

COVID-19 Survey in Five National Longitudinal Studies

Waves 1 and 2

User Guide (Version 2)

December 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: Waves 1 and 2 User Guide (Version 2)*. London: UCL Centre for Longitudinal Studies and MRC Unit for Lifelong Health and Ageing.

You should also acknowledge CLS following the guidance from

<https://cls.ucl.ac.uk/data-access-training/citing-our-data/>

This guide was published in December 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 Social Research Institute, University College London. It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ucl.ac.uk.

The 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
1.1 Background	1
2. Development	2
3. Fieldwork	2
3.1 Issued sample	2
3.2 Contact strategy	3
3.3 Response	5
4. Overview of questionnaire	7
4.1 Overview	7
5. Survey Research Data	21
5.1 Licencing	21
5.2 Identifiers	22
5.3 Variable names	24
5.4 Variable description	24
5.5 Missing values	25
5.6 Variable order	25
5.7 Coding of disclosive information	26
5.8 Data errors and inconsistencies	27
5.9 Weights variables	28
6. Derivation and implementation of non-response weights	29
6.1 Introduction	29
6.2 Target population and response	30
6.3 Derivation of non-response weights	34
6.4 Weights effectiveness	40
6.5 Implementation of non-response weights	42

6.6 References.....	48
7. Appendices.....	49
APPENDIX 1 - Non-response weights estimation.....	49
APPENDIX 2 – Restoring sample representativeness – further examples	64

1. Introduction

1.1 Background

The Centre for Longitudinal Studies (CLS) and the MRC Unit for Lifelong Health and Ageing (LHA) have carried out two online surveys of the participants of five national longitudinal cohort studies which have collected insights into the lives of study participants including their physical and mental health and wellbeing, family and relationships, education, work, and finances during the coronavirus pandemic. The Wave 1 Survey was carried out at the height of lockdown restrictions in May 2020 and focussed mainly on how participants' lives had changed from just before the outbreak of the pandemic in March 2020 until then. The Wave 2 survey was conducted in September/October 2020 and focussed on the period between the easing of restrictions in June through the summer into the autumn. A third wave of the survey will take place in early 2021.

The surveys were 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)

The Centre for Longitudinal Studies is funded by the [Economic and Social Research Council](#). The [Medical Research Council](#) funds the MRC Unit for Lifelong Health and Ageing.

2. Development

A consultation was carried out in advance of each wave, during which time academic researchers, Government departments, third sector representatives and funders made proposals for the content of the surveys. The scientific and technical development of the questionnaires 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).

3. Fieldwork

In Wave 1, interviews were completed between 4th and 30th May 2020. In Wave 2, interviews were completed between 10th September and 16th October 2020.

The Wave 1 Survey was programmed and administered by CLS/LHA using Qualtrics. The Wave 2 Survey was programmed and administered by Kantar Public.

3.1 Issued sample

At the time the Wave 1 survey was conducted, CLS/LHA were not able to send mass postal mailings, so the survey invitations had to be sent via email, meaning only those for whom an email address was held could be invited to participate.

The Wave 1 issued sample was therefore comprised of all NCDS, BCS70, NSHD, Next Steps and MCS cohort members for whom an email address was held, provided that they a) had not permanently withdrawn from the study b) were not 'permanently untraced' and c) were not known to have died.

At Wave 2 it was possible to send invitations via post, meaning that it was possible to include those for whom no email address was held. Cohort members with no email address were invited to take part provided that they had taken part in a recent major sweep of data collection or their address had been confirmed in recent years. As at Wave 1, cohort members who had permanently withdrawn from the study, were 'permanently untraced' or were known to have died were excluded from the survey. In addition, cohort members who had 'opted out' of the COVID project at Wave 1 were not invited to take part in Wave 2.

MCS parents were also invited to complete the surveys. At Wave 1, parents were invited to take part if they had taken part in the Age 17 Survey (MCS7) and provided an email address. At Wave 2, parents were invited to take part regardless of whether they had taken part in the Age 17 Survey and regardless of whether an email address was held (unless their families had permanently withdrawn, were permanently untraced, or the parent had opted out of the COVID project at Wave 1). Where cohort members had two parents, 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.

3.2 Contact strategy

In Wave 1, all communication with participants was conducted via email. Study-specific invitation emails, which included study branding and logos were sent by CLS/LHA from Qualtrics. Two email reminders were sent to NCDS, BCS70, Next Steps and MCS participants who had not started, or who had partially completed the survey. A single email reminder was sent to NSHD participants.

Wave 1 Contact Strategy

Date	Contact
4 th May 2020	Invitation email – NCDS
5 th May 2020	Invitation email - BCS70, Next Steps, MCS
11 th May 2020	Invitation email - NSHD Email reminder 1 – NCDS, BCS70, Next Steps
12 th May 2020	Email reminder 1 – MCS
15 th May 2020	Email reminder 2 – NCDS, BCS70, Next Steps, MCS
20 th May 2020	Email reminder 1 – NSHD
26 th May 2020	Web survey closed - NCDS, BCS70, Next Steps, MCS
30 th May 2020	Web survey closed – NSHD

In Wave 2, invitations were sent by both post and email (where email addresses were held). Invitations were followed by email reminders (where email addresses were held), two text message reminders (where mobile numbers were held) and a postal reminder (if no email address was held).

The Wave 2 issued sample was split into a soft-launch and main-stage. The soft-launch sample was comprised of 20% of issued cases (across all cohorts).

