and 3. They enable the mentioned individuals to be used to make inferences about the same UK adult population as the cross-sectional weights.

An adaptation of our methods for cross-sectional weights was used to produce the longitudinal weights. Rather than a single (LASSO selected) probit model of conditional survey response being utilised in combination with the wave 10 weight to compute the weight, chains of models of response in the current wave conditional on response to all previous waves were used. Specifically, the first model in the chain was of wave 1 response conditional on having a wave 10 UKHLS cross-sectional weight, and an interim weight was computed using the wave 10 weights and the model estimates. The second model in the chain was of wave 2 response conditional on wave 1 and UKHLS wave 10 response, and another weight was computed using the w1 interim weights and these model estimates. For wave 2, this was the weight then shared with respondents lacking wave 10 weights to produce the final longitudinal weight set: for wave, 3 another model of response conditional

on wave 2 and wave 1 and UKHLS wave 10 response was required before weights could be computed and shared, and so on for further waves.

When producing the longitudinal weights, the set of eligible individuals at the final, target wave was used as the set of eligible individuals for all response models in the chain. Hence, chains of models computed to produce final weights differed slightly depending on the final, target wave. Another variation on the methods used to produce cross-sectional weights was that for wave 2 response propensity models onward it was possible to utilise COVID-19 survey items as auxiliary variables, i.e. to use answers from previous wave COVID-19 surveys as predictors of subsequent wave unit (non-)response.

Users may wish to analyse data from combinations of COVID-19 survey waves other than those for which the above longitudinal weights were designed. For example, an analysis may utilise data from waves 1, 3 and 5. In these cases, the provided weights for the final, target wave will support valid inference, though from a potentially smaller sample: some (excluded from the analysis) individuals may have responded only to waves 1, 3, and 5. The alternative is for the user to utilise the information provided in this section to produce their own set of custom weights, based on a chain of models of response to wave 1 conditional on response to UKHLS wave 10, response to wave 3 conditional on response to wave 1 and UKHLS wave 10, and response to wave 5 conditional on response to wave 3 and wave 1 and UKHLS wave 10.

17.9 Longitudinal weights for using the Nov 2020 telephone data with the Nov 2020 web data

We also produce similar longitudinal weights for analyses of the combined November 2020 web and telephone survey data files. With these, response to the November 2020 (wave 6) web and telephone survey is modelled conditional on response to the earlier April / May 2020 web and telephone surveys (see Section 6). We produce two sets of weights here, in recognition of the fact that the earlier telephone survey may be combined with either the wave 1 or the wave 2 web surveys: respectively, these weights are called "cf_betaindin_lw_t1" and "cf_betaindin_lw_t2". They should be used in place of the standard "web only" weights ("cf_betaindin_lw") if the telephone responses are used as well as the web responses. The weights are also provided on the relevant telephone data file, and support inferences about the same population as the standard web-only cross-sectional weights ("cW_betaindin_xw").

17.10 Youth longitudinal weights

In addition, we produce longitudinal weights for analyses of the youth survey data (the "cW_youth_p" files). These weights, included in the cf_youth_p and ch_youth_p files and respectively named "cf_betayth_lw" and "ch_betayth_lw", scale individuals responding to both or all three youth survey waves to the same population as the youth cross-sectional

weights ("cW_betayth_xw"), and enable inferences to be made about that population. They are produced analogously to the weights described in the previous two sections, i.e. by modelling response to the wave two survey conditional on response to the wave one survey, using the same (reduced compared to the adult survey) set of auxiliary variables as for the youth cross-sectional weights (see section 17.3).

18. *Understanding Society* data from 2019

As discussed in Section 21, the *Understanding Society* COVID-19 data can be linked to the End User Licence (EUL) version of the annual *Understanding Society* data (SN 6614), which are also available from the UKDS. Wave 9 is the latest available wave of the annual interviews, with fieldwork conducted in 2017 to early 2019.

In order to bridge the gap between the latest data from the annual interviews and the COVID-19 data, we have made available interim data from the mainstage survey for households issued for interviews in 2019. These are from the wave 10 year 2 and the wave 11 year 1 samples. It is worth noting that a in some instances, even if a household is allocated and issued to a particular month, the resulting interview may not be obtained/completed until several months after the month of issue has passed. As a result, the datafiles also include some interviews conducted in 2020: 92% of the sample issued in 2019 were also interviewed in 2019, with 8% interviewed in the first few months of 2020. The 2019 data released with the COVID-19 study contain only sample members who have responded to the COVID-19 study plus other members of their households. That is, they do not contain all respondents who have completed *Understanding Society* interviews in 2019.

It should be noted that these data are interim. Since the interim data were prepared in a short time-frame, there may be unforeseen data issues although every effort has been made to limit these. The full wave 10 data have in the meantime also been released via the UK Data Service (SN6614) and the wave 11 data will be made available in November 2021.

18.1 Naming conventions for the 2019 annual interview data

Although the data are from the wave 10 and wave 11 interviews, the data are released as a combined dataset. All variables, with the exception of personal identifiers, have the wave prefix "jk_", denoting the combination of waves 10 (j) and 11 (k). All data files are named "jk_filename_cv".

18.2 Missing Values for the 2019 annual interview data

As the 2019 dataset is a combination of wave 10 and wave 11, there are situations when a variable may exist in one wave and not the other. In such cases the value -12 is used to indicate that the respondent is in wave 10 but the variable is only applicable to wave 11. The value -13 is used for the reverse situation.

18.3 Data files for the 2019 annual interview data

The following data files are provided with the 2019 dataset:
- jk_child_cv
- jk_egoalt_cv
- jk_hhresp_cv
- jk_hhsamp_cv
- jk_indall_cv
- jk_indresp_cv
- jk_indsamp_cv
- jk_newborn_cv
- jk_parstyle_cv
- jk_youth_cv

Note that not all of the files usually released with the annual *Understanding Society* data have been released with the 2019 dataset: the derived crosswave files have not been released, nor has the call record file. The standard full set of files will be made available as per usual with the official release of waves 10 and 11 respectively.

18.4 Data file content and supporting documentation

The files within the 2019 dataset and their structure are the same as for the data released in the annual Understanding Society study (SN 6614). The variables within each of the 2019 files can be identified (without downloading the data) from the Data Dictionary available from the UKDS here:
https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8644#!/documentation.

Information on the files, how they can be used together, and all other principles of the survey structure and usage can therefore be ascertained from the existing main User Guide which is available online here:
https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/

Additional useful annual Understanding Society survey documentation can also be found on the Understanding Society website, most notably the wave 10 and 11 questionnaires which can be accessed from here:
https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires

18.5 Derived income variables

The data include a set of derived income variables for the year 2019, which are included in the file "jk_indresp_cv". These derived income variables have been created by combining data from the mainstage survey for the year 2019 (namely the second half of wave 10 and

the first half of wave 11) and parameters of the tax and national insurance system, which are available externally.

