COVID-19 Survey in Five National Longitudinal Studies

Wave 1 User Guide (Version 1)

July 2020

Contact

Data queries: help@ukdataservice.ac.uk

Authors

Matt Brown, Alissa Goodman, Andrew Peters, George B. Ploubidis, Aida Sanchez, Richard Silverwood, Kate Smith.

How to cite this guide

Brown, M., Goodman, A., Peters, A., Ploubidis, G.B., Sanchez, A., Silverwood, R., Smith, K. (2020) *COVID-19 Survey in Five National Longitudinal Studies: Wave 1 User Guide (Version 1)*. London: UCL Centre for Longitudinal Studies and MRC Unit for Lifelong Health and Ageing.

This guide was first published in July 2020 by the UCL Centre for Longitudinal Studies (CLS) and the MRC Unit for Lifelong Health and Ageing (LHA).

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

The MRC Unit for Lifelong Health and Ageing at UCL (LHA) is home to three major studies: MRC National Survey of Health and Development, Southall And Brent REvisited Study and LINKAGE-Camden. For more information, visit https://www.ucl.ac.uk/cardiovascular/research/population-science-and-experimental-medicine/mrc-unit-lifelong-health-and-ageing-ucl

This document is available in alternative formats. Please contact the Centre for Longitudinal Studies. For questions and feedback about this user guide:

tel: +44 (0)20 7612 6875

email: clsfeedback@ucl.ac.uk

Contents

1. Introduction

1.1 Background

The Centre for Longitudinal Studies (CLS) and the MRC Unit for Lifelong Health and Ageing (LHA) carried out an online survey of the participants of five national longitudinal cohort studies in May 2020.

The aim of the survey was to collect insights into the lives of study participants including their physical and mental health and wellbeing, family and relationships, education, work, and finances during the lockdown. The questions focussed mainly on how participants' lives had changed from just before the outbreak of the pandemic in March 2020 up until their response to the survey at the height of the lockdown restrictions in May 2020.

The survey was sent to participants of all five of the national longitudinal cohort studies run at CLS and the LHA unit.   These studies have been following large nationally representative groups of people since birth, and their ages currently range from 19 through to 74. The studies included are:

- Millennium Cohort Study (born 2000-02) both cohort members and parents (MCS),
- Next Steps (born 1989-90) (NS),
- 1970 British Cohort Study (BCS70),
- 1958 National Child Development Study (NCDS), and
- MRC National Survey of Health and Development (NSHD, 1946 British birth cohort)

This survey (COVID-19 Wave 1) is the first COVID-19 data collection within these cohorts. A second survey (Wave 2) is planned for August 2020, and a third may be conducted in November 2020 (subject to funding).

This User Guide accompanies the deposit of the COVID-19 Wave 1 data at the UK Data Service.

2. Development

The development of the COVID-19 Wave 1 Survey took place during March and April 2020 and fieldwork was carried out in May 2020. A consultation was carried out in April 2020, during which time academic researchers, Government departments, third sector representatives and funders made proposals for the content for the survey. The scientific and technical development of the questionnaire was supported by members of the CLS and LHA teams, including Matt Brown, Darina Peycheva, Sierra Mesplie Cowan, Kate Smith, Bozena Wielgoszewska, David Bann, Jane Maddock, Morag Henderson, Andy Wong, Gaby Captur, Dan Davis and Praveetha Patalay. Final decisions on questionnaire content were taken by the PIs of the five studies and the Research Director of CLS (Professors Lisa Calderwood, Nish Chaturvedi, Emla Fitzsimons, Alissa Goodman, George B. Ploubidis and Alice Sullivan).

The questionnaire was programmed in Qualtrics by the CLS Survey Management Team. Once finalised the CLS version of the questionnaire was shared with LHA who made some minor amendments to the questionnaire (which are shown in the questionnaire documentation). Prior to launch the questionnaire was extensively tested within CLS and LHA.

3. Fieldwork

3.1 Issued sample and survey response

The COVID-19 Wave 1 Survey was conducted by web, and all those for whom an email address was held were be invited to take part. The survey was issued to web-only (rather than by mixed mode) due to restrictions on business operations at the time the survey was conducted.

All NCDS, BCS70, NSHD, Next Steps and MCS cohort members for whom an email address was held were selected for issue, provided that they a) had not permanently withdrawn from the study b) were not 'permanently untraced' and c) were not known to have died.

MCS parents were also invited to complete the survey, provided that they had taken part in the Age 17 Survey (MCS7) and an email address had been provided. Where two parents had participated, both were invited to take part. MCS cohort members and parents were all treated as individuals for the purpose of the survey – there were no links made between family members during the invitation process or within the questionnaire - however respondents can be linked for research purposes.

Emigrants for whom an email address was held were included in the issued sample. This includes study members living outside of Great Britain in the case of NCDS, BCS70 and Next Steps and those living outside the UK (i.e. including Northern Ireland) in the case of MCS.

Study specific emails, which included study branding and logos were sent from Qualtrics. Examples of these are available on request. The invite email was sent to NCDS cohort members on 4th May 2020 and the other cohorts on 5th May. Up to two reminder emails were sent to those who had not started, or those who had started but not completed the survey.   First reminders were sent to NCDS, BCS70 and Next Steps cohort members on 11th May and MCS cohort members and parents on the12th May. A second reminder was sent to all studies on the 15th May. Fieldwork for the first survey closed for all studies on the 26th May 2020. The invite email was sent to NSHD cohort members on 11th May 2020. First reminders were sent to NSHD cohort members on 20th May and fieldwork for the first survey closed on the 30th May 2020.   Charts showing how responses were spread across the fieldwork period are shown in Appendix 1.

The issued sample and response rates by CLS cohort are as follows:

Cohort	Issued sample (n)	Email bounced back	Failed email	Opted-Out*	Response**
NCDS	8943	696 (7.8%)	19 (0.2%)	159 (1.8%)	5178 (57.9%)
BCS70	10458	1077 (10.2%)	15 (0.1%)	115 (1.1%)	4223 (40.4%)
Next Steps	9380	473 (5.0%)	17 (0.2%)	107 (1.1%)	1907 (20.3%)
MCS (Cohort Members)	9946	724 (7.3%)	0 (0%)	95 (1.0%)	2645 (26.6%)
MCS (Parent)	9909	750 (7.6%)	0 (0%)	111 (1.1%)	2831 (28.6%)
NSHD	1843	80 (4.3%)	2 (0.1%)	29 (1.6%)	1258 (68.%)
TOTAL	50479	3800 (7.5%)	53 (0.1%)	616 (1.2%)	18042 (35.7%)

*Opt-outs were those who either phoned or emailed to request not to be contacted further, or who clicked the 'opt-out' button which was included in the invitation email.

**Response was defined as completion of the first block of the questionnaire ("Physical health since outbreak").

The 18,042 completed interviews include 576 completed by emigrants.

The total response rate pooled across cohorts with respect to the issued sample was 35.7%. Section 6 of this User Guides sets out further information about response, the achieved sample and derivation of weights.

4. Overview of questionnaire

4.1 Overview

The aim of the survey was to capture the health, social and economic consequences of the COVID-19 outbreak. It focussed on aspects of people's lives that needed to be captured immediately In Spring 2020, in order for the impact of the pandemic to be assessed and where possible, measures were chosen to maximise the use of the longitudinal measures already previously collected within the studies.

One survey was designed for all five cohorts, with the majority of questions being asked of all. However, a number of scales or questions were asked of specific cohorts only, primarily to enable longitudinal continuity with questions which had been included previously in major sweeps of each study. Some additional questions were added to the NSHD questionnaire.

At the end of the survey respondents were asked to sign up to the Zoe COVID-19 symptom tracker, and an opt-out was provided for participants who did not wish their symptom tracker data to be linked to the data held by the study.

It is estimated that the questionnaire took 25 minutes to complete on average.

A summary of the content is provided below. The full questionnaire, annotated with variable names, is available within this same data release and is also available on the CLS website.

Introduction	
Physical health	- Whether has had COVID-19 - Symptoms of COVID-19 - Self-rated general health (current & pre-COVID-19) - Long-standing health conditions - Whether routine appointments, surgery, cancer treatments were cancelled.

	• Medication (1946 birth cohort only) • Whether in defined vulnerable category • Extent of compliance with social distancing guidelines
Time use	Time use on typical weekday since outbreak: numeric hours (rounded to nearest hour): • Paid work • Volunteering / unpaid work (not for your household) • Home schooling your children • Other interactive activities with children • Caring for someone other than a child • Housework (e.g. cleaning, laundry, cooking, DIY) • Formal learning as part of a course • Physical activity / exercise • Leisure activities and hobbies (TV, gaming, reading, news, listening to music, gardening, online shopping, mealtime) • Socialising - talking, video-calling, messaging with non-household members • Travelling for work • Shopping or essential appointments • Personal care • Ill in bed • Other • Hours spent outside of home
Family and household	• Current household composition (household grid) • Children who do not live in household • Whether been changes to people who live with since outbreak

	• Whether started living with partner since outbreak
	• Changes to household since outbreak (e.g. whether chid moved in/has moved in with child)
	• Change in childcare & schooling arrangements (tailored questions, by age-band)
	• Whether in non-cohabiting relationship
	• Relationship satisfaction and conflict (current)
	• If study member or partner is pregnant: week of pregnancy
	• Number & age of children live with
	• Care/school attendance children under 4 pre/current outbreak. If attending school, reason (e.g. key worker)
	• If any children aged 5-16 physically attending school & reason.
	• Whether CM or anyone lives with usually received help, who from & hours pre outbreak
	• Change in help needed/received, hours & who from post outbreak
	• Number of rooms in house
	• Postcode
	• Access to garden
Financial Situation and benefits	• Subjective assessment of how managing financially pre and post outbreak
	• Food security, use of food banks
	• Receipt of benefits (self and/or partner) in 3 months before outbreak
	• New claims for benefits since outbreak
	• Use of mortgage/rent/debt holidays since outbreak

Employment & education (FE, HE, apprenticeships)	Pre-COVID19: (own & partner) • economic activity • apprenticeship type • hours • occupation (title, description) • contract type (fixed-term, zero-hours) (own and partner) • apprenticeship description In education or apprenticeship: • subject of study • institution name and town • course length • current year of study • how learning activity has changed: taking a break, online learning with/ without contact, drop-out • satisfaction with learning resources provided by institution, and whether has been able to continue studies effectively (0-10) • whether accepted a college/university place for September; name/town of college/university; whether still planning to do this, deferring, or no longer planning to do this. (MCS only) Since COVID-19: own/partner's: • economic activity • hours • work location • whether key worker status • apprenticeship description

Health behaviours	Behaviours pre-COVID 19 & current: • smoking (number of cigarettes) • vaping • alcohol (number and type of drinks and some aspects of problematic drinking) • physical activity (number of days did 30 mins or more) • diet (fruit & veg) • hours of sleep per night • weight
Mental health and social connectedness	• Contact with friends & family in past 7 days (telephone, video calls, email, text, electronic messaging) • Freq took part in online community activity in past 7 days • Freq gave help to anyone outside household affected by coronavirus in past 7 days • Social support • Loneliness • Overall life satisfaction • Mental health and wellbeing scales (capturing depression and anxiety). **NB Scales vary by cohort study; see scales section (4.1.2)** • Risk • Patience • Trust • Trust in government • Trust in political leaders • Self-assessed change in stress, interpersonal conflict & social trust

OPEN	Open question: the main ways the coronavirus outbreak has affected your life and/or your loved ones so far, and what you think the effects might be in the future.
Consent to link to symptom tracker app	Link to the Zoe COVID-19 symptom tracker, and an opt-out option for participants who did not wish their tracker data to be linked to the data held by the study.