Wave 2 Contact Strategy

Date	Contact
10 th September 2020	Soft launch – Invitation email
16 th September 2020	Soft-launch – Email/Text reminder 1
19 th September 2020	Main-stage – Invitation email
24 th September 2020	Main stage – Email/Text reminder 1
25 th September 2020	Soft-launch – Email/Text reminder 2 / Postal reminder 1
1 st October 2020	Soft-launch - Email reminder 3
2 nd October 2020	Main-stage – Email/Text reminder 2
8 th October 2020	Main-stage Email reminder 3
16 th October 2020	Soft-launch / Main-stage – Web survey closed

3.3 Response

The issued sample and response rates for Wave 1 and Wave 2 are shown below by cohort. Response was defined as completion of the first module of questions covering experience of COVID-19:

Cohort	Wave1		Wave 2	
	Issued sample (n)	Response*	Issued sample (n)	Response*
NCDS	8943	5178 (57.9%)	11655	6282 (53.9%)
BCS70	10458	4223 (40.4%)	12133	5320 (43.9%)
Next Steps	9380	1907 (20.3%)	11529	3664 (31.8%)
MCS (Cohort Members)	9946	2645 (26.6%)	13547	3274 (24.2%)
MCS (Parent)	9909	2831 (28.6%)	22321	5707 (25.7%)
NSHD	1843	1258 (68.2%)	2551	1569 (61.5%)
TOTAL	50479	18042 (35.7%)	73736	25816 (35.0%)

Of the 25,816 interviews completed at Wave 2, 14,819 interviews were completed by Wave 1 respondents and 10,997 by 'new' respondents. A breakdown by cohort is provided below:

Wave 2			
Cohort	Wave 2 Interviews	Wave 1 respondents	'New' respondents
NCDS	6282	4538 (72%)	1744 (28%)
BCS70	5320	3389 (64%)	1931 (36%)
Next Steps	3664	1541 (42%)	2123 (58%)
MCS (Cohort Members)	3274	1770 (54%)	1504 (46%)
MCS (Parent)	5707	2238 (39%)	3469 (61%)
NSHD	1569	1343 (86%)	226 (14%)
TOTAL	25816	14819 (57%)	10997 (43%)

In Wave 1, interviews were achieved with 499 emigrants. In Wave 2, 695 emigrant interviews were completed.

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 surveys was to capture the health, social and economic consequences of the COVID-19 outbreak. The surveys sought to understand the immediate and ongoing impact of the pandemic. Where possible, measures were chosen to maximise the use of the longitudinal measures already previously collected within the studies.

For each wave 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.

The Wave 1 survey focussed on the period between the outbreak of the pandemic in March 2020 and the time of completion in May. In many domains, the Wave 1 survey sought to collect pre and post-pandemic measures of activity or behaviours in order to evaluate the level of change. At Wave 2 the routing of the questionnaire differed between Wave 1 respondents and non-respondents. Wave 1 non-respondents were asked about their pre-pandemic behaviours and activities and, in some domains, how these had changed between March and May, whereas this information had already been provided by Wave 1 respondents. All participants were then asked about their recent behaviours and activities.

The Wave 2 survey repeated much of the content from Wave 1 but also introduced some new question areas. Some of the new question areas were only asked of Wave 1 respondents (because otherwise the Wave 1 non-respondent version of the questionnaire would have been too long).

At the end of both surveys respondents were asked to sign up to the [Zoe COVID-19 symptom tracker app](#). It was explained to participants that we hoped to be able to link the data collected by the app with the data collected

in the survey. Participants were able to opt-out from this if they did not wish this to happen.

It is estimated that the Wave 1 and Wave 2 questionnaires took 25-30 minutes to complete on average.

A summary of the content is provided below. The full questionnaires, annotated with variable names, are available within this same data release and are also available on the [CLS website](#).

Section	Topic	W1	W2
Physical health	Whether has had COVID-19	Y	Y
	Whether has had COVID-19 test & results	N	Y
	Symptoms of COVID-19	Y	Y
	Self-rated general health	Y	Y
	Long-standing health conditions	Y	Y
	Whether routine appointments, surgery, cancer treatments were cancelled or postponed & whether rearranged & taken place.	Y	Y
	Medication	Y (1946 only)	Y
	Whether in defined vulnerable category	Y	Y
	Extent of compliance with social distancing guidelines	Y	N
Time use	Time use on typical weekday since outbreak	Y	Y
Family and household	Current household composition (household grid)	Y	Y
	Children who do not live in household	Y	Y
	Changes in household composition	Y	Y
	Change in childcare & schooling arrangements (tailored questions, by age-band)	Y	Y
	Whether in non-cohabiting relationship	Y	Y
	Relationship satisfaction and conflict	Y	Y
	Family conflict	N	Y
	Whether study member or partner is pregnant: week of pregnancy	Y	Y
	Number & age of children live with	Y	Y
	Care/school attendance children under 4 pre/current outbreak. If attending school, reason (e.g. key worker)	Y	Y
	If any children aged 5-16 physically attending school & reason.	Y	Y
Household care needs and receipt of care	Y	Y	