The 2019 derived income variables include (net and gross) monthly employee earnings from main job and income from self-employment. The release also includes a variable, "jk_paidwork", which contains the employment status of the respondent in the week before the interview with categories: employee, self-employee, unemployed, and missing.

The 2019 derived income variables differ from the ones in the main *Understanding Society* release in one key dimension: missing values have not been imputed in the 2019 versions. For this reason, there are, as of yet, no variables containing household income estimates. The complete set of *Understanding Society* income variables - including household income estimates - will become available as part the main *Understanding Society* wave 10 and wave 11 releases in 2020 and 2021, respectively.

Complete details of how *Understanding Society* derived income variables are constructed, alongside data quality assessments, are available in Fisher et al. (2019). An overview of *Understanding Society* income variables is also available in the main *Understanding Society* User Guide (Institute for Social and Economic Research 2020).

Table 18.1: 2019 Derived Income variables in the 2019 annual interview data

Variable name	Description	Main survey variable equivalent
jk_paygu_dv	Usual gross pay per month	W_paugu_dv
jk_paynu_dv	Usual net pay per month	W_paynu_dv
jk_payg_dv	Gross pay per month: last payment.	W_*paygl*
jk_payn_dv	Net pay per month: last payment.	W_*paynl*
jk_payu_dv	Usual pay per month if differs from last.	W_payu
jk_seearngrs_dv	Self-employment earnings – gross	W_seearngrs_dv
jk_seearnnet_dv	Self-employment earnings – net	W_seeartnnet_dv
Jk_paidwork	Employment status in week before interview	derived from main survey variables: *jbhas*, *jboff*, *jbsemp.*
Jk_topcoded	Income amount(s) top coded indicator	n/a

Income amounts have been top-coded in order to prevent disclosure. The top coding thresholds have been set to match those as set out in Section 11.5 (each income source is top-coded at a monthly equivalent of £4,000 net per week – or where a variable is gross, a gross equivalent is used). Individuals for whom any of the labour income or earnings amounts have been top coded can be identified with the variable "jk_topcoded".

The derived income variables have the same variable names as their counterparts in the main *Understanding Society* release, but with the addition of the "jk_" prefix. The derived income variables are summarised in Table 18.1.

19. Geographical *Understanding Society* COVID-19 datasets

Datasets containing the *Understanding Society* COVID-19 geographical identifiers LSOA 2011 (SN 8663) and Local Authority District (SN 8664) are available from the UK Data Service (see Section 4 for access details). These datasets can be used to link the annual *Understanding Society* COVID-19 dataset (SN 8644) to the LSOA 2011 or Local Authority District location for each respondent.

19.1 Geographical datasets for the COVID-19 web and telephone surveys

A geographical lookup file is provided for every wave, for both the web and telephone parts of the survey. Each file contains pidp (the linking variable) and the appropriate geographical identifier.

The geographical identifiers of the respondents have been calculated by looking up their postcode (not available in the released data) against the February 2020 edition of the ONS Postcode Directory for Waves 1 to 7 and against the February 2021 edition for Waves 8 and 9.

19.2 Geographical datasets for the 2019 annual interview data

A single geographical lookup file is provided for the 2019 dataset. The file contains an entry for every respondent in the main *Understanding Society* COVID-19 survey. Four variables are provided:
- pidp – cross-wave person identifier
- j_hidp – household identifier for wave 10
- k_hidp – household identifier for wave 11
- jk_lsoa11 or jk_oslaua – the geographical identifier

Linkages from the 2019 dataset can therefore be through the person identifier (pidp) or household identifier (w_hidp). As any given respondent (identified by pidp) can only be in either wave 10 or wave 11 in the 2019 dataset, j_hidp will contain either the household identifier for the wave or inapplicable (-8) if they are in wave 11, and vice versa for k_hidp.

The geographical identifiers of the respondents have been calculated by looking up their wave 10 postcode (not available in the released data) against the May 2019 version of the ONS Postcode Directory or their wave 11 postcode against the February 2020 edition as appropriate.

20. *Understanding Society* COVID-19 School Codes dataset

A dataset containing the *Understanding Society* COVID-19 School Codes (SN 8730) is available from the UK Data Service (see Section 4 for access details). The dataset consists of two data files that enable identification of children's schools in the COVID-19 study.

20.1 School data for the COVID-19 web survey

School identification information for children was requested from both parents in the wave 5 web survey. These data have been released in the file "ce_schcodest_w" that includes the variables documented in Table 20.1.

Table 20.1: Variables in the COVID-19 School Data file (ce_schcodest_w)

Variable	Description
pidp_c	Cross-wave Person Identifier, child
ce_schcodest_m	School reference number, mother
ce_countrycode_m	Country of school, mother
ce_schcodest_f	School reference number, father
ce_countrycode_f	Country of school, father
ce_schnamecpt_m	Private school name, mother
ce_schtowncpt_m	Private school town, mother
ce_schnamecpt_f	Private school name, father
ce_schtowncpt_f	Private school town, father

"pidp_c" is the personal identifier that should be used to link "ce_schcodest_w" to the child file "ce_schild_w".

The school reference number variables contain the official school codes of the child's school. This consists of the DfE number in England and Wales, the SEED number in Scotland and the Department of Education number in Northern Ireland. These codes can be used to link to external data available on schools.

The variables for private schools consist of free text fields exactly as entered by survey respondents. It should be noted that private school identification information is not available for the COVID-19 2019 dataset as it is not currently available for any mainstage *Understanding Society* wave.

As the school information was asked of both parents, there are two versions of each variable: the report from the mother (*_m) and the report from the father(*_f).

20.2 School data for the 2019 annual interview data

In the mainstage *Understanding Society* survey, school identification information for children is requested from one parent. These data are only collected in odd numbered waves of the annual interviews. Since the 2019 COVID-19 dataset is comprised of elements of Waves 10 and 11, child school data are only available for the Wave 11 sample members.

These data have been released in the file "jk_schcodest_cv" that includes the variables documented in Table 20.2.

Table 20.2: Variables in the *Understanding Society* 2019 School Data file (jk_schcodest_cv)

Variable	Description
pidp	Cross-wave Person Identifier, child
k_hidp	Household identifier
jk_schcodest	School reference number
jk_countrycode	Country of school

"pidp" is the personal identifier that should be used to link "jk_schcodest" to other files in the 2019 dataset. "k_hidp" can be used to link to household level files in the 2019 dataset.

The school reference number variable contains the official school codes of the child's school. This consists of the DfE number in England and Wales, the SEED number in Scotland and the Department of Education number in Northern Ireland. These codes can be used to link to external data available on schools.

Data identifying private schools is currently not available in the 2019 data files.

21. Linking to other *Understanding Society* datasets

The data in the *Understanding Society* COVID-19 study can be linked to the annual *Understanding Society* study available from the UK Data Service.