4.1.2 Scales

The UCL COVID-19 questionnaire included several established scales which are listed below. Some scales were cohort specific.

4.1.2.1 Short Social Provisions Scale (3-items) (NS & MCS only)

Cutrona CE, Russell DW. The provisions of social support and adaptation to stress. Advance in Personal Relationships. 1987;1:37–67

Three items were included from the 10-item Social Provisions Scale (Cutrona 1987). The Social Provisions Scale measures the availability of social support.

Next Steps and MCS cohort members were asked to think about their current relationships with friends, family members, community members and so on. They were asked to indicate the extent to which each statement described their current relationship with other people from the following responses:

1. Very true
2. Partly true
3. Not true at all

Variable Name	Questions	Cohort
CW1_SOCPROV_1	I have family and friends who help me feel safe, secure and happy	NS & MCS

Variable Name	Questions	Cohort
CW1_SOCPROV_2	There is someone I trust whom I would turn to for advice if I were having problems	NS & MCS
CW1_SOCPROV_3	There is no one I feel close to	NS & MCS

4.1.2.2 UCLA Loneliness Scale (3-items) (All)

Three items from the 20-item UCLA loneliness scale were asked of all cohort members. They were asked to give the frequency in response to questions about current loneliness and related emotional sates from the following response options:

1. Hardly ever
2. Some of the time
3. Often

In addition, a fourth item (How often do you feel lonely?) was included which is not part of the UCLA scale, but has been used in NCDS62 survey.

Variable Name	Questions	Cohort
CW1_LONELY_1	How often do you feel that you lack companionship?	ALL
CW1_LONELY_2	How often do you feel left out?	ALL
CW1_LONELY_3	How often do you feel isolated from others?	ALL

4.1.2.3 Kessler 6 (MCS only)

Archives of General Psychiatry. 60(2), 184-189. Information on scoring and interpretation of this scale can be found at http://www.hcp.med.harvard.edu/ncs/k6_scales.php.

The Kessler 6 (K6) scale is a quantifier of non-specific psychological distress. It consists of six questions about depressive and anxiety symptoms that a person has experienced in the last 30 days.

MCS cohort members were asked six questions on how they had felt over the last 30 days with a self-report scale of five possible answers plus don't know/don't wish to answer (which was not shown on screen unless an item was left blank):

1. All of the time
2. Most of the time
3. Some of the time
4. A little of the time
5. None of the time

Variable name	Question	Cohort
CW1_PHDE	During the last 30 days, about how often did you feel so depressed that nothing could cheer you up?	MCS
CW1_PHHO	During the last 30 days, about how often did you feel hopeless?	MCS
CW1_PHRF	During the last 30 days, about how often did you feel restless or fidgety?	MCS
CW1_PHEE	During the last 30 days, about how often did you feel that everything was an effort?	MCS
CW1_PHWO	During the last 30 days, about how often did you feel worthless?	MCS

Variable name	Question	Cohort
CW1_PHNE	During the last 30 days, about how often did you feel nervous?	MCS

4.1.2.4 Warwick-Edinburgh Mental Wellbeing Scale (Short WEMWBS) (MCS only)

Copyright: Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) © NHS Health Scotland, The University of Warwick and University of Edinburgh, 2006, all right reserved.

The 7-item short WEMWBS is a mental wellbeing scale. It provides a single summary score indicating overall wellbeing. Permission was granted to use the scale.

The MCS cohort members were asked to select the answer that best described their experience over the past two weeks for seven statements:

1. None of the time
2. Rarely
3. Some of the time
4. Often
5. All of the time

Variable name	Question	Cohort
CW1_WEMWBS_1	I've been feeling optimistic about the future	MCS
CW1_WEMWBS_2	I've been feeling useful	MCS
CW1_WEMWBS_3	I've been feeling relaxed	MCS
CW1_WEMWBS_4	I've been dealing with problems well	MCS
CW1_WEMWBS_5	I've been thinking clearly	MCS
CW1_WEMWBS_6	I've been feeling close to other people	MCS
CW1_WEMWBS_7	I've been able to make up my own mind about things	MCS

Scoring:

https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/howto/

4.1.2.5 Malaise inventory (9-item) (NCDS & BCS70 only)

Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, health, and behaviour.* London: Longman.

The questions in the Malaise inventory measure levels of psychological distress, or depression.

NCDS and BCS70 cohort members were asked how they were feeling generally in response to the 9 questions with the response options:

1. Yes
2. No

Variable name	Question	Cohort
CW1_MALAISE_1	Do you feel tired most of the time?	NCDS & BCS70
CW1_MALAISE_2	Do you often feel miserable or depressed?	NCDS & BCS70
CW1_MALAISE_3	Do you often get worried about things?	NCDS & BCS70
CW1_MALAISE_4	Do you often get in a violent rage?	NCDS & BCS70
CW1_MALAISE_5	Do you often suddenly become scared for no good reason?	NCDS & BCS70
CW1_MALAISE_6	Are you easily upset or irritated?	NCDS & BCS70
CW1_MALAISE_7	Are you constantly keyed up and jittery?	NCDS & BCS70

Variable name	Question	Cohort
CW1_MALAISE_8	Does every little thing get on your nerves and wear you out?	NCDS & BCS70
CW1_MALAISE_9	Does your heart often race like mad?	NCDS & BCS70

4.1.2.6 GHQ-12 (Next Steps and 1946 cohort)

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

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

Variable name	Question	Cohort
CW1_GHQ121	Have you recently been able to concentrate on what you're doing?	NS & NSHD
CW1_GHQ122	Have you recently lost much sleep over worry?	NS & NSHD
CW1_GHQ123	Have you recently felt that you are playing a useful part in things?	NS & NSHD
CW1_GHQ124	Have you recently felt capable of making decisions about things?	NS & NSHD
CW1_GHQ125	Have you recently felt constantly under strain?	NS & NSHD
CW1_GHQ126	Have you recently felt you couldn't overcome your difficulties?	NS & NSHD

Variable name	Question	Cohort
CW1_GHQ127	Have you recently been able to enjoy your normal day to day activities?	NS & NSHD
CW1_GHQ128	Have you recently been able to face up to your problems?	NS & NSHD
CW1_GHQ129	Have you recently been feeling unhappy or depressed?	NS & NSHD
CW1_GHQ1210	Have you recently been losing confidence in yourself?	NS & NSHD
CW1_GHQ1211	Have you recently been thinking of yourself as a worthless person?	NS & NSHD
CW1_GHQ1212	Have you recently been feeling reasonably happy, all things considered?	NS & NSHD

The cohort member's score on the General Health Questionnaire 12 point scale (GHQ12) is derived by summing responses to the twelve GHQ12 questions (GHQ121 to GHQ1212). This is scored according to the 0-0-1-1 method, in which the first two possible responses to each question are assigned a value of 0 and the third and fourth responses with a value of 1, resulting in a maximum possible score of 12 for this variable. A higher score on this scale indicates a greater likelihood of mental ill health.

4.1.2.9 GAD-2 (Generalised Anxiety Disorder 2-item) (ALL)

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

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

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

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

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

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

0 = Not at all
1 = Several days
2 = More than half the days
3 = Nearly every day

Variable name	Question	Cohort
CW1_GAD2PHQ2_1	Feeling nervous, anxious or on edge	ALL
CW1_GAD2PHQ2_2	Not being able to stop or control worrying	ALL

4.1.2.10 PHQ-2 (Patient Health Questionnaire 2-item) (ALL)

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

The PHQ-2 enquires about the frequency of depressed mood and anhedonia over the past two weeks. The PHQ-2 includes the first two items of the PHQ-9

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

1. Not at all
2. Several days

3. More than half the days

4. Nearly every day

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

0 = Not at all

1 = Several days

2 = More than half the days

3 = Nearly every day

Variable name	Question	Cohort
CW1_GAD2PHQ2_3	Little interest or pleasure in doing things	ALL
CW1_GAD2PHQ2_4	Feeling down, depressed or hopeless	ALL

5. Survey Research Data

The research data from the survey have been supplied to the UK Data Service under End User Licence for the CLS studies (NCDS, BCS70, NS and MCS) and under Special Licence for the 1946 birth cohort study (NSHD).

All four CLS cohort studies are included in the same dataset.

Information about the variable names, number of cases, labelling of variables, identifiers and derived variables is given below.

Study	Data Owner	No. of cases*	UKDS Data Licencing
NCSD (1958)	CLS	5178	End user Licence
BCS70	CLS	4223	End user Licence
Next Steps	CLS	1907	End user Licence
MCS CMs	CLS	2645	End user Licence
MCS Parents	CLS	2831	End user Licence

Study	Data Owner	No. of cases*	UKDS Data Licencing
NSHD (1946)	LHA	1258	Special Licence

* The deposited data only includes the cases who completed the first block of the questionnaire ("Physical health since outbreak").

5.1 Licencing

All datasets are available from the UK Data Service (UKDS).

All users of the data need to be registered with the UKDS (details of how to do this are available at https://www.ukdataservice.ac.uk/get-data/how-to-access/registration).

The CLS data under the End User Licence can be downloaded once the access conditions have been ticked.

The LHA data under the Special Licence can be accessed by downloading a Special License application form. Once the form has been reviewed by UKDS and accepted by the LHA the data will be available to download.

5.2 Identifiers

Individual identifiers

All four CLS-based cohort studies are included in the same dataset, each with their standard research IDs that allow them to be linked to the other study data available at the UKDS.

The NSHD dataset has been pseudo-anonymised with an ID created exclusively for this project. If you wish to link other NSHD data to this web survey dataset, contact NSHD at: https://skylark.ucl.ac.uk/NSHD/doku.php.

For NCDS, BCS70 and Next Steps, the data for each cohort member is displayed with one case per row.

MCS data are displayed in long format, where MCSID identifies each family and PNUM identifies each family member. Therefore, for families with several

cohort members there will be several rows per family (MCDSID), but one row per family member (PNUM*). This is the same format as other MCS data deposits at UKDS.