Section	Topic	W1	W2
Housing	Number of rooms in house	Y	Y
	Postcode	Y	Y
	Access to garden	Y	Y
	Tenure	Y	Y
Financial situation	Subjective assessment of how managing financially pre and post outbreak	Y	Y
	Food security, use of food banks	Y	N
	Receipt of benefits (self and/or partner) in 3 months before outbreak	Y	Y
	New claims for benefits since outbreak	Y	Y
	Use of mortgage/rent/debt holidays since outbreak	Y	Y
	Giving /receiving financial help	N	Y
Employment & education (FE, HE, apprenticeships)	Pre-COVID19 Economic activity – cohort member and partner	Y	Y
	<i>Employment – cohort member and partner</i>		
	Hours	Y	Y
	Occupation (title, description)	Y	Y
	Contract type (fixed-term, zero-hours)	Y	Y
	<i>Education – cohort member only</i>		
	Subject of study	Y	Y
	Institution name and town	Y	Y
	Course length	Y	Y
	Current year of study	Y	Y
	How learning activity has changed: taking a break, online learning with/ without contact, drop-out	Y	Y
	Satisfaction with learning resources provided by institution, and whether has been able to continue studies effectively (0-10)	Y	Y
	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)	Y	Y
	Whether has returned to college/university (MCS only)	N	Y
	Post-COVID19 Economic activity – cohort member and partner	Y	Y
	<i>Employment</i>		
	Hours	Y	Y
	Work location	Y	Y

Section	Topic	W1	W2
	Key worker status	Y	N
	Job satisfaction	N	Y
	Home working satisfaction	N	Y
Health behaviours	Smoking (number of cigarettes)	Y	Y
	Vaping	Y	Y
	Alcohol (number and type of drinks and some aspects of problematic drinking)	Y	Y
	Physical activity (number of days did 30 mins or more)	Y	Y
	Diet (fruit & veg)	Y	Y
	Hours of sleep per night	Y	Y
	Weight	Y	Y
Social contact, social support and loneliness	Contact with friends & family in past 7 days (telephone, video calls, email, text, electronic messaging)	Y	Y
	Frequency gave help in past 7 days to anyone outside household affected by coronavirus	Y	Y
	Participation in online community activity	Y	Y
	Provision of help to others	Y	Y
	Social support	Y	Y
	Loneliness	Y	Y
Mental health	Overall life satisfaction	Y	Y
	Self-assessed mental health	N	Y
	Control over life	N	Y
	Mental health and wellbeing scales (capturing depression and anxiety). NB Scales vary by cohort study; see scales section (4.1.2)	Y	Y
	Self-assessed change in stress, interpersonal conflict & social trust	Y	N
	Optimism	N	Y
Risk, Patience, Trust	Risk	Y	Y
	Patience	Y	Y
	Trust	Y	Y
	Trust in government & political leaders	Y	Y
Life events	Life Events	N	Y
Child Loop	<i>For children aged 4-18 living in household:</i>	N	Y

Section	Topic	W1	W2
	Summer term: Whether enrolled in school, school year and school type: whether attended school in person & amount attended; home learning: hours per day, online lesson provision; home schooling help (self/partner); learning resources; effect on academic progress	N	Y
	Autumn term: whether enrolled in school, details if new school, whether will attend in person & number of days,	N	Y
	Child's mental health pre covid & current	N	Y
OPEN	Open question on impact of COVID	Y	N
Consent to link to symptom tracker app	Request to download ZOE symptoms tracker app	Y	Y

4.1.2 Scales

The UCL COVID-19 questionnaire included several established scales which are listed below. Some scales were cohort specific. All scales detailed below were included in both Waves 1 and 2 with the exception of the Revised Life Orientation Test LOT-R (measuring optimism) which was added at Wave 2.

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
CW*_SOCPROV_1	I have family and friends who help me feel safe, secure and happy	NS & MCS
CW*_SOCPROV_2	There is someone I trust whom I would turn to for advice if I were having problems	NS & MCS
CW*_SOCPROV_3	There is no one I feel close to	NS & MCS

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

Daniel W. Russell (1996) UCLA Loneliness Scale (Version 3): Reliability, Validity, and Factor Structure, *Journal of Personality Assessment*, 66:1, 20-40, DOI: [10.1207/s15327752jpa6601_2](https://doi.org/10.1207/s15327752jpa6601_2)

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 states 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
CW*_LONELY_1	How often do you feel that you lack companionship?	ALL
CW*_LONELY_2	How often do you feel left out?	ALL
CW*_LONELY_3	How often do you feel isolated from others?	ALL

4.1.2.3 Kessler 6 (MCS only)

Kessler, R.C., Barker, P.R., Colpe, L.J., Epstein, J.F., Gfroerer, J.C., Hiripi, E., Howes, M.J, Normand, S-L.T., Manderscheid, R.W., Walters, E.E., Zaslavsky, A.M. (2003). Screening for serious mental illness in the general population. 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
CW*_PHDE	During the last 30 days, about how often did you feel so depressed that nothing could cheer you up?	MCS
CW*_PHHO	During the last 30 days, about how often did you feel hopeless?	MCS
CW*_PHRF	During the last 30 days, about how often did you feel restless or fidgety?	MCS

Variable name	Question	Cohort
CW*_PHEE	During the last 30 days, about how often did you feel that everything was an effort?	MCS
CW*_PHWO	During the last 30 days, about how often did you feel worthless?	MCS
CW*_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 & NSHD 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
CW*_WEMWBS_1	I've been feeling optimistic about the future	MCS & NSHD
CW*_WEMWBS_2	I've been feeling useful	MCS & NSHD
CW*_WEMWBS_3	I've been feeling relaxed	MCS & NSHD
CW*_WEMWBS_4	I've been dealing with problems well	MCS & NSHD
CW*_WEMWBS_5	I've been thinking clearly	MCS & NSHD
CW*_WEMWBS_6	I've been feeling close to other people	MCS & NSHD
CW*_WEMWBS_7	I've been able to make up my own mind about things	MCS & NSHD