- The End User Licence (EUL) version, SN 6614, can be found here, however, please refer to the *Understanding Society* website first if you require more information on it.
- There is also a Special Licence version of the annual survey data (SN 6931) which contains additional variables not available in the in the EUL version (such as month of birth, detailed country of birth variable etc), and non-top-coded versions of income and earnings variables.
- Special Licence files to identify different levels of geographies for survey members in the annual study are also available from the UK Data Service. LSOA 2011 and Local Authority District files specifically produced for the *Understanding Society* COVID-19 study are available (see Section 19 for more details). Note that access to Special Licence datasets requires users to apply for the data via a process which requires them to justify why access is required before approval is granted. It should also be noted that other restrictions may also have to be met.
- Special Licence files to identify children's schools from the annual interviews (SN 7182) are also available from the UK Data Service.

As detailed in Section 18, the 2019 data from the annual *Understanding Society* dataset has been supplied with the COVID-19 data. The 2019 dataset has been especially prepared in order to provide the latest available annual survey data for the COVID-19 respondents, taken from either wave 10 or wave 11, and released here in advance of being available in SN 6614. The 2019 data allows researchers to bridge the gap to the currently available wave of the main study data, wave 9, which ended in May 2019. Researchers can therefore link the main *Understanding Society* COVID-19 data to the 2019 data and then to waves 1 to 9 already released in SN 6614.

Linkage between the *Understanding Society* COVID-19 study and the annual *Understanding Society* study at the respondent level is made by using the Cross-wave Person Identifier variable "pidp". The key household identifier variables, "W_hidp", which uniquely identify households within waves, are made available in the COVID-19 survey data files and can be used to link to households in the either the 2019 dataset or the annual survey data (SN 6614).

22. Known data issues and corrections

The speed at which the *Understanding Society* COVID-19 study was deployed, and data subsequently released, has meant that we were not able to do the amount of pre-testing that we would normally do prior to fielding a questionnaire, nor the amount of checking we would normally do prior to release. However, our quality assurance activities continue after the first release of the data.

22.1 Routing problems in the employment module

In continuing to check the data we have discovered some routing problems in the employment module, beginning with Wave c (June). These are of two types:

i)   There is subtle flaw in the way a feed-forward is specified that is only manifest after 2 waves of data collection.

ii)   There was a problem with the way the field agency's scripting handled non-integer responses to hours questions.

The impact of these two issues is that some respondents were not asked a question they should have, while other were asked a question they should been routed past.

A few variables which we now deem unfit for purpose have therefore been dropped from the released data (see Table 22.1 for the complete list).

For the "hourschange*" variables which were also affected by the routing, but to a lesser extent, we have released the variables but made corrections to reflect the errors in routing: where a respondent should have been asked a question but was not, we recode the variable to -9 "Missing" (where previously it was -8 "Inapplicable"). For the same variables, if a responded was asked a question they should been routed out of, we have recoded to -8 "Inapplicable".

The variables and waves affected, and the corrections we have made are summarized in the following Table.

Table 22.1: Variables dropped or corrected in the INDRESP files

Variable	Waves affected	Action taken
hrschange1, 2 and 3	c, d, e	recoded
hrschangeup1	d, e	recoded
empchange	d, e	withdrawn
jbindustry	e	withdrawn
		NB: fine at waves c, d

22.2 Occupation

The coding of occupation in surveys draws on multiple pieces of information (job title, industry, firm size, managerial duties) and even using automated coding CASCOT software requires numerous judgements by human coders. The fieldwork for the COVID-19 web surveys – and subsequent occupation coding – is done by a different agency than the main annual interview. Moreover, the main survey annual interviewing makes use of dependent interviewing in collecting occupation and this was not possible in the COVID-19 web surveys. This has led to differences in coded occupations and may have led to potentially spurious occupation changes between pre-COVID and COVID period, which require further investigation. We have, therefore, withdrawn these variables collected in the employment module, while we investigate and we will update users at the next COVID-19 data release. The variables temporarily withdrawn are:

ce_jbsoc00_cc, ce_jbmngr, ce_jbsize

We remind users that, past and future, labour market and occupational histories for this period will subsequently be captured by the main annual interview, and we have released data from 2019 interviews (waves 10 and 11) for pre-pandemic measures (see Section 18).

Note however that in the first two waves of Understanding Society the occupation description was first coded to the code frame of SOC 2000. From Wave 3 onward, descriptions of new jobs were first coded to the code frame of SOC 2010 and the SOC 2000 was computed from a look-up file. The SOC 2010 code is currently unavailable for Wave 1 and Wave 2. SOC 2010 codes are [-9] missing for jobs that were already held in Wave 1 or 2 and (re-)confirmed in later waves.

22.3 Pregnancy in the November 2020 telephone survey (wave 6)

In the long-term health conditions module there was a script error in the question "pregnow". This meant the nine people who should have been asked the question were not. These cases have been set to -9 "Missing".

22.4 ff_sempgovt in the November 2020 web survey (wave 6)

The feed-forward variable ff_sempgovt (Government support for self-employed) is based on the question sempgovt: *"Have you been told by HMRC that you are eligible for government support for the selfemployed?" (1 Yes 2 No).* The original specification for ff_sempgovt was:

0. Not eligible for government support: IF sempgovt is not 1 in any of the previous survey months
1. Eligible for government support: IF sempgovt = 1 in any of the previous survey months

In September the question sempgovt was replaced by sempgovt2: *"Have you received government support for the self-employed?" (1 Yes, in the first round up to 14 July 2 Yes, in the second round from 14 July 3 Yes in both rounds 4 No)*. However the specification of ff_sempgovt was erroneously not updated. As a result, in the November web survey ff_sempderived took value '1' if sempgovt or sempgovt2 had taken value 1 in any of the previous surveys. The correct specification would have been for ff_sempgovt to take value 1 if sempgovt = 1 in any previous survey or sempgovt2 = 1, 2, 3 in any previous survey.

22.5 Feed-forward variables in the November 2020 web and telephone surveys (wave 6)

As documented in Section 11.2, the feed-forward variables are updated each wave: the answers of respondents who completed the full survey are incorporated into the feed-forward variables for the next wave. For sample members who only completed the survey partially or not at all, the feed-forward variables remain unchanged. However, in wave 6 this was done inconsistently: the feed-forward variables were updated for all respondents, including those who only completed a partial survey in wave 5. The affected 84 cases can be identified by merging the variable ce_outcome from the xsample file to the cf_indresp_w file, to identify wave 5 partial respondents who completed the wave 6 survey.

22.6 Tutoring in the November 2020 telephone survey (wave 6)

In the schooling module, the question "tutoring" was specified as a 'select all that apply' question in the web survey. In the telephone survey it was erroneously specified as a single choice question.

References

Benzeval, M., Burton, J., Crossley, T.F., Fisher, P., Jäckle, A., Low, H. and Read, B. (2020) The Idiosyncratic Impact of an Aggregate Shock: The Distributional Consequences of COVID-19. *Understanding Society Working Paper 2020-09*, Colchester: University of Essex.