Cohort identifier

Variable CW1_COHORT allows the identification of the data by cohort study, and for MCS whether it is the CM or parent respondent. It is set as follows:

1 = NCDS

2 = BCS70

3 = Next Steps

4 = MCS CM

5 = MCS Parent

6 = NSHD

Emigrant identifier

An emigrant flag (CW1_EMIGRANT) distinguishes between UK-based respondents and those living overseas. This information is acquired from the initial contact data used to select the participants at the point at which the survey began. It may therefore not be fully accurate as previous emigrants may have now returned and recent emigrants may not have notified the studies.

5.3 Variable names

In order to identify the data as belonging to COVID-19 wave 1 of data collection, the names of the variables are the original question names from the questionnaire, preceded by "CW1_". This is to allow the longitudinal matching of variables to subsequent data collection waves where only the wave number will change.

The variable names on the dataset have also been adjusted for question grids and for multi-code questions, where the question names are followed by the value number or order in the grid.

5.4 Variable description

Variable labels

The variable labels are based on the question wording from the survey questionnaire. Where necessary, labels have been modified in an effort to ensure they are comprehensible and accurate.

The labels from most variables include either a "Pre-C19" or "Post-C9" prefix to indicate whether the questions refers to the respondent's lives before or after the coronavirus outbreak. Labels that do not incorporate either prefix refer to broader timescales information and generally cover the time leading up to and following lockdown.

In addition, labels include the name of the scale used (e.g. "MALAISE:").

Value labels

The value labels are based on the answers from the questionnaire and have been individually reviewed and amended, where necessary.

5.5 Missing values

Missing values are consistently labelled as follows:

-1 = Not applicable

-8 = No information

Not applicable (-1) indicates that a question was left unanswered because the routing of the questionnaire did not reach that item.

No information (-8) indicates that the question was left blank where an answer is expected. This would cover a situation where the participant skipped the question (don't want to answer/don't know") or because of a technical issue.

21

5.6 Variable order

The order in which variables appear in the dataset is:

- IDs for each cohort
- Cohort study
- Sex
- Emigrant status
- Date of survey completion
- Answers to the questions in the order in which there were asked on the questionnaire. Newly coded variables that replace a disclosive question appear in the position of that original question (e.g. region appears in the position of postcode).
- Region of residence
- Weights

5.7 Coding of disclosive information

In addition to the pseudo-anonymisation, all text variables that contained detailed information provided by the respondents have been removed from the research dataset. This includes job titles, job descriptions, exact names of education institutions, town name, postcodes and the final open-ended question.

These variables have been replaced by less the disclosive coded variables, as follows:

Education

Two variables have been coded based on the open ended questions provided by the respondents:

- CW1_STUDYQUALDV: pre-COVID qualification level
- CW1_EDUQUALDV: post-COVID qualification level.

Employment

- SOC2010: Standard Occupational Classification, 3 digits (SOC Minor)

- NS-SEC: National Statistics Socio-economic classification, operational subcategory

NS-SEC was derived from SOC based on the simplified method described by ONS here:

https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010#deriving-the-ns-sec-full-reduced-and-simplified-methods

Geography

- Region of residence based on address details provided in survey

5.8 Data errors and inconsistencies

Users should be aware of the following data corrections.

Benefits

On the online questionnaire, the pre-COVID benefits grid (CW1_BENEFITB_1-14) included 'Pension Credit' twice (options 3 and 11). The data for these two values have been merged into the variable corresponding to option 3 (CW1_BENEFITB_3). Variable with option 11 (CW1_BENEFITB_11) has been removed from the final dataset.

It should be noted that some participants selected only one on these options, and some selected both.

Smoking

Participants who reported they currently smoked (CW1_SMOKING) were asked for the number of cigarettes smoked pre-COVID (CW1_NUMCIGSPP) and post-COVID (CW1_NUMCIGSSP).

The survey design did not allow participants to enter the value 0 for CW1_NUMCIGSPP (pre-COVID) so any potential respondents who only started smoking after the outbreak will not appear in the data.

Fruit and Vegetables

Many open-text numeric questions allowed for decimal input. While the majority of non-whole number responses were only 1 decimal place and have been left in the data, a number of unusual near 0 values occurred for CW1_FRTVEGPP and CW1_FRTVEGSP. These have been set to -8 (No information).

Self-reported weight

Participants could choose to provide their weight in stones (CW1_WGHTSTP_4) and pounds (CW1_WGHTSTP_5). There are 14 pounds in a stone.

In this survey no upper limit was set on how many pounds could be entered in the pounds field, and a number of respondents entered a value higher than 14. In some cases they left the weight in stones (CW1_WGHTSTP_4) empty, suggesting that the full weight was provided in pounds (CW1_WGHTSTP_5). However, the data is left untouched in order to leave any inference to data users.

5.9 Weights variables

The variables containing the calculated weights are as follows:

Variable name	Variable description
CW1_DESIGNWEIGHT	Weight: Design weight
CW1_SAMPPSU	Sampling: School (primary sampling unit)
CW1_SAMPSTRATUM	Sampling: Stratum
CW1_PTTYPE2	Stratum within Country
CW1_SPTN00	Fieldwork point number incorporating superwards
CW1_NH2	Population Correction Factor (for use in Stata)
Cw1_WEIGHT2	MCS Weight to use on whole UK analyses

Variable name	Variable description
CW1_COMBWT	Combined weight (design weight x web survey non-response weight) – final

Please refer to the Weights section below for a detailed explanation on how these were derived.

6. Derivation and implementation of non-response weights

6.1 Introduction

Non-response is common in longitudinal surveys. Missing values mean less efficient estimates because of the reduced size of the analysis sample, but also introduce the potential for bias since respondents are often systematically different from non-respondents. To support researchers in producing robust analysis, we have developed comprehensive advice on how to deal with missing data (1). The approaches we recommend to researchers capitalise on the rich data cohort members provided over the years before their non-response. These include well known methods such as Multiple Imputation (MI), Inverse Probability Weighting (IPW), and Full Information Maximum Likelihood (FIML). To correct for non-response in the COVID-19 Wave 1 Survey and facilitate analysis in all cohorts, non-response weights are provided, so that IPW analysis can be undertaken, either in isolation or in combination with MI.

This section of the User Guide describes the derivation and implementation of non-response weights for the COVID-19 Wave 1 Survey.

The weights were created and documented by Richard Silverwood and George B. Ploubidis, and the development of datasets for creating the weights was undertaken by Aase Villadsen, Martina Narayanan, Brian Dodgeon and Bozena Wielgoszewska.

6.2 Target population and response

For the purposes of weighting in NSHD, NCDS and BCS70, we have defined the target population of each cohort as individuals born in the specified birth period of the cohort who are alive and still residing in the UK. The COVID-19 Wave 1 Survey was also issued to a relatively small number of cohort members who had already emigrated from the UK, however we do not allocate weights to these individuals, and they are not used in the derivation of the non-response weights.

We note that for MCS and Next Steps, information on mortality and emigration was not available, and we therefore did not adjust the target populations to take deaths or emigrations into account. We expect mortality in both cohorts to be very low, and rates of emigration are also unlikely to be very significant. However to the extent that the target population in MCS and Next Steps may have been overestimated due to these factors, this would lead to a (likely, minor) underestimation of response relative to target in these cohorts.

The COVID-19 Wave 1 Survey target population and responses within the target population, as well as the web survey issued sample, in each cohort are presented in Table 1. Note that details of the COVID-19 Wave 1 Survey issued sample and total response are provided in section 3 of this User Guide. The differences in responses between Table 1 and section 3 reflect responses outside of the target population (i.e. cohort members who had already emigrated from the UK). In MCS there was an additional exclusion from the target population: only singletons and one twin or triplet from each twin pair/triplet set were included (i.e. second twin and second/third triplets were excluded).

Table 1. COVID-19 Wave 1 Survey target population and responses within the target population by cohort.

Cohort	Issued sample (n)	Response within the issued sample*	Cohort members within the target population (alive and still residing in the UK)	Response within the target population**
NCDS	8943	5178 (57.9%)	15291	5119 (33.5%)
BCS70	10458	4223 (40.4%)	17486	4132 (23.6%)
Next Steps	9380	1907 (20.3%)	15286	1876 (12.3%)
MCS (Cohort Members)	9946	2645 (26.6%)	19243	2609 (13.6%)
NSHD	1843	1258 (68.%)	3758	1170 (31.1%)

* Response was defined as completion of the first block of the questionnaire ("Physical health since outbreak")

** Mortality and emigration data not available for Next Steps and MCS.

The total response rate of all cohort members with respect to the target population was 21%, which is as expected lower than the response rate for cohort members with respect to the issued sample of 37.5% (note this differs to the total response rate given in Section 3.1, since no weights have been derived for MCS parents and thus their response is not included in the response rate given here). The response rate of cohort members within the issued sample is comparable to that of similar web surveys undertaken at this time (e.g. Understanding Society COVID19 Web Survey, 38.7%).

6.3 Derivation of non-response weights

The derivation of the COVID-19 Wave 1 Survey non-response weights was implemented in each cohort separately but following a common approach:

1. Within the sample corresponding to the target population (those alive and living in the UK), model COVID-19 Wave 1 Survey response conditional on a common set of covariates using logistic regression. The selection of covariates was informed from results of the CLS Missing Data Strategy (2, 3) and their *a priori* assumed association with the probability of response and/or with key COVID-19 Web Wave 1 Survey variables.

2. For COVID-19 Wave 1 Survey respondents, predict the probability of response from the model.

3. Calculate the COVID-19 Wave 1 Survey non-response weight as the inverse of the probability of response.

4. Examine the distribution of derived non-response weights across cohorts to decide whether truncation may be desirable; apply truncation if so.

5. Calibrate the COVID-19 Web Wave 1 Survey non-response weights so that they sum to the number of COVID-19 Wave 1 Survey respondents in each cohort.

The variables included in the COVID-19 Wave 1 Survey response model in stage 1 are listed in Table 2. We aimed to use broadly the same set of variables in each cohort to ensure consistency in the non-response weight derivation. However, it was not possible to include identical sets of variables due to data being collected at different ages and using different questions, and occasionally due to certain variables not been collected at all in some cohorts. Given that the non-response weight derivation was implemented separately in each cohort, such relatively minor differences were not deemed likely to be important.

Table 2. Variables included in the COVID-19 Wave 1 Survey response model in each cohort.