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
CW*_MALAISE_1	Do you feel tired most of the time?	NCDS & BCS70
CW*_MALAISE_2	Do you often feel miserable or depressed?	NCDS & BCS70
CW*_MALAISE_3	Do you often get worried about things?	NCDS & BCS70
CW*_MALAISE_4	Do you often get in a violent rage?	NCDS & BCS70
CW*_MALAISE_5	Do you often suddenly become scared for no good reason?	NCDS & BCS70
CW*_MALAISE_6	Are you easily upset or irritated?	NCDS & BCS70
CW*_MALAISE_7	Are you constantly keyed up and jittery?	NCDS & BCS70
CW*_MALAISE_8	Does every little thing get on your nerves and wear you out?	NCDS & BCS70
CW*_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
CW*_GHQ121	Have you recently been able to concentrate on what you're doing?	NS & NSHD
CW*_GHQ122	Have you recently lost much sleep over worry?	NS & NSHD
CW*_GHQ123	Have you recently felt that you are playing a useful part in things?	NS & NSHD
CW*_GHQ124	Have you recently felt capable of making decisions about things?	NS & NSHD
CW*_GHQ125	Have you recently felt constantly under strain?	NS & NSHD
CW*_GHQ126	Have you recently felt you couldn't overcome your difficulties?	NS & NSHD
CW*_GHQ127	Have you recently been able to enjoy your normal day to day activities?	NS & NSHD
CW*_GHQ128	Have you recently been able to face up to your problems?	NS & NSHD
CW*_GHQ129	Have you recently been feeling unhappy or depressed?	NS & NSHD
CW*_GHQ1210	Have you recently been losing confidence in yourself?	NS & NSHD
CW*_GHQ1211	Have you recently been thinking of yourself as a worthless person?	NS & NSHD
CW*_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
CW*_GAD2PHQ2_1	Feeling nervous, anxious or on edge	ALL
CW*_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
CW*_GAD2PHQ2_3	Little interest or pleasure in doing things	ALL
CW*_GAD2PHQ2_4	Feeling down, depressed or hopeless	ALL

4.1.2.11 Revised Life Orientation Test LOT-R (3 optimism items) (ALL) (WAVE 2 only)

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67(6), 1063-1078.

The Revised Life Orientation Test (LOT-R) is a 10-item scale that measures how optimistic or pessimistic people feel about the future. The three items from the scale which are used to measure optimism were included.

Respondents were asked how much they agreed or disagreed with three statements with the response options:

1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree

Variable name	Question	Cohort
CW2_OPTMSM1	In uncertain times I usually expect the best	ALL
CW2_OPTMSM2	I'm always optimistic about my future	ALL
CW2_OPTMSM3	Overall, I expect more good things to happen to me than bad	ALL

5. Survey Research Data

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 research data from the surveys 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). However, MCS data for families containing triplets are only available under Secure Access because of the potential for disclosivity. This affects 5 cohort members and 2 parents in wave 1 and 8 cohort members and 5 parents in MCS in wave 2. Numbers in this guide represent the full data collected.

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

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 Licence application form. Once the form has been reviewed by UKDS and accepted by the LHA the data will be available to download.

Study	Data Owner	W1 cases*	W2 cases*	UKDS Data Licencing
NCSD (1958)	CLS	5178	6282	End User Licence
BCS70	CLS	4223	5320	End User Licence
Next Steps	CLS	1907	3664	End User Licence
MCS CMs**	CLS	2645 (2640)	3274 (3266)	End User Licence
MCS Parents**	CLS	2831 (2829)	5707 (5702)	End user Licence
NSHD (1946)	LHA	1258	1569	Special Licence

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

** The EUL data excludes triplet families

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 (CW*_CNUM00/CW*_PNUM00). This is the same format as other MCS data deposits at UKDS.

Cohort identifier

Variable CW*COHORT allows the identification of the data by cohort study, and for MCS whether it is the CM or parent respondent. This variable is referred to in the questionnaire documentation as 'GROUP'. It is set as follows:

1 = NCDS

2 = BCS70

3 = Next Steps

4 = MCS CM

5 = MCS Parent

6 = NSHD web

A second variable referred to in the questionnaire documentation but not included in the data deposit is COHORTID which does not distinguish between MCS cohort members and parents.

1 = NCDS

2 = BCS70

3 = Next Steps

4 = MCS

5 = NSHD web

Other identifiers

The wave 2 data includes a flag to identify respondents who previously completed the wave 1 survey (CW2_W1OUTCOME). A number of wave 2

questions were asked based on whether or not the cohort member completed wave 1.

An emigrant flag (CW*_EMIGRANT) distinguishes between UK-based respondents and those living overseas.

5.3 Variable names

In order to identify which wave of COVID-19 data collection variables belong to, the names of the variables are the original question names from the questionnaire, preceded by “CW1_” for wave 1 and “CW2_” for wave 2. This is to allow the longitudinal matching of variables to subsequent data collection waves where only the wave number is different.

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 many variables include either a “Pre-C19” or “Post-C19” 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.

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

Wave 1

In wave 1, missing values are consistently labelled as follows:

-1 = Not applicable

-8 = No information

Not applicable (-1) indicate that a question was left unanswered because the routing of the questionnaire did not reach that item. This is the same across both waves.

No information (-8) in wave 1 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.