Benzeval, M., Burton, J., Crossley, T.F., Fisher, P., Gardiner, C., Jäckle, A. and Moore, J. (2021) High Frequency Online Data Collection in an Annual Household Panel Study: Some Evidence on Bias Prevention and Bias Adjustment. *Understanding Society Working Paper 2021-03*, Colchester: University of Essex.

Berthoud, R., Fumagalli, L., Lynn, P. and Platt, L. (2009) Design of the Understanding Society Ethnic Minority Boost Sample, *Understanding Society Working Paper 2009-02*, Colchester: University of Essex.

Fisher, P., Fumagalli, L., Buck, N. and Avram, S. (2019) Understanding Society and its income data*. Understanding Society Working Paper 2019-08*, Colchester: University of Essex.

Institute for Social and Economic Research (2020) *Understanding Society*: Waves 1-9, 2009-2018, and Harmonised BHPS: Waves 1-18, 1991-2009, User Guide. Colchester: University of Essex. https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/ [Accessed 26/05/2020].

Lynn, P. (2009) Sample design for Understanding Society, *Understanding Society Working Paper 2009-01*. Colchester: University of Essex.

Lynn, P., Nandi, A., Parutis, V. and Platt, L. (2017) Design and Implementation of a High Quality Probability Sample of Immigrants and Ethnic Minorities: Lessons Learnt. *Understanding Society Working Paper 2017-11*. Colchester: University of Essex.

Valliant, R., Dever, J.A. and Kreuter, F. (2013) *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer, 2013.

Appendix A: SOC 2010 code frame

Code	Description
1	**MANAGERS, DIRECTORS AND SENIOR OFFICIALS**
11	**CORPORATE MANAGERS AND DIRECTORS**
111	**Chief Executives and Senior Officials**
1115	Chief executives and senior officials
1116	Elected officers and representatives
112	**Production Managers and Directors**
1121	Production managers and directors in manufacturing
1122	Production managers and directors in construction
1123	Production managers and directors in mining and energy
113	**Functional Managers and Directors**
1131	Financial managers and directors
1132	Marketing and sales directors
1133	Purchasing managers and directors
1134	Advertising and public relations directors
1135	Human resource managers and directors
1136	Information technology and telecommunications directors
1139	Functional managers and directors n.e.c.
115	**Financial Institution Managers and Directors**
1150	Financial institution managers and directors
116	**Managers and Directors in Transport and Logistics**
1161	Managers and directors in transport and distribution
1162	Managers and directors in storage and warehousing
117	**Senior Officers in Protective Services**
1171	Officers in armed forces
1172	Senior police officers
1173	Senior officers in fire, ambulance, prison and related services
118	**Health and Social Services Managers and Directors**
1181	Health services and public health managers and directors
1184	Social services managers and directors
119	**Managers and Directors in Retail and Wholesale**
1190	Managers and directors in retail and wholesale
12	**OTHER MANAGERS AND PROPRIETORS**
121	**Managers and Proprietors in Agriculture Related Services**
1211	Managers and proprietors in agriculture and horticulture
1213	Managers and proprietors in forestry, fishing and related services
122	**Managers and Proprietors in Hospitality and Leisure Services**
1221	Hotel and accommodation managers and proprietors
1223	Restaurant and catering establishment managers and proprietors
1224	Publicans and managers of licensed premises
1225	Leisure and sports managers
1226	Travel agency managers and proprietors
124	**Managers and Proprietors in Health and Care Services**
1241	Health care practice managers
1242	Residential, day and domiciliary care managers and proprietors
125	**Managers and Proprietors in Other Services**
1251	Property, housing and estate managers
1252	Garage managers and proprietors
1253	Hairdressing and beauty salon managers and proprietors
1254	Shopkeepers and proprietors – wholesale and retail
1255	Waste disposal and environmental services managers

1259	Managers and proprietors in other services n.e.c.
2	**PROFESSIONAL OCCUPATIONS**
21	**SCIENCE, RESEARCH, ENGINEERING AND TECHNOLOGY PROFESSIONALS**
211	**Natural and Social Science Professionals**
2111	Chemical scientists
2112	Biological scientists and biochemists
2113	Physical scientists
2114	Social and humanities scientists
2119	Natural and social science professionals n.e.c.
212	**Engineering Professionals**
2121	Civil engineers
2122	Mechanical engineers
2123	Electrical engineers
2124	Electronics engineers
2126	Design and development engineers
2127	Production and process engineers
2129	Engineering professionals n.e.c.
213	**Information Technology and Telecommunications Professionals**
2133	IT specialist managers
2134	IT project and programme managers
2135	IT business analysts, architects and systems designers
2136	Programmers and software development professionals
2137	Web design and development professionals
2139	Information technology and telecommunications professionals n.e.c.
214	**Conservation and Environment Professionals**
2141	Conservation professionals
2142	Environment professionals
215	**Research and Development Managers**
2150	Research and development managers
22	**HEALTH PROFESSIONALS**
221	**Health Professionals**
2211	Medical practitioners
2212	Psychologists
2213	Pharmacists
2214	Ophthalmic opticians
2215	Dental practitioners
2216	Veterinarians
2217	Medical radiographers
2218	Podiatrists
2219	Health professionals n.e.c.
222	**Therapy Professionals**
2221	Physiotherapists
2222	Occupational therapists
2223	Speech and language therapists
2229	Therapy professionals n.e.c.
223	**Nursing and Midwifery Professionals**
2231	Nurses
2232	Midwives
23	**TEACHING AND EDUCATIONAL PROFESSIONALS**
231	***Teaching and Educational Professionals***
2311	Higher education teaching professionals
2312	Further education teaching professionals
2314	Secondary education teaching professionals
2315	Primary and nursery education teaching professionals

2316	Special needs education teaching professionals
2317	Senior professionals of educational establishments
2318	Education advisers and school inspectors
2319	Teaching and other educational professionals n.e.c.
24	**BUSINESS, MEDIA AND PUBLIC SERVICE PROFESSIONALS**
241	**Legal Professionals**
2412	Barristers and judges
2413	Solicitors
2419	Legal professionals n.e.c.
242	**Business, Research and Administrative Professionals**
2421	Chartered and certified accountants
2423	Management consultants and business analysts
2424	Business and financial project management professionals
2425	Actuaries, economists and statisticians
2426	Business and related research professionals
2429	Business, research and administrative professionals n.e.c.
243	**Architects, Town Planners and Surveyors**
2431	Architects
2432	Town planning officers
2433	Quantity surveyors
2434	Chartered surveyors
2435	Chartered architectural technologists
2436	Construction project managers and related professionals
244	**Welfare Professionals**
2442	Social workers
2443	Probation officers
2444	Clergy
2449	Welfare professionals n.e.c.
245	**Librarians and Related Professionals**
2451	Librarians
2452	Archivists and curators
246	**Quality and Regulatory Professionals**
2461	Quality control and planning engineers
2462	Quality assurance and regulatory professionals
2463	Environmental health professionals
247	**Media Professionals**
2471	Journalists, newspaper and periodical editors
2472	Public relations professionals
2473	Advertising accounts managers and creative directors
3	**ASSOCIATE PROFESSIONAL AND TECHNICAL OCCUPATIONS**
31	**SCIENCE, ENGINEERING AND TECHNOLOGY ASSOCIATE PROFESSIONALS**
311	**Science, Engineering and Production Technicians**
3111	Laboratory technicians
3112	Electrical and electronics technicians
3113	Engineering technicians
3114	Building and civil engineering technicians
3115	Quality assurance technicians
3116	Planning, process and production technicians
3119	Science, engineering and production technicians n.e.c.
312	**Draughtspersons and Related Architectural Technicians**
3121	Architectural and town planning technicians
3122	Draughtspersons
313	**Information Technology Technicians**
3131	IT operations technicians