	NSHD	NCDS	BCS70	Next Steps	MCS
Sex	Birth	Birth	Birth	Age 14	9 months
Ethnicity	-	-	-	Age 14	9 months Age 3
Parental social class	Age 4	Birth	Birth	Age 14	9 months Age 11
Number of rooms at home/persons per room	Birth	Birth	Birth	-	9 months
Cognitive ability	Age 8	Age 7	Age 10	-	Age 11
Early life mental health	Age 13 & 15	Age 16	Age 16	Age 15	Age 11 Age 14
Voting	Age 26	Age 42	Age 42	Age 20	NA
Membership in organisations	Age 43	Age 42	Age 42	Age 26	Age 14
Internet access prior to web survey	Age 69	Age 50	Age 46	Age 26	Age 14
Consent for biomarkers	Age 60-64[B]	Age 44	Age 46	-	-
Consent for linkages	Age 60-64[B]	-	-	Age 26	-
Educational qualifications	Age 26	Age 42	Age 42	Age 26	9 months[A]

	NSHD	NCDS	BCS70	Next Steps	MCS
Economic activity	Age 60-64	Age 50	Age 46	Age 26	Age 14[A]
Partnership status	Age 69	Age 50	Age 46	Age 26	Age 14
Psychological distress	Age 69	Age 50	Age 46	Age 26	Age 14
BMI	Age 69	Age 50	Age 46	Age 26	Age 11
Self-rated health	Age 69	Age 50	Age 46	Age 26	Age 14
Smoking status	Age 69	Age 50	Age 46	Age 26	Age 14
Maternal mental health[C]	-	-	-	-	9 months
Social capital/social support	Age 69	Age 50	Age 46	Age 26	Age 14
Income	Age 69	Age 55	Age 42	Age 26	Age 14[A]
Number of non-responses across all previous sweeps	Birth – age 69	Birth – age 55	Birth – age 42	Age 14 – age 26	9 months – age 14

[A] Main respondent, >90% mothers. [B] Excluded from final model due to collinearity. [C] Also available in BCS70 at age 16 but not included in model.

Missing values in the above variables were handled using multiple imputation (MI), conducted in each cohort separately. The imputation model for each cohort included the above variables, COVID-19 Web Wave 1 Survey response and, for relevant cohorts (NSHD, Next Steps and MCS), the design weight. Five imputed datasets were created using chained equations. Such a small number of imputations was deemed sufficient as only point estimates (the probability of COVID-19 Web Wave 1 Survey response) were to be estimated from the MI analysis (more imputations would certainly be required

for inference). Models for COVID-19 Web Wave 1 Survey response were fitted in each imputed dataset and combined using standard rules (estimated models reported in the Appendix 2). From these models, the probability of COVID-19 Web Wave 1 Survey response was predicted for each respondent, with the non-response weight calculated as the inverse of the response probability. The distributions of the resultant COVID-19 Web Wave 1 Survey non-response weights are presented in Table 3.

Table 3. Distributions of the COVID-19 Wave 1 Survey non-response weight (prior to truncation and calibration).

Percentile	NSHD	NCDS	BCS70	Next Steps	MCS
0%	1.1	1.1	1.2	1.5	1.6
5%	1.2	1.2	1.4	2.1	2.1
25%	1.5	1.4	1.8	3.0	2.6
50%	1.9	1.7	2.3	4.3	3.7
75%	3.0	2.4	3.6	7.2	6.3
95%	9.6	6.4	10.5	27.5	18.0
100%	136.1	150.7	133.6	233.2	424.8

Test analyses were conducted in each cohort at different levels of weight truncation which suggested that truncation to 50 could provide some improvement in precision without undue introduction of bias. COVID-19 Web Wave 1 Survey non-response weights were therefore truncated to 50 in each cohort.

The COVID-19 Web Wave 1 Survey non-response weights were then calibrated so that they sum to the number of COVID-19 Web Wave 1 Survey respondents in each cohort by multiplying them by the ratio of the number of responses to the total of the uncalibrated non-response weights. The distributions of the resultant calibrated non-response weights are presented in Table 4.

Table 4. Distributions of the truncated and calibrated COVID-19 Wave 1 Survey non-response weights.

Percentile	NSHD	NCDS	BCS70	Next Steps	MCS
0%	0.34	0.44	0.32	0.20	0.27
5%	0.39	0.48	0.38	0.28	0.34
25%	0.47	0.55	0.47	0.39	0.43
50%	0.59	0.66	0.62	0.57	0.62
75%	0.94	0.92	0.94	0.96	1.04
95%	3.01	2.48	2.78	3.66	2.97
100%	15.75	19.52	13.22	6.65	8.23

6.4 Weights effectiveness

To examine the effectiveness of the derived non-response weights in restoring sample representativeness we conducted several analyses, one of which is presented here (and several more are also presented in Appendix 3). We considered the distribution of sex in each cohort, which is observed at baseline in virtually all cohort members. We compared the distribution of sex across all cohort members to the distribution of the same variable in COVID-19 Wave 1 Survey respondents only (to assess the extent of bias caused by non-response) and in COVID-19 Wave 1 Survey respondents after the application of the non-response weights (to assess to what extent the bias due to non-response could be overcome). The results are presented in Fig. 1. The extent of bias in the estimated percentage of female cohort members caused by non-response to the COVID-19 Wave 1 Survey varied across cohorts, but was substantial in most cases. However, the application of the non-response weights greatly reduced this bias in all cohorts, essentially completely eliminating it in NSHD, NCDS, BCS70 and MCS so that the sample representativeness with respect to this variable was restored. Whilst the truncated version of the non-response weights were not as effective in eliminating the bias in Next Steps, the untruncated version performed much

better, albeit with a wider confidence interval (results not shown). Although this analysis illustrates the performance of the non-response weights with respect to sex observed at baseline, it does not form a "test" of the performance of the non-response weights in general. In analyses of other variables (e.g. number of rooms, psychological distress, results available in Appendix 3) we found the non-response weights to perform similarly well (or better in the case of Next Steps), but this may not be the case for all variables of interest.
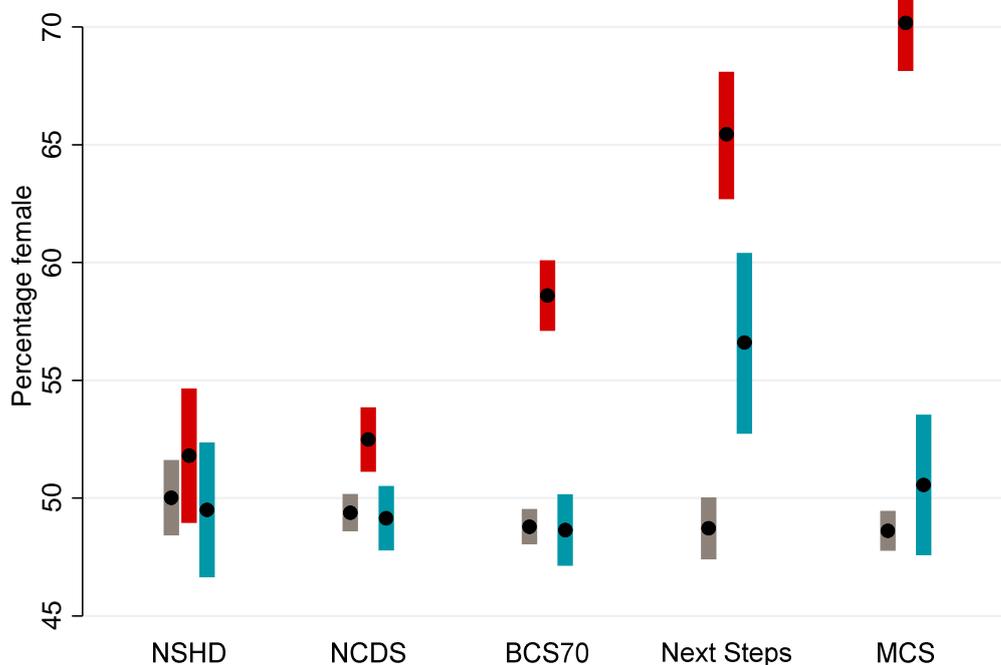
Fig. 1. Percentage female in each cohort under different estimation approaches. **Grey**: using observed baseline data from the whole cohort; red: using observed baseline data from COVID-19 Wave 1 Survey respondents only – unweighted (NCDS and BCS70) or using design weight only (NSHD, Next Steps and MCS); **blue:** using observed baseline data from COVID-19 Wave 1 Survey respondents only – weighted using non-response weights (in addition to design weights as appropriate).

6.5 Implementation of non-response weights

COVID-19 Web Wave 1 Survey non-response weights are provided as part of the COVID-19 Web Wave 1 Survey dataset. In cohorts where the study design means that design weights must be applied in any analyses (NSHD, Next Steps and MCS), the non-response weights have already been combined with the design weights ("CW1_INF", "CW1_DESIGNWEIGHT" and "CW1_WEIGHT2", respectively) to produce a combined weight ("CW1_COMBWT"). In cohorts without design weights (NCDS and BCS70), the same variable name ("CW1_COMBWT") has been used for consistency but is simply the COVID-19 Web Wave 1 Survey non-response weight.

We will illustrate how to use the COVID-19 Web Wave 1 Survey non-response weights by estimating the proportion of individuals reporting having coronavirus in each cohort, using the variable "CW1_COVID19". This variable is initially coded 1 "Yes, confirmed by a positive test", 2 "Yes, based on strong personal suspicion", 3 "Unsure" and 4 "No". We will combine the first two categories and combine the last two categories to produce a binary variable coded 0 "No" and 1 "Yes".

```
. recode CW1_COVID19 1/2=1 3/4=0

. label define CW1_COVID19_lab 0 "No" 1 "Yes"

. label values CW1_COVID19 CW1_COVID19_lab
```

The illustrative analyses are conducted in Stata (version 16), but could be conducted similarly in other statistical software packages. We will use the command `proportion` to estimate the proportions and specify the use of Agresti-Coull confidence intervals (4), as these are the generally preferred option in this setting.

NSHD

In NSHD there is a design weight ("CW1_INF") to take into account, but recall that this is already included in the COVID-19 Web Wave 1 Survey combined weight ("CW1_COMBWT").

```
. proportion CW1_COVID19 [pweight=CW1_COMBWT] if CW1_GROUP==6,
  citype(agresti)


Proportion estimation              Number of obs   =      1,170



----------------------------------------------------------------
                    |                          Agresti-Coull
                    | Proportion   Std. Err.    [95% Conf. Interval]
------------+---------------------------------------------------
CW1_COVID19  |

         No |   .9790111    .0060985       .9689493     .9859378

        Yes |   .0209889    .0060985       .0140622     .0310507

----------------------------------------------------------------
```

The estimated proportion of NSHD cohort members with coronavirus is 2.1%, with 95% confidence interval 1.4% - 3.1%.

NCDS

In NCDS there is no study design to take into account, so the analysis simply includes the COVID-19 Web Wave 1 Survey weight ("CW1_COMBWT").

```
. proportion CW1_COVID19 [pweight=CW1_COMBWT] if CW1_COHORT==1,
  citype(agresti)


Proportion estimation              Number of obs   =      5,118
```

```
----------------------------------------------------------------
            |                           Agresti-Coull
            | Proportion   Std. Err.    [95% Conf. Interval]
------------+---------------------------------------------------
CW1_COVID19 |
         No |   .9422861    .0051806      .9355507     .9483582
        Yes |   .0577139    .0051806      .0516418     .0644493
----------------------------------------------------------------
```

The estimated proportion of NCDS cohort members with coronavirus is 5.8%, with 95% confidence interval 5.2% - 6.4%.