Wave 2

Wave 2 distinguishes between unanswered questions, as respondents could specify "Don't know" or "Don't want to answer". Therefore, in wave 2 missing values are consistently labelled as follows:

-1 = Not applicable

-8 = Don't know

-9 = Don't want to answer

5.6 Variable order

The order in which variables appear in the dataset is:

- IDs for each cohort
- Cohort study
- Outcomes (wave 2 only)
- 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 (wave 1)

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

- SIC: Standard Industrial Classification for main respondent and partner (CW*(P)SIC3), 3 digits.
- SOC2010/2020: Standard Occupational Classification for main respondent and partner (CW*(P)SOC2010), 3 digits (SOC Minor). Low count (<10) potentially disclosive codes moved to corresponding major (2 digit) categories. SOC2020 is only available for wave 2
- NS-SEC: National Statistics Socio-economic classification. Three variables each for main respondent and partner covering; operational subcategories (CW*(P)NSSEC2010SB), operational categories

(CW*_(P)NSSEC2010OP) and analytic classes

(CW*_(P)NSSEC2010AN)

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

<https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatistics socioeconomicclassificationnssecbasedonsoc2010#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 and details

Benefits (wave 1)

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 (wave 1)

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 (wave 1/wave 2)

In wave 1 of data collection 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).

For the second wave, unlike wave 1, an upper limit (20) was set on daily portions of fruit and vegetables (CW2_FRTVEGPP, CW2_FRTVEGSP). A comparatively high number of respondents gave the highest possible answer of 20, suggesting many would have gone above if they could. Users are advised to consider setting their own cut-off because of this, particularly when comparing to the wave 1 equivalents.

Self-reported weight (wave 1)

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.

Child School Autumn term (wave 2)

Respondents provided a response for each child in their household about whether or not they started autumn term at school (CW2_SCAUTT_1-10). Two cases had responses out of range which were set to -8 (Don't know)

5.9 Weights variables

The variables containing the calculated weights are as follows:

Variable name	Variable description
CW*_DESIGNWEIGHT	Weight: Design weight
CW*_SAMPPSU	Sampling: School (primary sampling unit)
CW*_SAMPSTRATUM	Sampling: Stratum
CW*_PTTYPE2	Stratum within Country
CW*_SPTN00	Fieldwork point number incorporating superwards
CW*_NH2	Population Correction Factor (for use in Stata)
CW*_WEIGHT2	MCS Weight to use on whole UK analyses
CW*_COMBWT	Combined weight (design weight x web survey non-response weight) – final

Please refer to the Weights section 6 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 and 2 surveys 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 and 2 surveys. Non-response weight derivation was undertaken using a very similar approach at each wave.

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 and 2 surveys were 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. Non-response weights are not derived for the parents of MCS cohort members as parents are not the focus of the study.

The Wave 1 and 2 target populations and responses within the target populations, as well as within the issued samples, are presented in Table 1. Note that details of the issued samples 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).

The total response rate of all cohort members with respect to the target population was 20.8% in Wave 1 and 27.7% in Wave 2, which is as expected lower than the response rates for cohort members with respect to the issued sample of 37.5% and 39.1% respectively (note these differ from the total response rates 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 rates of cohort members within the issued samples are comparable to those of similar web surveys undertaken at similar times (e.g. Understanding Society COVID-19 Web Survey Waves 1 and 4, 38.7% and 38.0% respectively).

Table 1. COVID-19 Wave 1 and 2 surveys target population and responses within the target population by cohort.

Cohort	Wave 1				Wave 2			
	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**	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%)	15,291	5119 (33.5%)	11,655	6282 (53.9%)	15,291	6228 (40.7%)
BCS70	10,458	4223 (40.4%)	17,486	4132 (23.6%)	12,133	5320 (43.9%)	17,486	5236 (29.9%)
Next Steps	9380	1907 (20.3%)	15,770***	1876 (11.9%)	11,529	3664 (31.8%)	15,770***	3609 (22.9%)

	Wave 1				Wave 2			
MCS (Cohort Members)	9946	2645 (26.6%)	19,243	2609 (13.6%)	13,547	3274 (24.2%)	19,243	3233 (16.8%)
NSHD	1843	1258 (68.3%)	3758	1170 (31.1%)	2551	1569 (61.5%)	3758	1488 (39.6%)
Total	40,570	15,211 (37.5%)	71,548	14,906 (20.8%)	51,415	20,109 (39.1%)	71,548	19,794 (27.7%)

* 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.

*** Next Steps includes original sample only (i.e. not ethnic minority boost sample).

6.3 Derivation of non-response weights

The derivation of the COVID-19 Wave 1 and 2 survey non-response weights was implemented in each cohort separately but following a common approach. For each wave separately, we proceeded as follows:

1. Within the sample corresponding to the target population (those alive and living in the UK), model COVID-19 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 survey variables.
2. For COVID-19 survey respondents, predict the probability of response from the model.
3. Calculate the COVID-19 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 survey non-response weights so that they sum to the number of COVID-19 survey respondents in each cohort.

The variables included in the 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. The variables used in the Wave 1 and 2 response models were identical, with the exception of the addition of Wave 1 response to the Wave 2 response model.

Table 2. Variables included in the COVID-19 Wave 1 and 2 survey response models 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
Economic activity	Age 60-64	Age 50	Age 46	Age 26	Age 14 ^A

	NSHD	NCDS	BCS70	Next Steps	MCS
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
Response to COVID-19 Wave 1 survey^D	Age 74	Age 62	Age 50	Age 30	Age 19

^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. ^D Included in Wave 2 response model only.

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, response at the Wave under consideration 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 Wave 1 and 2 survey response) were

to be estimated from the MI analysis (more imputations would certainly be required for inference).

Models for COVID-19 survey response were fitted in each imputed dataset and combined using standard rules. Estimated models are reported in Appendix 1 but coefficients should be interpreted with caution since inclusion of independent variables from across the life course will likely lead to over-adjustment for earlier variables. In particular, the inclusion of Wave 1 survey response, which is itself affected by the other independent variables in the model, in the model for Wave 2 response will likely lead to over-adjustment for all other independent variables. We note that this since our focus is prediction of response for the derivation of non-response weights, over-adjustment is not an issue. However, if estimation of the association of specific variables with response was the aim, appropriate adjustment (e.g. not for variables in the causal pathway between the independent variable of interest and response) should be employed. From these models, the probability of COVID-19 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 Wave 1 and 2 non-response weights are presented in Table 3.