3132	IT user support technicians
32	**HEALTH AND SOCIAL CARE ASSOCIATE PROFESSIONALS**
321	**Health Associate Professionals**
3213	Paramedics
3216	Dispensing opticians
3217	Pharmaceutical technicians
3218	Medical and dental technicians
3219	Health associate professionals n.e.c.
323	**Welfare and Housing Associate Professionals**
3231	Youth and community workers
3233	Child and early years officers
3234	Housing officers
3235	Counsellors
3239	Welfare and housing associate professionals n.e.c.
33	**PROTECTIVE SERVICE OCCUPATIONS**
331	**Protective Service Occupations**
3311	NCOs and other ranks
3312	Police officers (sergeant and below)
3313	Fire service officers (watch manager and below)
3314	Prison service officers (below principal officer)
3315	Police community support officers
3319	Protective service associate professionals n.e.c.
34	**CULTURE, MEDIA AND SPORTS OCCUPATIONS**
341	**Artistic, Literary and Media Occupations**
3411	Artists
3412	Authors, writers and translators
3413	Actors, entertainers and presenters
3414	Dancers and choreographers
3415	Musicians
3416	Arts officers, producers and directors
3417	Photographers, audio-visual and broadcasting equipment operators
342	**Design Occupations**
3421	Graphic designers
3422	Product, clothing and related designers
344	**Sports and Fitness Occupations**
3441	Sports players
3442	Sports coaches, instructors and officials
3443	Fitness instructors
35	**BUSINESS AND PUBLIC SERVICE ASSOCIATE PROFESSIONALS**
351	**Transport Associate Professionals**
3511	Air traffic controllers
3512	Aircraft pilots and flight engineers
3513	Ship and hovercraft officers
352	**Legal Associate Professionals**
3520	Legal associate professionals
353	**Business, Finance and Related Associate Professionals**
3531	Estimators, valuers and assessors
3532	Brokers
3533	Insurance underwriters
3534	Finance and investment analysts and advisers
3535	Taxation experts
3536	Importers and exporters
3537	Financial and accounting technicians
3538	Financial accounts managers

3539	Business and related associate professionals n.e.c.
354	**Sales, Marketing and Related Associate Professionals**
3541	Buyers and procurement officers
3542	Business sales executives
3543	Marketing associate professionals
3544	Estate agents and auctioneers
3545	Sales accounts and business development managers
3546	Conference and exhibition managers and organisers
355	**Conservation and Environmental Associate Professionals**
3550	Conservation and environmental associate professionals
356	**Public Services and Other Associate Professionals**
3561	Public services associate professionals
3562	Human resources and industrial relations officers
3563	Vocational and industrial trainers and instructors
3564	Careers advisers and vocational guidance specialists
3565	Inspectors of standards and regulations
3567	Health and safety officers
4	**ADMINISTRATIVE AND SECRETARIAL OCCUPATIONS**
41	**ADMINISTRATIVE OCCUPATIONS**
411	**Administrative Occupations: Government and Related Organisations**
4112	National government administrative occupations
4113	Local government administrative occupations
4114	Officers of non-governmental organisations
412	**Administrative Occupations: Finance**
4121	Credit controllers
4122	Book-keepers, payroll managers and wages clerks
4123	Bank and post office clerks
4124	Finance officers
4129	Financial administrative occupations n.e.c.
413	**Administrative Occupations: Records**
4131	Records clerks and assistants
4132	Pensions and insurance clerks and assistants
4133	Stock control clerks and assistants
4134	Transport and distribution clerks and assistants
4135	Library clerks and assistants
4138	Human resources administrative occupations
415	**Other Administrative Occupations**
4151	Sales administrators
4159	Other administrative occupations n.e.c.
416	**Administrative Occupations: Office Managers and Supervisors**
4161	Office managers
4162	Office supervisors
42	**SECRETARIAL AND RELATED OCCUPATIONS**
421	**Secretarial and Related Occupations**
4211	Medical secretaries
4212	Legal secretaries
4213	School secretaries
4214	Company secretaries
4215	Personal assistants and other secretaries
4216	Receptionists
4217	Typists and related keyboard occupations
5	**SKILLED TRADES OCCUPATIONS**
51	**SKILLED AGRICULTURAL AND RELATED TRADES**
511	**Agricultural and Related Trades**

5111	Farmers
5112	Horticultural trades
5113	Gardeners and landscape gardeners
5114	Groundsmen and greenkeepers
5119	Agricultural and fishing trades n.e.c.
52	**SKILLED METAL, ELECTRICAL AND ELECTRONIC TRADES**
521	**Metal Forming, Welding and Related Trades**
5211	Smiths and forge workers
5212	Moulders, core makers and die casters
5213	Sheet metal workers
5214	Metal plate workers, and riveters
5215	Welding trades
5216	Pipe fitters
522	**Metal Machining, Fitting and Instrument Making Trades**
5221	Metal machining setters and setter-operators
5222	Tool makers, tool fitters and markers-out
5223	Metal working production and maintenance fitters
5224	Precision instrument makers and repairers
5225	Air-conditioning and refrigeration engineers
523	**Vehicle Trades**
5231	Vehicle technicians, mechanics and electricians
5232	Vehicle body builders and repairers
5234	Vehicle paint technicians
5235	Aircraft maintenance and related trades
5236	Boat and ship builders and repairers
5237	Rail and rolling stock builders and repairers
524	**Electrical and Electronic Trades**
5241	Electricians and electrical fitters
5242	Telecommunications engineers
5244	TV, video and audio engineers
5245	IT engineers
5249	Electrical and electronic trades n.e.c.
525	**Skilled Metal, Electrical and Electronic Trades Supervisors**
5250	Skilled metal, electrical and electronic trades supervisors
53	**SKILLED CONSTRUCTION AND BUILDING TRADES**
531	**Construction and Building Trades**
5311	Steel erectors
5312	Bricklayers and masons
5313	Roofers, roof tilers and slaters
5314	Plumbers and heating and ventilating engineers
5315	Carpenters and joiners
5316	Glaziers, window fabricators and fitters
5319	Construction and building trades n.e.c.
532	**Building Finishing Trades**
5321	Plasterers
5322	Floorers and wall tilers
5323	Painters and decorators
533	**Construction and Building Trades Supervisors**
5330	Construction and building trades supervisors
54	**TEXTILES, PRINTING AND OTHER SKILLED TRADES**
541	**Textiles and Garments Trades**
5411	Weavers and knitters
5412	Upholsterers
5413	Footwear and leather working trades