BCS70

In BCS70 there is similarly no study design to take into account, so the analysis simply includes the COVID-19 Web Wave 1 Survey weight ("CW1_COMBWT").

```
. proportion CW1_COVID19 [pweight=CW1_COMBWT] if CW1_COHORT==2,
  citype(agresti)


Proportion estimation             Number of obs   =      4,131


----------------------------------------------------------------
            |                           Agresti-Coull
            | Proportion   Std. Err.    [95% Conf. Interval]
------------+---------------------------------------------------
CW1_COVID19 |
         No |   .8983936    .0089662      .8887996     .9072473
        Yes |   .1016064    .0089662      .0927527     .1112004
----------------------------------------------------------------
```

The estimated proportion of BCS70 cohort members with coronavirus is 10.2%, with 95% confidence interval 9.3% - 11.1%.

Next Steps

In Next Steps we must also account for the primary sampling unit ("CW1_SAMPPSU") and strata ("CW1_SAMPSTRATUM") of the study design. Recall that the Next Steps design weight ("CW1_DESIGNWEIGHT") is already included in the COVID-19 Web Wave 1 Survey combined weight ("CW1_COMBWT"). We first svyset the data, then conduct the analysis using the svy prefix.

```
. svyset CW1_SAMPPSU [pweight=CW1_COMBWT], strata(CW1_SAMPSTRATUM)


      pweight: CW1_COMBWT

          VCE: linearized

  Single unit: missing

    Strata 1: CW1_SAMPSTRATUM

        SU 1: CW1_SAMPPSU

      FPC 1: <zero>



. svy: proportion CW1_COVID19 if CW1_COHORT==3, citype(agresti)

(running proportion on estimation sample)



Survey: Proportion estimation



Number of strata =       37      Number of obs    =      1,876

Number of PSUs   =      589      Population size = 1,868.2311

                                 Design df        =       552


--------------------------------------------------------------
```

```
                |          Linearized       Agresti-Coull

                | Proportion    Std. Err.    [95% Conf. Interval]

------------+-------------------------------------------------

CW1_COVID19  |

         No |    .8915577     .012161      .8652231      .9132975

        Yes |    .1084423     .012161      .0867025      .1347769

----------------------------------------------------------------
```

The estimated proportion of Next Steps cohort members with coronavirus is 10.8%, with 95% confidence interval 8.7% - 13.5%.

MCS

In MCS we must again account for the primary sampling unit ("CW1_SPTN00") and strata ("CW1_PTTYPE2") of the study design, and additionally apply a finite population correction ("CW1_NH2"). Recall that the MCS design weight ("CW1_WEIGHT2") is already included in the COVID-19 Web Wave 1 Survey combined weight ("CW1_COMBWT"). We first svyset the data, then conduct the analysis using the svy prefix.

```
. svyset CW1_SPTN00 [pweight=CW1_COMBWT], strata(CW1_PTTYPE2)
  fpc(CW1_NH2)


      pweight: CW1_COMBWT

         VCE: linearized

  Single unit: missing

    Strata 1: CW1_PTTYPE2

       SU 1: CW1_SPTN00

      FPC 1: CW1_NH2


. svy: proportion CW1_COVID19 if CW1_COHORT==4, citype(agresti)

(running proportion on estimation sample)
```

```
Survey: Proportion estimation


Number of strata =        9       Number of obs   =       2,609

Number of PSUs   =      391       Population size = 2,687.3503

                                  Design df       =        382



----------------------------------------------------------------

              |              Linearized        Agresti-Coull

              | Proportion   Std. Err.     [95% Conf. Interval]

--------------+-------------------------------------------------

CW1_COVID19   |

          No  |  .9451542    .0061459       .931708     .9560996

         Yes  |  .0548458    .0061459       .0439004    .068292

----------------------------------------------------------------
```

The estimated proportion of MCS cohort members with coronavirus is 5.5%, with 95% confidence interval 4.4% - 6.8%.

6.6 References

1.   Silverwood R, Narayanan M, Dodgeon B, Ploubidis G. Handling missing data in the National Child Development Study: User Guide. London: UCL Centre for Longitudinal Studies; 2020.

2.   Mostafa T, Narayanan M, Pongiglione B, Dodgeon B, Goodman A, Silverwood RJ, et al. Improving the plausibility of the missing at random assumption in the 1958 British birth cohort: A pragmatic data driven approach. CLS Working Paper 2020/6. London: UCL Centre for Longitudinal Studies; 2020.

3.   Silverwood RJ, Calderwood L, Sakshaug JW, Ploubidis GB. A data driven approach to understanding and handling non-response in the Next Steps cohort. CLS Working Paper 2020/5. London: UCL Centre for Longitudinal Studies; 2020.

4.   Agresti A, Coull BA. Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions. The American Statistician. 1998;52(2):119-26.