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 in both Waves 1 and 2. Both sets of non-response weights were therefore truncated to 50 in each cohort.

The non-response weights were then calibrated so that they sum to the number of COVID-19 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 3. Distributions of the COVID-19 Wave 1 and 2 Survey non-response weight (prior to truncation and calibration).

	Wave 1					Wave 2				
Percentile	NSHD	NCDS	BCS70	Next Steps	MCS	NSHD	NCDS	BCS70	Next Steps	MCS
0%	1.1	1.1	1.2	1.5	1.6	1.0	1.0	1.1	1.1	1.1
5%	1.2	1.2	1.4	2.1	2.1	1.0	1.0	1.1	1.1	1.2
25%	1.5	1.4	1.8	3.0	2.6	1.0	1.1	1.1	1.2	1.4
50%	1.9	1.7	2.3	4.3	3.7	1.1	1.1	1.3	2.6	2.1
75%	3.0	2.4	3.6	7.2	6.3	2.0	1.9	3.0	4.2	6.0
95%	9.6	6.4	10.5	27.5	18.0	6.9	5.8	8.2	12.8	14.6
100%	136.1	150.7	133.6	233.2	424.8	81.4	94.9	80.8	100.5	324.4

Table 4. Distributions of the truncated and calibrated COVID-19 Wave 1 and 2 survey non-response weights.

	Wave 1					Wave 2				
Percentile	NSHD	NCDS	BCS70	Next Steps	MCS	NSHD	NCDS	BCS70	Next Steps	MCS
0%	0.34	0.44	0.32	0.20	0.27	0.44	0.47	0.37	0.25	0.23
5%	0.39	0.48	0.38	0.28	0.34	0.45	0.48	0.38	0.26	0.26
25%	0.47	0.55	0.47	0.39	0.43	0.45	0.49	0.40	0.28	0.29
50%	0.59	0.66	0.62	0.57	0.62	0.47	0.52	0.46	0.61	0.44
75%	0.94	0.92	0.94	0.96	1.04	0.89	0.89	1.08	1.00	1.25
95%	3.01	2.48	2.78	3.66	2.97	3.04	2.63	2.91	3.06	3.03
100%	15.75	19.52	13.22	6.65	8.23	21.89	22.90	17.68	11.93	10.41

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 (with several more in Appendix 2). We considered the distribution of sex in each cohort, which is observed at baseline in virtually all cohort members. For Wave 1 and 2 separately, we compared the distribution of sex across all cohort members to the distribution of the same variable in COVID-19 survey respondents only (to assess the extent of bias caused by non-response) and in COVID-19 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 and 2 for Wave 1 and 2, respectively.

Wave 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 (Fig. 1). 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).

Wave 2: The extent of bias caused by non-response to the Wave 2 survey was generally a little less than in Wave 1, but remained substantial in most cases (Fig. 2). The application of the non-response weights greatly reduced this bias in all cohorts, essentially completely eliminating it so that the sample representativeness with respect to this variable was restored.

Although these analyses illustrate 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 (see Appendix 2) we found the non-response weights to perform similarly well, but this may not be the case for all variables of interest.

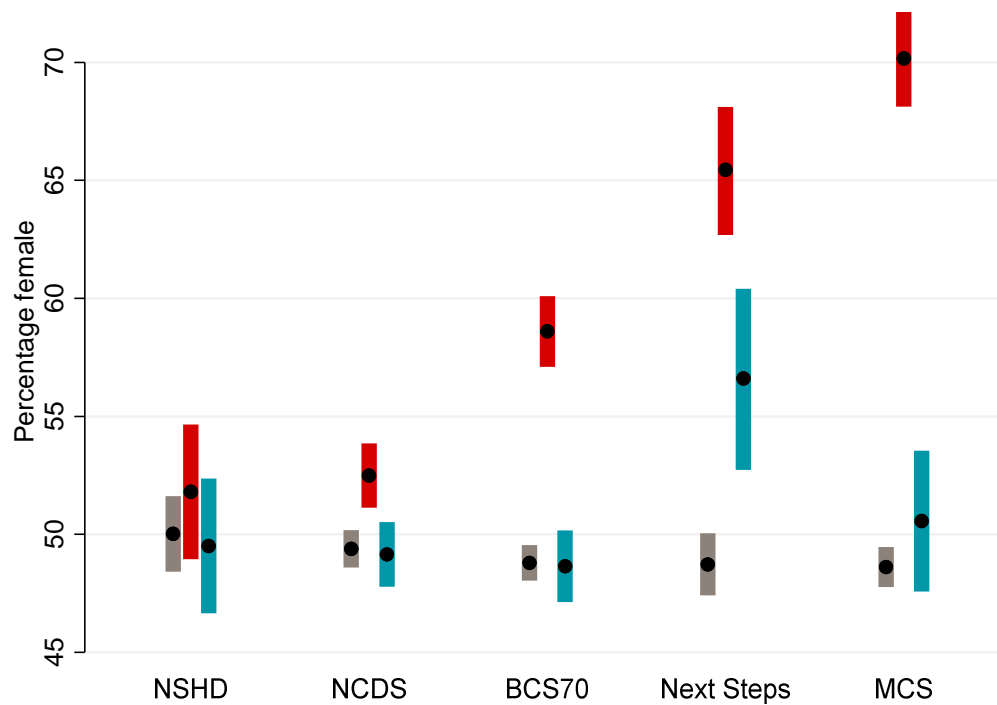


Fig. 1. Percentage female in each cohort under different estimation approaches. **Grey:** using observed baseline data from the whole cohort; **red:** using observed baseline data from COVID-19 Wave 1 survey respondents only – unweighted (NCDS and BCS70) or using design weight only (NSHD, Next Steps 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).