5414	Tailors and dressmakers
5419	Textiles, garments and related trades n.e.c.
542	**Printing Trades**
5421	Pre-press technicians
5422	Printers
5423	Print finishing and binding workers
543	**Food Preparation and Hospitality Trades**
5431	Butchers
5432	Bakers and flour confectioners
5433	Fishmongers and poultry dressers
5434	Chefs
5435	Cooks
5436	Catering and bar managers
544	**Other Skilled Trades**
5441	Glass and ceramics makers, decorators and finishers
5442	Furniture makers and other craft woodworkers
5443	Florists
5449	Other skilled trades n.e.c.
6	**CARING, LEISURE AND OTHER SERVICE OCCUPATIONS**
61	**CARING PERSONAL SERVICE OCCUPATIONS**
612	**Childcare and Related Personal Services**
6121	Nursery nurses and assistants
6122	Childminders and related occupations
6123	Playworkers
6125	Teaching assistants
6126	Educational support assistants
613	**Animal Care and Control Services**
6131	Veterinary nurses
6132	Pest control officers
6139	Animal care services occupations n.e.c.
614	**Caring Personal Services**
6141	Nursing auxiliaries and assistants
6142	Ambulance staff (excluding paramedics)
6143	Dental nurses
6144	Houseparents and residential wardens
6145	Care workers and home carers
6146	Senior care workers
6147	Care escorts
6148	Undertakers, mortuary and crematorium assistants
62	**LEISURE, TRAVEL AND RELATED PERSONAL SERVICE OCCUPATIONS**
621	**Leisure and Travel Services**
6211	Sports and leisure assistants
6212	Travel agents
6214	Air travel assistants
6215	Rail travel assistants
6219	Leisure and travel service occupations n.e.c.
622	**Hairdressers and Related Services**
6221	Hairdressers and barbers
6222	Beauticians and related occupations
623	**Housekeeping and Related Services**
6231	Housekeepers and related occupations
6232	Caretakers
624	**Cleaning and Housekeeping Managers and Supervisors**
6240	Cleaning and housekeeping managers and supervisors

7	SALES AND CUSTOMER SERVICE OCCUPATIONS
71	**SALES OCCUPATIONS**
711	**Sales Assistants and Retail Cashiers**
7111	Sales and retail assistants
7112	Retail cashiers and check-out operators
7113	Telephone salespersons
7114	Pharmacy and other dispensing assistants
7115	Vehicle and parts salespersons and advisers
712	**Sales Related Occupations**
7121	Collector salespersons and credit agents
7122	Debt, rent and other cash collectors
7123	Roundspersons and van salespersons
7124	Market and street traders and assistants
7125	Merchandisers and window dressers
7129	Sales related occupations n.e.c.
713	**Sales Supervisors**
7130	Sales supervisors
72	**CUSTOMER SERVICE OCCUPATIONS**
721	**Customer Service Occupations**
7211	Call and contact centre occupations
7213	Telephonists
7214	Communication operators
7215	Market research interviewers
7219	Customer service occupations n.e.c.
722	**Customer Service Managers and Supervisors**
7220	Customer service managers and supervisors
8	**PROCESS, PLANT AND MACHINE OPERATIVES**
81	**PROCESS, PLANT AND MACHINE OPERATIVES**
811	**Process Operatives**
8111	Food, drink and tobacco process operatives
8112	Glass and ceramics process operatives
8113	Textile process operatives
8114	Chemical and related process operatives
8115	Rubber process operatives
8116	Plastics process operatives
8117	Metal making and treating process operatives
8118	Electroplaters
8119	Process operatives n.e.c.
812	**Plant and Machine Operatives**
8121	Paper and wood machine operatives
8122	Coal mine operatives
8123	Quarry workers and related operatives
8124	Energy plant operatives
8125	Metal working machine operatives
8126	Water and sewerage plant operatives
8127	Printing machine assistants
8129	Plant and machine operatives n.e.c.
813	**Assemblers and Routine Operatives**
8131	Assemblers (electrical and electronic products)
8132	Assemblers (vehicles and metal goods)
8133	Routine inspectors and testers
8134	Weighers, graders and sorters
8135	Tyre, exhaust and windscreen fitters
8137	Sewing machinists

8139	Assemblers and routine operatives n.e.c.
814	**Construction Operatives**
8141	Scaffolders, stagers and riggers
8142	Road construction operatives
8143	Rail construction and maintenance operatives
8149	Construction operatives n.e.c.
82	**TRANSPORT AND MOBILE MACHINE DRIVERS AND OPERATIVES**
821	**Road Transport Drivers**
8211	Large goods vehicle drivers
8212	Van drivers
8213	Bus and coach drivers
8214	Taxi and cab drivers and chauffeurs
8215	Driving instructors
822	**Mobile Machine Drivers and Operatives**
8221	Crane drivers
8222	Fork-lift truck drivers
8223	Agricultural machinery drivers
8229	Mobile machine drivers and operatives n.e.c.
823	**Other Drivers and Transport Operatives**
8231	Train and tram drivers
8232	Marine and waterways transport operatives
8233	Air transport operatives
8234	Rail transport operatives
8239	Other drivers and transport operatives n.e.c.
9	**ELEMENTARY OCCUPATIONS**
91	**ELEMENTARY TRADES AND RELATED OCCUPATIONS**
911	**Elementary Agricultural Occupations**
9111	Farm workers
9112	Forestry workers
9119	Fishing and other elementary agriculture occupations n.e.c.
912	**Elementary Construction Occupations**
9120	Elementary construction occupations
913	**Elementary Process Plant Occupations**
9132	Industrial cleaning process occupations
9134	Packers, bottlers, canners and fillers
9139	Elementary process plant occupations n.e.c.
92	**ELEMENTARY ADMINISTRATION AND SERVICE OCCUPATIONS**
921	**Elementary Administration Occupations**
9211	Postal workers, mail sorters, messengers and couriers
9219	Elementary administration occupations n.e.c.
923	**Elementary Cleaning Occupations**
9231	Window cleaners
9232	Street cleaners
9233	Cleaners and domestics
9234	Launderers, dry cleaners and pressers
9235	Refuse and salvage occupations
9236	Vehicle valeters and cleaners
9239	Elementary cleaning occupations n.e.c.
924	**Elementary Security Occupations**
9241	Security guards and related occupations
9242	Parking and civil enforcement occupations
9244	School midday and crossing patrol occupations
9249	Elementary security occupations n.e.c.
925	**Elementary Sales Occupations**

9251	Shelf fillers
9259	Elementary sales occupations n.e.c.
926	**Elementary Storage Occupations**
9260	Elementary storage occupations
927	**Other Elementary Services Occupations**
9271	Hospital porters
9272	Kitchen and catering assistants
9273	Waiters and waitresses
9274	Bar staff
9275	Leisure and theme park attendants
9279	Other elementary services occupations n.e.c.