7. Appendices

APPENDIX 1 – Cumulative Response by Cohort

Reminders indicated by red lines

Cumulative response rate (%) - BCS

Cumulative response rate (%) - Next Steps

Cumulative response rate (%) - MCS cohort members

Cumulative response rate (%) - MCS parents

Cumulative response rate (%) - NSHD
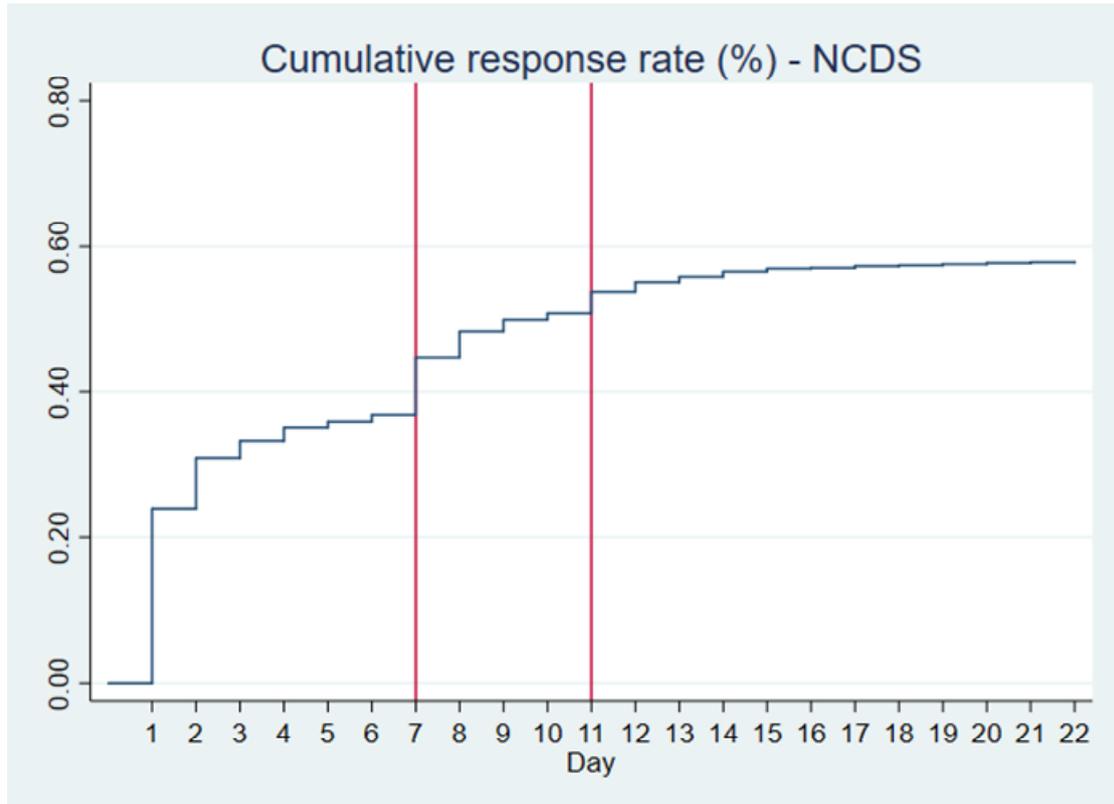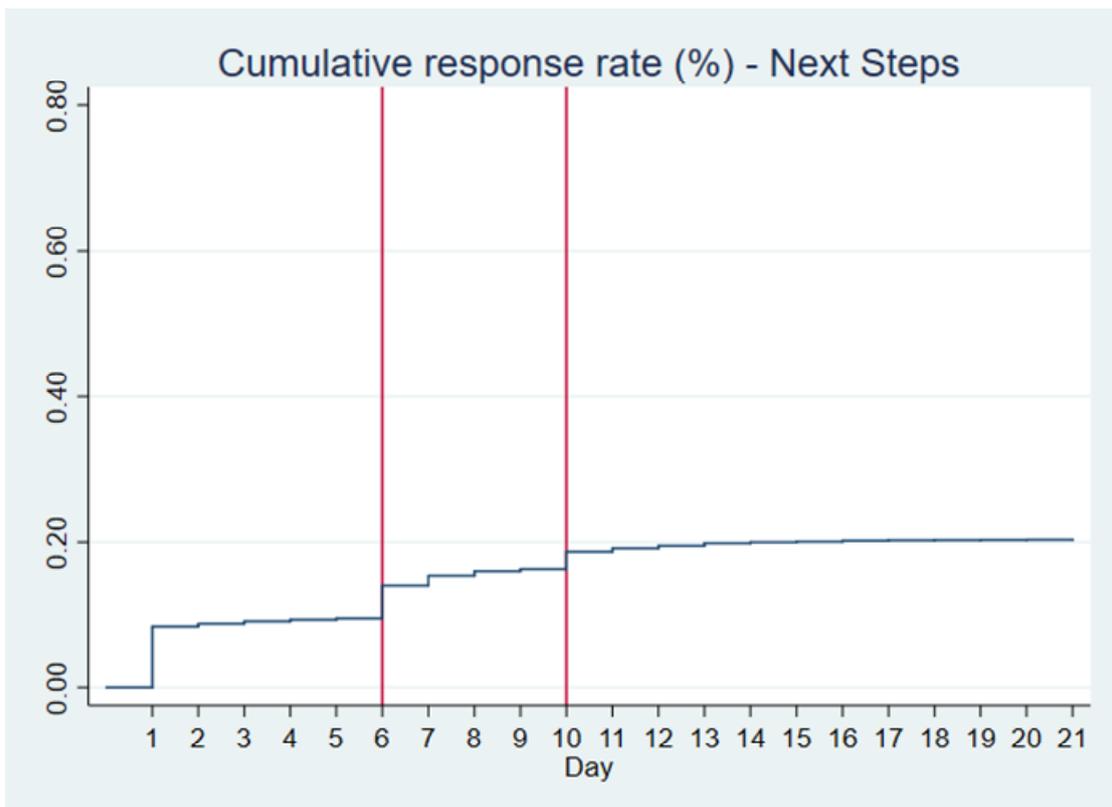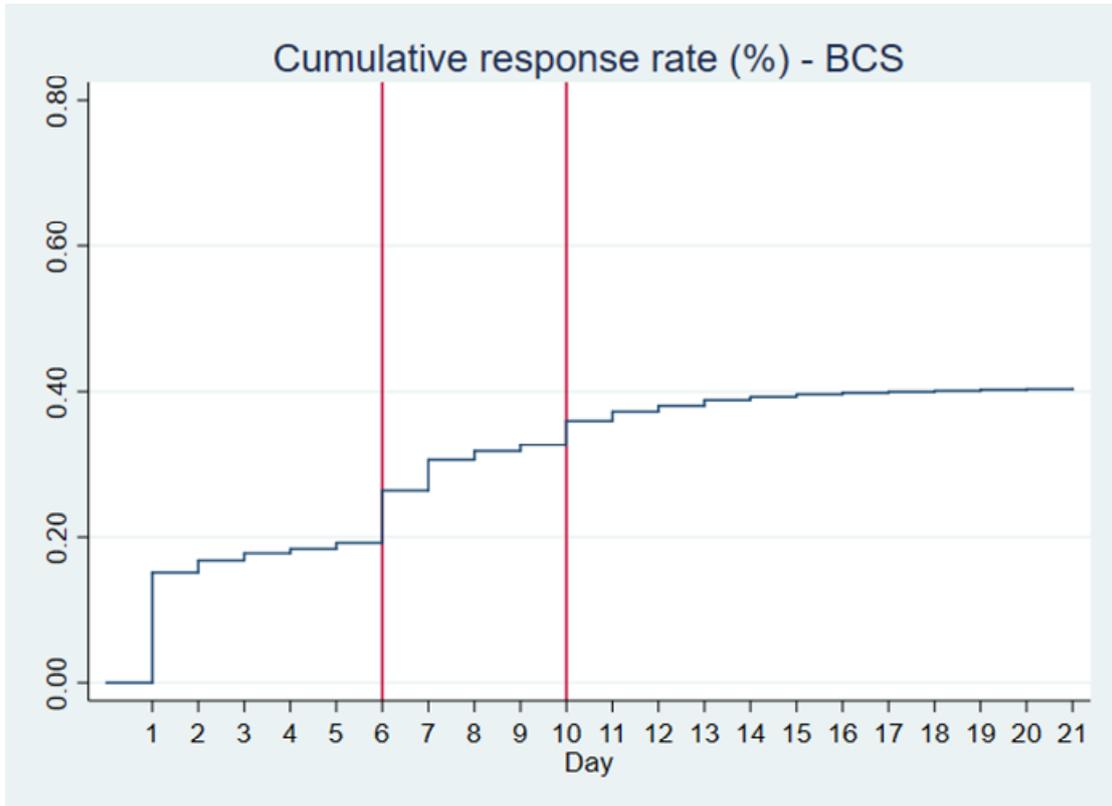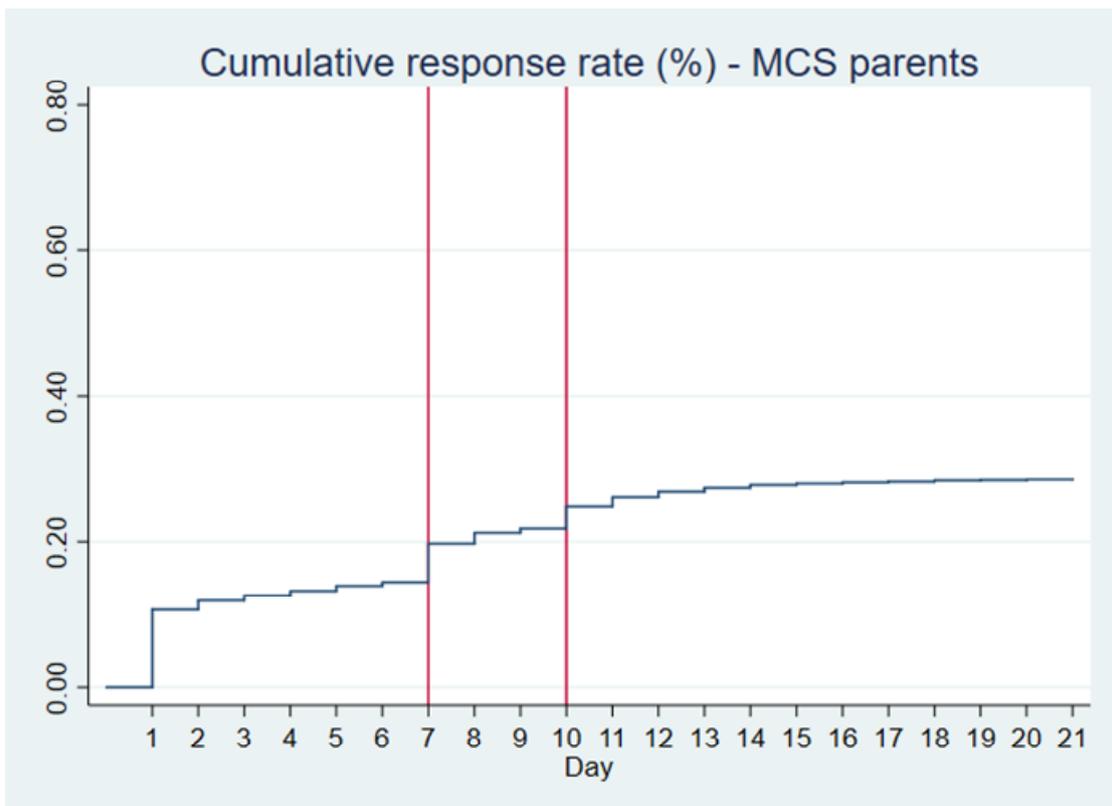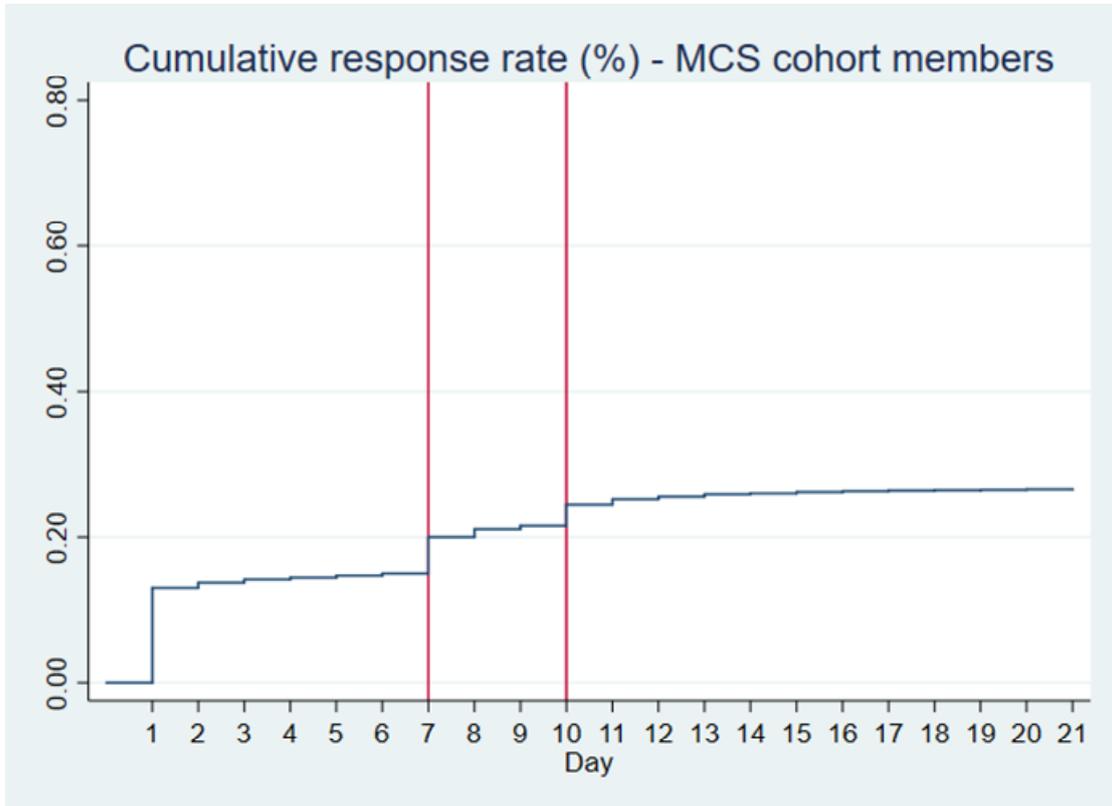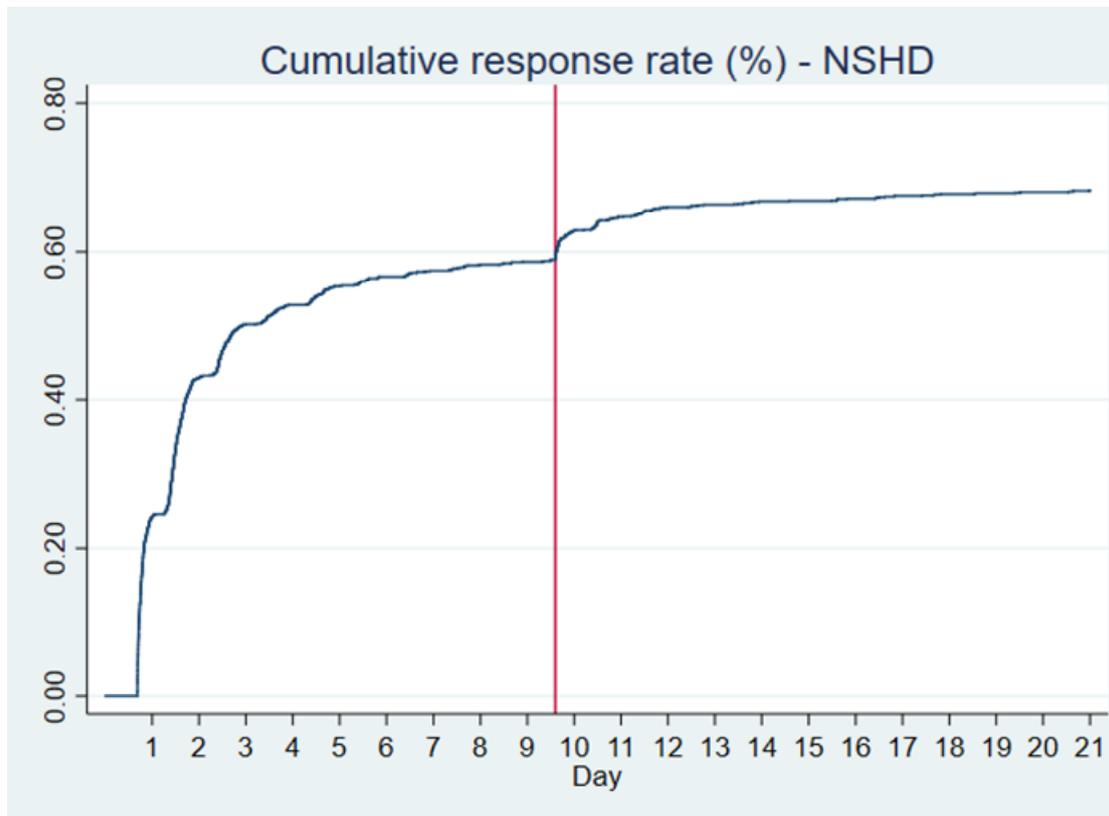
APPENDIX 2 - Non-response weights estimation

Table A1. Estimated COVID-19 Web Wave 1 Survey response model in NSHD (n = 3,758).