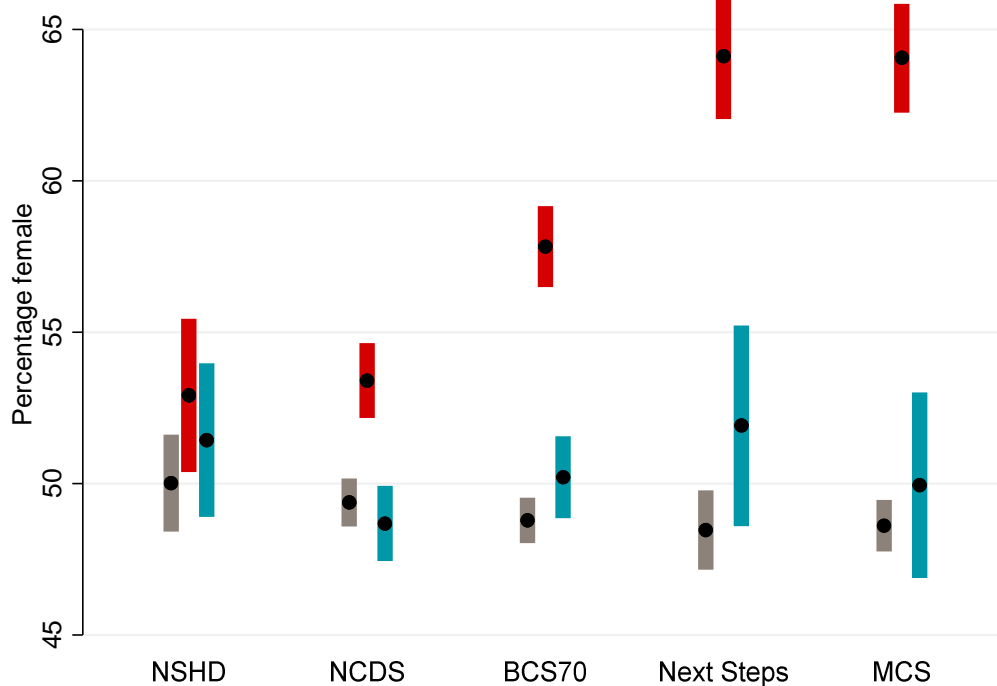


Fig. 2. Percentage female in each cohort under different estimation approaches. **Grey:** using observed baseline data from the whole cohort; **red:** using observed baseline data from COVID-19 Wave 2 survey respondents only – unweighted (NCDS and BCS70) or using design weight only (NSHD, Next Steps and MCS); **blue:** using observed baseline data from COVID-19 Wave 2 survey respondents only – weighted using non-response weights (in addition to design weights as appropriate).

6.5 Implementation of non-response weights

COVID-19 Wave 1 and 2 survey non-response weights are provided as part of the COVID-19 Wave 1 and 2 surveys 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 (“CW*_INF”, “CW*_DESIGNWEIGHT” and “CW*_WEIGHT2”, respectively, where “CW*” means “CW1” or “CW2” as appropriate) to produce a combined weight (“CW*_COMBWT”). In cohorts without design weights (NCDS and BCS70), the same variable name

("CW*_COMBWT") has been used for consistency but is simply the COVID-19 Wave 1 or 2 survey non-response weight.

We will illustrate how to use the COVID-19 Wave 2 survey non-response weights by estimating the proportion of individuals reporting having coronavirus in each cohort at Wave 2, using the variable "CW2_COVID19" (the equivalent analysis of coronavirus at Wave 1 could be conducted using the variable "CW1_COVID19", but is omitted here in the interests of space). 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 CW2_COVID19 -9/-1=. 1/2=1 3/4=0
. label define CW2_COVID19_lab 0 "No" 1 "Yes"
. label values CW2_COVID19 CW2_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 ("CW2_INF") to take into account, but recall that this is already included in the COVID-19 Wave 2 survey combined weight ("CW2_COMBWT").

```
. proportion CW2_COVID19 [pweight=CW2_COMBWT] if CW2_GROUP==6,
    citype(agresti)
```

```
Proportion estimation          Number of obs   =       1,485
```

	Agresti-Coull			
	Proportion	Std. Err.	[95% Conf. Interval]	
-----+-----				
CW2_COVID19				
No	.9778868	.0050314	.9689836	.9843239
Yes	.0221132	.0050314	.0156761	.0310164

The estimated proportion of NSHD cohort members with coronavirus is 2.2%, with 95% confidence interval 1.6% - 3.1%.

NCDS

In NCDS there is no study design to take into account, so the analysis simply includes the COVID-19 Wave 2 survey weight ("CW2_COMBWT").

```
. proportion CW2_COVID19 [pweight=CW2_COMBWT] if CW2_GROUP==1,
  citype(agresti)
```

Proportion estimation Number of obs = 6,185

	Agresti-Coull			
	Proportion	Std. Err.	[95% Conf. Interval]	
-----+-----				
CW2_COVID19				
No	.931368	.0063471	.92479	.9374106
Yes	.068632	.0063471	.0625894	.07521

The estimated proportion of NCDS cohort members with coronavirus is 6.9%, with 95% confidence interval 6.3% - 7.5%.