Appendix B: Stata code to create derived variables from COVID-19 data

Note: the wave prefixes in variable names for the current wave (ca_, cb_, cc_, etc) have been omitted from the code below for simplicity.

* GHQ scales: Likert Scoring (scghq1_dv) and Caseness Scoring (scghq2_dv)

```
foreach w in scghqa scghqb scghqc scghqd scghqe scghqf scghqg scghqh scghqi ///
       scghqj scghqk scghql {
       recode `w' -9/-1 = .
       gen `w'l = `w'-1 /*Likert Scoring*/
       gen `w'c = `w' /*Caseness Scoring*/
       recode `w'c 1/2 = 0 3/4 = 1
       }
```

```
gen scghq1_dv = scghqal+scghqbl+scghqcl+scghqdl+scghqel+scghqfl+scghqgl+ ///
       scghqhl+scghqil+scghqjl+scghqkl+scghqll
replace scghq1_dv = -9 if scghq1_dv ==.
lab def scghq1_dv -9 "Missing" -8 "Inapplicable" -2 "Refusal" -1 "Don't know"
lab val scghq1_dv scghq1_dv
label variable scghq1_dv "Subjective wellbeing (GHQ): Likert"
```

```
gen scghq2_dv = scghqac+scghqbc+scghqcc+scghqdc+scghqec+scghqfc+scghqgc+ ///
       scghqhc+scghqic+scghqjc+scghqkc+scghqlc
replace scghq2_dv=-9 if scghq2_dv==.
lab def scghq2_dv -9 "Missing" -8 "Inapplicable" -2 "Refusal" -1 "Don't know"
lab val scghq2_dv scghq2_dv
label variable scghq2_dv "Subjective wellbeing (GHQ): Caseness"
```

* SDQ subscales and total score

Note: the wave prefix in variable names for the current wave (cd_) and the variable suffixes indicating responses provided by a mother (_m) or a father (_f) have been omitted from the code below for simplicity.

/*SDQ Scoring*/

/* Step 1:
Treat a value of 4 which means "can't say" the same as don't know and so recode it to -1
Recode all negative values (-9 -8 -7 -2 -1) which represent missing values to system missings (.)
Adjust original scores of the composite variables by reducing these by 1

For a sub-set of variables the scale needs to be reversed to make it consistent with the other variables. In these cases reverse the value order.

Note: the Stata command revv is not one of Stata's pre-installed commands and so will need to be installed. To do that type
ssc install revv
/

```
foreach var in chsdqa chsdqb chsdqc chsdqd chsdqe chsdqf chsdqh ///
            chsdqi chsdqj chsdql chsdqm chsdqo chsdqp chsdqq ///
            chsdqr chsdqs chsdqt chsdqv chsdqw chsdqx  {
      recode `var' (4 = -1) // assume can't say is the same as don't know.
      recode `var' (-9/-1 = .)
      gen `var'r=`var'-1
}
```

```
*Reverse Scoring*
foreach var in chsdqg chsdqk chsdqn chsdqu chsdqy  {
      recode `var' 4 = -1 // assume can't say is the same as don't know.
      recode `var' -9/-1 = .
      revv `var'
      gen `var'r=rv_`var'-1
}
```

/*Creating SDQ Subscales. Note: Each subscale is computed by multiplying the average score of its composite measures by 5 (number of items in each subscale), rounded to the nearest integer. Subscales can be computed as long as there are no missings on at least 3 out of 5 items forming the scale*/

```
egen n_emotion=robs(chsdqcr chsdqhr chsdqmr chsdqpr chsdqxr)
egen chsdqes_dv =rmean(chsdqcr chsdqhr chsdqmr chsdqpr chsdqxr) ///
            if n_emotion>2 & n_emotion~=.
replace chsdqes_dv =round(chsdqes_dv*5)
```

```
egen n_conduct=robs(chsdqer chsdqgr chsdqlr chsdqrr chsdqvr)
egen chsdqcp_dv=rmean(chsdqer chsdqgr chsdqlr chsdqrr chsdqvr)  ///
            if n_conduct>2 & n_conduct~=.
replace chsdqcp_dv=round(chsdqcp_dv*5)
```

```
egen n_hyper=robs(chsdqbr chsdqjr chsdqor chsdqur chsdqyr)
egen chsdqha_dv=rmean(chsdqbr chsdqjr chsdqor chsdqur chsdqyr) ///
            if n_hyper>2 & n_hyper~=.
replace chsdqha_dv=round(chsdqha_dv*5)
```

```
egen n_peer=robs(chsdqfr chsdqkr chsdqnr chsdqsr chsdqwr)
egen chsdqpp_dv=rmean(chsdqfr chsdqkr chsdqnr chsdqsr chsdqwr) ///
            if n_peer>2 & n_peer~=.
```

```
replace chsdqpp_dv=round(chsdqpp_dv*5)

egen n_prosoc=robs(chsdqar chsdqdr chsdqir chsdqqr chsdqtr)
egen chsdqps_dv=rmean(chsdqar chsdqdr chsdqir chsdqqr chsdqtr) ///
                if n_prosoc>2 & n_prosoc~=.
replace chsdqps_dv=round(chsdqps_dv*5)
```

/*SDQ Total Difficulties Score. Note: Total difficulties score is a sum of 4 SDQ subscales: Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, Peer Relationship Problems */

```
gen chsdqtd_dv = chsdqes_dv + chsdqcp_dv + chsdqha_dv + chsdqpp_dv

label variable chsdqes_dv "SDQ Subscale: Emotional Symptoms"
label variable chsdqcp_dv  "SDQ Subscale: Conduct Problems"
label variable chsdqha_dv  "SDQ Subscale: Hyperactivity/Inattention"
label variable chsdqpp_dv  "SDQ Subscale: Peer Relationship Problems"
label variable chsdqps_dv  "SDQ Subscale: Prosocial"
label variable chsdqtd_dv  "SDQ Total Difficulties Score"

foreach var in chsdqes_dv chsdqcp_dv chsdqha_dv chsdqpp_dv chsdqps_dv chsdqtd_dv  {
        recode `var' (.=-9)
        lab define `var' -9 "missing", replace
        lab define `var' -8 "inapplicable", add
        lab val `var' `var'
        }
```

Appendix C: Cover letter and instructions for serology testing

C.1: Cover letter

Dear Participant,

Thank you for your continuing help with Understanding Society, and the COVID-19 study. Over the past year Understanding Society participants like you have made a tremendous contribution to COVID-19 research by completing the regular surveys which have increased the understanding of coronavirus symptoms and the impact the pandemic is having on people's lives.