	OR	95% CI
Sex		
Male	1.00	
Female	1.25	1.03, 1.53
Voting		
Didn't vote	1.00	
Voted	1.02	0.79, 1.32
Internet access prior to web survey		
Never	1.00	
Not never	1.72	1.40, 2.11
Self-rated health		
Excellent/very good	1.00	

Good	0.61	0.50, 0.75
Fair/poor	0.38	0.28, 0.51
Income quintile		
1	1.00	
2	1.30	0.98, 1.74
3	1.61	1.21, 2.15
4	1.71	1.27, 2.31
5	1.90	1.41, 2.57
Parental social class		
Professional/intermediate	1.00	
Skilled	1.03	0.84, 1.26
Partly-/unskilled	0.88	0.67, 1.15
Early life mental health: Conduct problems		
Absent	1.00	
Mild	1.21	0.96, 1.53
Severe	1.07	0.69, 1.64
Early life mental health: Emotional problems		
Absent	1.00	
Mild	0.90	0.74, 1.09
Severe	0.82	0.61, 1.11
Membership in organisations		
None	1.00	
1	1.20	0.99, 1.45
2+	1.22	0.95, 1.56
Educational qualifications		
None attempted	1.00	
Up to GCE 'O' Level	2.14	1.61, 2.85
GCE 'A' Level	2.44	1.88, 3.16
First or higher degree	3.11	1.92, 5.05
Economic activity		

	OR	95% CI
Still in main occupation	1.00	
Retired but still earning	1.06	0.78, 1.43
Fully retired/unemployed/housewife	0.88	0.67, 1.14
Partnership status		
Single & never married	1.00	
Married	2.97	1.70, 5.19
Separated/divorced/widowed	2.41	1.29, 4.52
Smoking status		
Current Smoker	1.00	
Ex-smoker	2.33	1.63, 3.35
Never smoked	1.92	1.33, 2.78
Social capital/social support: Frequency of meeting family and friends		
Never/almost never	1.00	
Fairly frequently	1.18	0.93, 1.50
Very frequently	1.22	0.93, 1.59
Number of persons per room (per person)	0.90	0.78, 1.03
Cognitive ability	1.47	1.29, 1.68
Psychological distress	0.98	0.95, 1.02
Body mass index (kg/m^2)	1.01	0.98, 1.03
Number of non-responses across all previous sweeps	0.85	0.82, 0.87

Table A2. Estimated COVID-19 Web Wave 1 Survey response model in NCDS (n = 15,291).

	OR	95% CI
Sex		
Male	1.00	
Female	1.12	1.03, 1.22
Voting		
Didn't vote	1.00	

Voted	1.07	0.97, 1.19
Membership in organisations		
No	1.00	
Yes	1.25	1.13, 1.38
Membership in unions		
No	1.00	
Yes	1.10	0.99, 1.23
Internet access prior to web survey		
Yes	1.00	
No	0.35	0.30, 0.40
Consent for biomarkers		
Yes	1.00	
No	0.42	0.14, 1.21
Economic activity		
Currently employed	1.00	
Not currently employed	0.83	0.71, 0.97
Self-rated health		
Excellent/very good	1.00	
Good	0.88	0.80, 0.98
Fair/poor	0.78	0.66, 0.91
Income quintile		
1	1.00	
2	1.06	0.90, 1.24
3	1.19	1.01, 1.40
4	1.25	1.08, 1.46
5	1.36	1.11, 1.66
Parental social class		
Professional/managerial	1.00	
Intermediate	0.94	0.84, 1.04
Partly-/unskilled	0.87	0.76, 1.00

Educational qualifications		
None	1.00	
NQV Level 1-3	1.13	0.96, 1.33
NVQ Level 4-5	1.52	1.27, 1.83
Partnership status		
Single & never married	1.00	
Married/civil partner	1.29	1.11, 1.50
Separated/divorced/widowed	1.09	0.91, 1.30
Smoking status		
Never	1.00	
Former	1.01	0.91, 1.12
Current	0.78	0.69, 0.89
Social capital/social support: How often visit friends/have friends visit		
Never	1.00	
Fairly frequently	0.94	0.83, 1.06
Very frequently	0.84	0.73, 0.95
Social capital/social support: Have people around to listen to problems and feelings		
A little/not at all	1.00	
Somewhat	1.01	0.84, 1.21
A great deal	1.00	0.85, 1.18
Social capital/social support: Whether most people can be trusted		
Most people can be trusted	1.00	
Can't be too careful	0.88	0.80, 0.96
Other/depends	0.80	0.68, 0.95
Number of persons per room (per person)	0.91	0.86, 0.96
Cognitive ability	1.43	1.34, 1.52
Early life mental health (int)	0.92	0.85, 0.98
Early life mental health (ext)	1.19	0.97, 1.46

	OR	95% CI
Psychological distress	1.02	1.00, 1.05
Body mass index (kg/m²)	1.00	0.99, 1.01
Number of non-responses across all previous sweeps	0.62	0.61, 0.64

Table A3. Estimated COVID-19 Web Wave 1 Survey response model in BCS70 (n = 17,486).

	OR	95% CI
Sex		
Male	1.00	
Female	1.69	1.55, 1.85
Voting		
Didn't vote	1.00	
Voted	1.30	1.12, 1.50
Consent for biomarkers		
No to one/both	1.00	
Yes to both	1.17	1.00, 1.36
Economic activity		
Currently employed	1.00	
Not currently employed	0.83	0.71, 0.97
Self-rated health		
Excellent/very good	1.00	
Good	0.87	0.77, 0.99
Fair/poor	0.81	0.71, 0.93
Income quintile		
1	1.00	
2	1.16	0.99, 1.36
3	1.30	1.12, 1.50
4	1.45	1.21, 1.75
5	1.43	1.13, 1.80
Parental social class		

Professional/managerial	1.00	
Intermediate	0.95	0.84, 1.06
Partly-/unskilled	0.96	0.85, 1.10
Membership in organisations		
No organisations	1.00	
1 organisation	1.23	1.08, 1.40
2+ organisations	1.21	1.02, 1.44
Internet access prior to web survey		
None/little	1.00	
Medium	1.22	1.09, 1.38
Lots	1.30	1.14, 1.48
Educational qualifications		
None	1.00	
NQV Level 1-3	1.30	1.09, 1.55
NVQ Level 4-5	1.44	1.19, 1.74
Partnership status		
Never married/in CP	1.00	
Married/CP	1.07	0.95, 1.21
Separated/divorced/widowed	0.96	0.81, 1.13
Smoking status		
Never	1.00	
Former	0.93	0.84, 1.03
Current	0.78	0.66, 0.93
Social capital/social support: Frequency of meeting family and friends		
Never/rarely	1.00	
Fairly frequently	0.88	0.79, 0.99
Very frequently	0.77	0.68, 0.86
Social capital/social support: Have people around to listen to problems		
A little/not at all	1.00	

	OR	95% CI
Somewhat	1.09	0.87, 1.38
A great deal	1.04	0.81, 1.34
Number of rooms at home (per room)	1.01	0.98, 1.04
Cognitive ability	1.36	1.26, 1.46
Early life mental health	1.01	0.99, 1.02
Psychological distress	0.97	0.94, 0.99
Body mass index (kg/m^2)	1.01	1.01, 1.02
Number of non-responses across all previous sweeps	0.66	0.64, 0.67

Table A4. Estimated COVID-19 Web Wave 1 Survey response model in Next Steps (n = 15,286).

	OR	95% CI
Sex		
Male	1.00	
Female	2.12	1.90, 2.38
Voting		
Didn't vote	1.00	
Voted	0.76	0.67, 0.86
Membership in organisations		
Yes	1.00	
No	0.91	0.80, 1.03
Economic activity		
Currently employed	1.00	
Not currently employed	0.79	0.67, 0.94
Self-rated health		
Excellent/very good	1.00	
Good	0.88	0.77, 1.01
Fair/poor	0.84	0.69, 1.04
Income quintile		
1	1.00	

2	1.13	0.92, 1.38
3	1.20	1.00, 1.45
4	1.31	1.05, 1.64
5	1.68	1.34, 2.10
Parental social class		
Managerial	1.00	
Intermediate	0.90	0.78, 1.03
Routine/semi-routine	0.77	0.66, 0.90
Never worked	0.65	0.49, 0.86
Internet access prior to web survey		
None	1.00	
Little	1.07	0.86, 1.32
Lot	1.29	1.04, 1.59
Consent for linkages		
None	1.00	
Some	1.39	1.18, 1.63
All	1.66	1.44, 1.92
Educational qualifications		
None	1.00	
NQV Level 1-3	1.62	1.12, 2.34
NVQ Level 4-5	2.10	1.46, 3.04
Partnership status		
None	1.00	
Spouse/civil partner	1.01	0.83, 1.24
Cohabiting partner	1.05	0.92, 1.21
Smoking status		
Never	1.00	
Former	0.79	0.67, 0.93
Current	0.73	0.63, 0.86

	OR	95% CI
Social capital/social support: How often meet up with family and friends		
Very frequently	1.00	
Fairly frequently	1.42	1.26, 1.60
Rarely/never	1.44	1.17, 1.78
Social capital/social support: Have people around to listen to problems		
A little/not at all	1.00	
Somewhat	0.84	0.63, 1.11
A great deal	0.70	0.54, 0.90
Ethnicity		
White	1.00	
Indian/Pakistani/Bangladeshi	0.55	0.45, 0.68
Black Caribbean/Black African	0.36	0.27, 0.48
Mixed/Other	0.68	0.55, 0.85
Early life mental health	1.02	1.00, 1.05
Psychological distress	1.01	0.99, 1.04
Body mass index (kg/m^2)	1.01	1.00, 1.03
Social capital/social support: Trust scale	0.99	0.97, 1.02
Number of non-responses across all previous sweeps	0.67	0.64, 0.70

Table A5. Estimated COVID-19 Web Wave 1 Survey response model in MCS (n = 19,243).