In the last survey you agreed to receive an antibody testing kit. **In this pack you will find:**

- **A box containing the blood sample test kit**
- **An Instruction Sheet**

Please check that you are using the kit addressed to you and not to anyone else in the household.

Please try and complete your test within a few days of receiving the kit. It is important to have people taking this test as soon as possible, providing a snapshot of the antibody results of Understanding Society participants close to the time of the last survey.

Please read the instruction sheet thoroughly and watch the video carefully before beginning the home test. It provides all the information on how to complete the test and will help ensure that the test is successful. Participation in this part of the study is completely voluntary, and if you do not want to do this, then you do not have to. You can watch a video explaining how to take a sample here: https://www.youtube.com/watch?v=okTozcGMDlU

Once you have collected your sample please make sure that you have labelled it and completed the Return Form. Please follow the instructions which show how to label the test correctly and return it to the laboratory for analysis. You should get your results in 2-3 weeks.

You can also read more information on the Understanding Society Participants website: https://www.understandingsociety.ac.uk/participants/serology

Professor Michaela Benzeval – Director, *Understanding Society*

Understanding Society is being conducted in accordance with the Data Protection Act. This means your personal details will be kept strictly confidential and you and your household will not be

identifiable from the data. To view our privacy policy, see
https://www.understandingsociety.ac.uk/participants/coronastudy

Instructions

See our instructions video
www.gov.uk/taking-antibody-blood-sample

What's in your kit?

Collection tube

Plaster ×2

Alcohol wipe

Cleansing wipe

Lancet ×3

Mailer

Return form

Tube label (on return form)

Clear plastic bag

Cardboard box

❗ Grab a bowl of warm water and a clean tissue before you start

IRAS project ID: 296134

V2.4p - last reviewed October 2020

PREPARE AND SET UP

Before starting, make sure you…

✓ **Read these instructions in full**

✓ Watch how to do your test
Go to www.gov.uk/taking-antibody-blood-sample

✓ Drink lots of water
Being hydrated is the best thing you can do to prepare — drink 2 glasses of water at least 30 minutes before your test.

Do not do your test, or post your test, on a Saturday – it might sit in the post too long. Any other day is fine.

✓ Set aside enough time
Give yourself about 30 minutes to prepare, get set up, and collect your sample — especially if it's your first time.

✓ Do some light exercise
Do jumping jacks, or swing your arm and clench your fist for a minute (if it's safe for you to do so) – this helps get your blood moving.

1. Prepare

1. Check that the kit is **addressed to you** before starting.
2. Lay your kit out on a low table.
3. Grab a clean tissue and a bowl of warm water.
4. Wash your hands with soap for 20 seconds and dry with a clean towel.

2. Set up

1. Remove the cap from the tube.
2. Stand the tube upright using the slot in the **top of the cardboard box.**
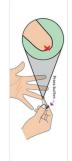3. Prepare the plaster so that it's ready to pop on your finger.

3. Get blood flowing

1. Put your **whole hand in a bowl of warm water** (about bath water temperature) for 2 minutes. Then dry your hand with a clean tissue.
2. Stand up for the whole process (but if you feel faint, sit down - and if you live with someone, ask for their help). Your hand must be lower than your elbow to keep the blood flowing.

Questions? Go to www.gov.uk/taking-antibody-blood-sample

v1

25/2/2021

COLLECT YOUR SAMPLE

1. Use the lancet

Press button

1. Clean your finger with the alcohol wipe — we recommend using your ring finger on your non-dominant hand.

2. To remove the protective tip at the bottom of the lancet: **twist the tip (one full turn)** and then **pull it all the way** out.

3. Place your hand face up on a hard surface (like a table top).

4. Press the lancet down on the outer centre of your fingertip.

5. **Press down firmly** on the purple **button until you hear a click.** You can only use each lancet once.

2. Fill the tube

1. Wipe away the first drop of blood with a clean tissue.

2. **Massage your hand firmly** in a downward motion every 3 to 4 seconds. Start from your palm and work down your finger. But **do not squeeze the piercing** or it might ruin your sample.

3. If the blood stops flowing, use pressure to wipe the pricked finger to stimulate blood flow. If this does not work, and it's been **more than 2 minutes, re-warm your hand** and prick another finger using a new lancet.

3. Mix the tube

Fill the tube to the **600 line.** If you only fill to the 400 line, that's OK. Place the cap onto the tube. Push it on **until it clicks.**

1. Gently **turn the tube upside down at least 10 times.** Make sure the blood touches the sides of the tube.

2. Do not worry if the blood solidifies when it's mixed. This is supposed to happen.

3. Apply pressure on finger to stop the bleeding, clean with a cleansing wipe, and apply a plaster.

4. Clean up any blood spills immediately and disinfect the surrounding area.

LABEL AND POST

1. **Peel your tube label** from the return form and **wrap** around your tube. Make sure the printed code is visible.

2. Get the clear plastic bag labelled 95 kPa Specimen transport. Put **the tube and lancets through the slot** — found in between the yellow sticker. With the upper part of the bag flat, **remove the sticker and fold the flap** onto the sticky backing. **Seal well by** pressing from the centre to the outside.

3. Fill in the **date and time on your return form.** Put the form and

clear plastic bag (containing tube) into the box your kit arrived in and close. Put the box into the prepaid mailer.

4. **Post your sample in a priority postbox** with the NHS logo. Find your nearest priority postbox at *www.royalmail.com/services-near-you.* If you're not able to safely get to a priority postbox, use your local postbox.

Symbols

☀ Keep away from sunlight	CE European Conformity	📖 Consult instructions for use	
☂ Keep away from rain	LOT Batch code	🏭 Manufacturer	
⊘ Do not re-use			
25°C Temperature limit 25°C			
IVD In vitro diagnostic medical device	⧗ Check tubes for use-by dates		

🛈 • This is an in vitro diagnostic medical device.
- Only use the equipment provided in your kit.
- This test must be carried out and labelled by the person who ordered the test.
- Watch out for signs of infection. If your wound site becomes red, hot, or swollen, seek medical advice.
- This kit is for people aged 18 and above.
- Apply firm pressure and raise your hand if bleeding does not stop. If this does not work, seek medical advice.
- If you feel faint, lie down and raise your legs. If you continue to feel unwell, seek medical advice.
- Risk of increased bleeding is higher if you take anticoagulation medication.
- Keep kit away from children, animals, sunlight, fire and smoking.
- Do not post your sample if temperature is above 25°C outside.
- It's highly unlikely, but there is a small risk of complications — including scarring, a haematoma (collection of blood), nerve damage, and necrosis (premature death of skin cells).

⚠ **Cannot get enough blood?**
For help go to **www.gov.uk/taking-antibody-blood-sample**. If you need to dispose of a partial sample, use the instructions to pack it up as you would with a normal sample (including the envelope). Then dispose of it all in household waste.