	OR	95% CI
Sex		
Male	1.00	
Female	2.93	2.66, 3.24
Membership in organisations		
At least once a month	1.00	
Less than once a month	0.86	0.78, 0.95
Economic activity		

Currently employed	1.00	
Not currently employed	1.01	0.88, 1.15
Smoking status		
Never smoked	1.00	
Current/former/tried	0.61	0.53, 0.71
Social capital/social support: Family and friends who help me feel safe, secure and happy		
Very true	1.00	
Partly true/not true at all	1.14	0.99, 1.32
Social capital/social support: Someone I trust whom I would turn to if I had problems		
Very true	1.00	
Partly true/not true at all	1.02	0.89, 1.17
Social capital/social support: No one I feel close to		
Very/partly true	1.00	
Not true at all	1.23	1.02, 1.47
Self-rated health		
Excellent/very good	1.00	
Good	1.00	0.90, 1.11
Fair/poor	0.97	0.82, 1.13
Income quintile		
1	1.00	
2	1.29	1.09, 1.54
3	1.27	1.05, 1.55
4	1.41	1.17, 1.71
5	1.40	1.15, 1.69
Parental social class (9 months)		
Managerial	1.00	
Intermediate	0.88	0.76, 1.02
Routine/semi-routine	0.89	0.76, 1.03
Parental social class (age 11)		

|		
Managerial	1.00	
Intermediate	0.95	0.84, 1.07
Routine/semi-routine	0.84	0.70, 1.00
Internet access prior to web survey		
Little/none	1.00	
Medium	1.00	0.89, 1.14
Lots	1.01	0.89, 1.15
Educational qualifications		
None	1.00	
NQV Level 1-3	1.19	1.00, 1.42
NVQ Level 4-5	1.38	1.13, 1.69
Partnership status		
None	1.00	
Spouse/civil partner	1.18	1.01, 1.37
Separated/divorced/widowed	1.26	1.05, 1.51
Ethnicity		
White	1.00	
Indian/Pakistani/Bangladeshi/Other Asian/Chinese	1.17	0.99, 1.39
Black Caribbean/Black African/Other Black	1.05	0.79, 1.41
Mixed/Other ethnic group	0.89	0.68, 1.16
Number of rooms at home (per room)	0.97	0.94, 1.00
Cognitive ability	1.37	1.27, 1.46
Early life mental health	0.98	0.97, 0.99
Psychological distress	1.01	1.00, 1.02
Body mass index (kg/m^2)	1.00	0.99, 1.02
Maternal mental health	0.98	0.95, 1.00
Number of non-responses across all previous sweeps	0.42	0.39, 0.44
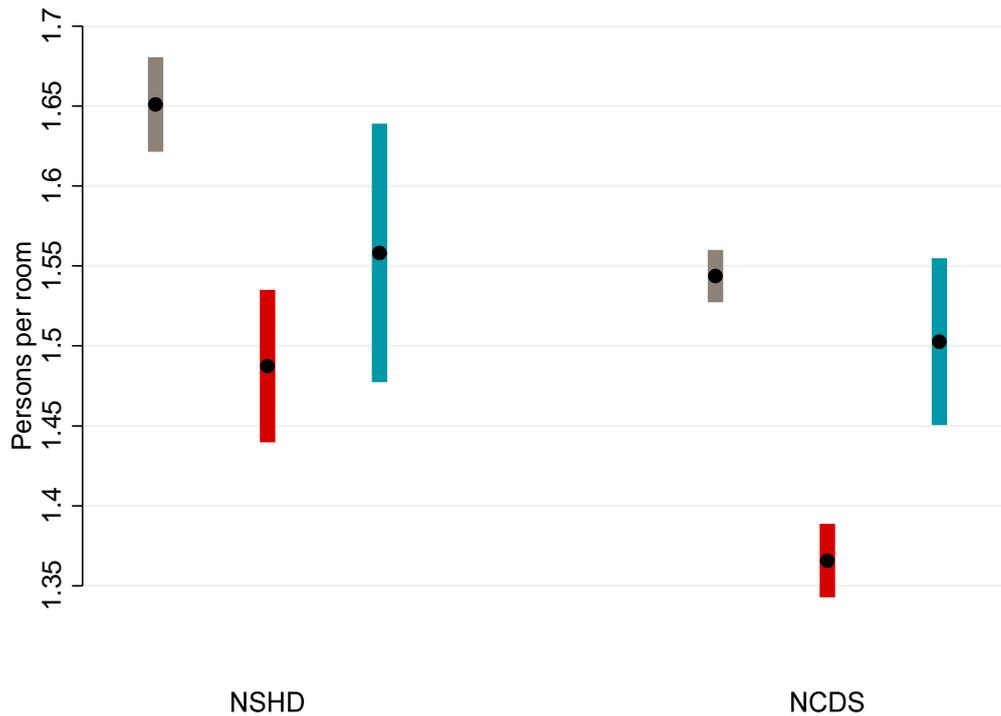
Fig. A5-1. Percentage of persons per room in NSHD and NCDS under different estimation approaches. **Grey:** using observed baseline data from the whole cohort; **red:** using observed baseline data from COVID-19 Wave 1 Survey respondents only – unweighted (NCDS) or using design weight only (NSHD); **blue**: using observed baseline data from COVID-19 Wave 1 Survey respondents only – weighted using non-response weights (in addition to design weights as appropriate).
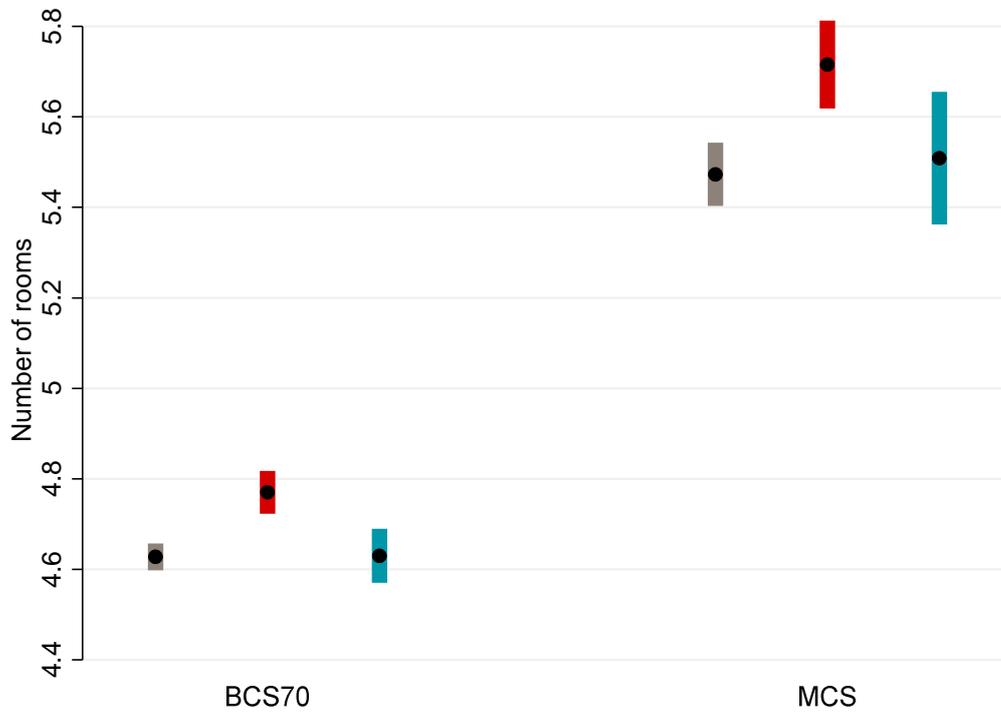
Fig. A5 -2. Percentage of number of rooms in BCS70 and MCS under different estimation approaches. **Grey**: using observed baseline data from the whole cohort; **red**: using observed baseline data from COVID-19 Wave 1 Survey respondents only – unweighted (BCS70) or using design weight only (MCS); **blue**: using observed baseline data from COVID-19 Wave 1 Survey respondents only – weighted using non-response weights (in addition to design weights as appropriate).
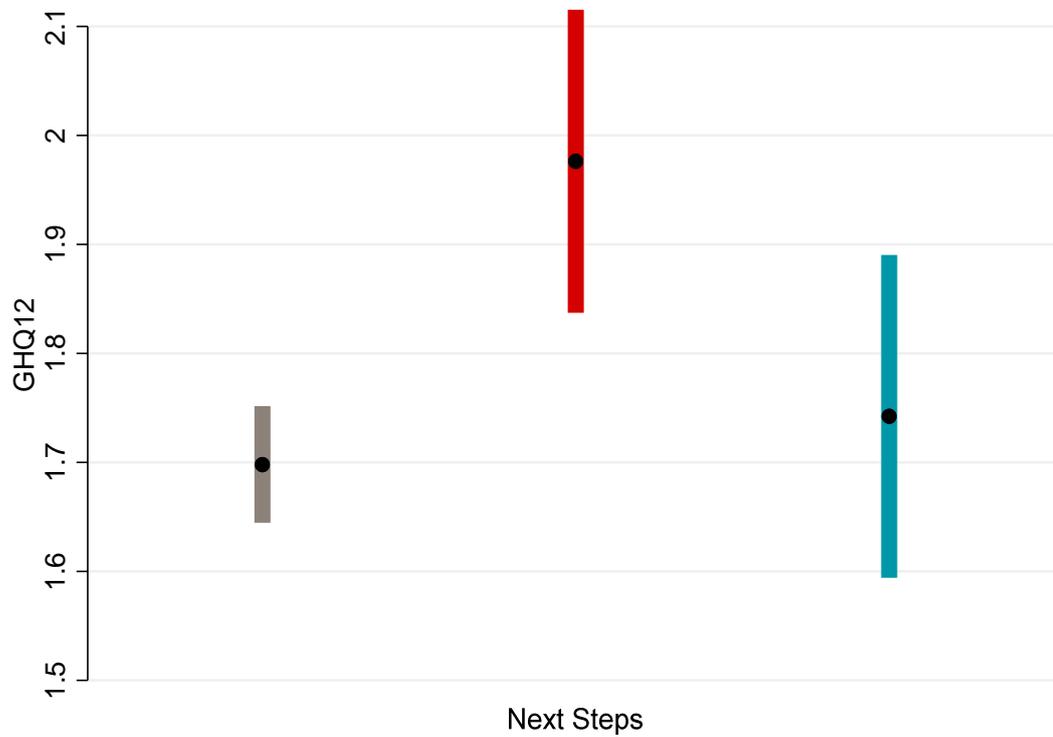
Fig. A5 - 3. GHQ12 psychological distress score in Nest Steps under different estimation approaches. **Grey**: using observed baseline data from the whole cohort; **red**: using observed baseline data from COVID-19 Wave 1 Survey respondents using design weight only; **blue**: using observed baseline data from COVID-19 Wave 1 Survey respondents only – weighted using non-response weights in addition to design weights.