BCS70

In BCS70 there is similarly no study design to take into account, so the analysis simply includes the COVID-19 Wave 2 survey weight (“CW2_COMBWT”).

```
. proportion CW2_COVID19 [pweight=CW2_COMBWT] if CW2_GROUP==2,
    citype(agresti)
```

```
Proportion estimation          Number of obs   =          5,199
```

```
-----+-----
                |                               Agresti-Coull
                | Proportion   Std. Err.   [95% Conf. Interval]
-----+-----+-----
CW2_COVID19 |
    No      |   .8888752   .0075138   .8800385   .8971376
    Yes     |   .1111248   .0075138   .1028624   .1199615
-----+-----
```

The estimated proportion of BCS70 cohort members with coronavirus is 11.1%, with 95% confidence interval 10.3% - 12.0%.

Next Steps

In Next Steps we must also account for the primary sampling unit (“CW2_SAMPPSU”) and strata (“CW2_SAMPSTRATUM”) of the study design. Recall that the Next Steps design weight (“CW2_DESIGNWEIGHT”) is already included in the COVID-19 Wave 2 Survey combined weight (“CW2_COMBWT”). We first svyset the data, then conduct the analysis using the svy prefix.

```
. svyset CW2_SAMPPSU [pweight=CW2_COMBWT], strata(CW2_SAMPSTRATUM)
```

```

pweight: CW2_COMBWT
      VCE: linearized
Single unit: missing
Strata 1: CW2_SAMPSTRATUM
      SU 1: CW2_SAMPPSU
      FPC 1: <zero>

. svy: proportion CW2_COVID19 if CW2_GROUP==3, cotype(agresti)
(running proportion on estimation sample)

```

Survey: Proportion estimation

```

Number of strata =      37      Number of obs   =      3,545
Number of PSUs   =      647      Population size = 3,515.1557
                                   Design df        =      610

```

```

-----
          |                Linearized          Agresti-Coull
          | Proportion  Std. Err.  [95% Conf. Interval]
-----+-----
CW2_COVID19 |
      No |   .8844699   .0100528   .8631965   .9028214
      Yes |   .1155301   .0100528   .0971786   .1368035
-----

```

The estimated proportion of Next Steps cohort members with coronavirus is 11.6%, with 95% confidence interval 9.7% - 13.7%.

MCS

In MCS we must again account for the primary sampling unit (“CW2_SPTN00”) and strata (“CW2_PTTYE2”) of the study design, and additionally apply a finite population correction (“CW2_NH2”). Recall that the MCS design weight (“CW2_WEIGHT2”) is already included in the COVID-19 Wave 2 Survey combined weight (“CW2_COMBWT”). We first svyset the data, then conduct the analysis using the svy prefix.

```
. svyset CW2_SPTN00 [pweight=CW2_COMBWT], strata(CW2_PTTYE2)
    fpc(CW2_NH2)
```

```
    pweight: CW2_COMBWT
```

```
        VCE: linearized
```

```
Single unit: missing
```

```
Strata 1: CW2_PTTYE2
```

```
    SU 1: CW2_SPTN00
```

```
    FPC 1: CW2_NH2
```

```
. svy: proportion CW2_COVID19 if CW2_GROUP==4, cotype(agresti)
(running proportion on estimation sample)
```

```
Survey: Proportion estimation
```

```
Number of strata =          9      Number of obs   =       3,158
Number of PSUs   =       394      Population size = 3,197.8323
                                   Design df        =          385
```

```
-----
|                               Linearized      Agresti-Coull
```

	Proportion	Std. Err.	[95% Conf. Interval]	
-----+-----				
CW2_COVID19				
No	.8749703	.0118362	.8497628	.8964835
Yes	.1250297	.0118362	.1035165	.1502372

The estimated proportion of MCS cohort members with coronavirus is 12.5%, with 95% confidence interval 10.4% - 15.0%.

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 - Non-response weights estimation

Table A1. Estimated COVID-19 Wave 1 and 2 survey response models in NSHD (n = 3,758).

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Sex				
Male	1.00		1.00	
Female	1.25	1.03, 1.53	1.44	1.12, 1.84
Voting				
Didn't vote	1.00		1.00	
Voted	1.02	0.79, 1.32	1.45	0.75, 2.78
Internet access prior to web survey				
Never	1.00		1.00	
Not never	1.72	1.40, 2.11	1.91	1.38, 2.65
Self-rated health				
Excellent/very good	1.00		1.00	
Good	0.61	0.50, 0.75	0.90	0.71, 1.16
Fair/poor	0.38	0.28, 0.51	0.62	0.42, 0.90
Income quintile				
1	1.00		1.00	
2	1.30	0.98, 1.74	1.49	1.00, 2.24
3	1.61	1.21, 2.15	1.67	1.08, 2.58
4	1.71	1.27, 2.31	2.00	1.28, 3.13
5	1.90	1.41, 2.57	1.95	1.31, 2.90
Parental social class				
Professional/intermediate	1.00		1.00	
Skilled	1.03	0.84, 1.26	0.91	0.70, 1.19
Partly-/unskilled	0.88	0.67, 1.15	0.89	0.64, 1.23

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Early life mental health: Conduct problems				
Absent	1.00		1.00	
Mild	1.21	0.96, 1.53	0.80	0.56, 1.16
Severe	1.07	0.69, 1.64	0.80	0.47, 1.36
Early life mental health: Emotional problems				
Absent	1.00		1.00	
Mild	0.90	0.74, 1.09	1.18	0.94, 1.48
Severe	0.82	0.61, 1.11	0.93	0.66, 1.31
Membership in organisations				
None	1.00		1.00	
1	1.20	0.99, 1.45	1.21	0.91, 1.61
2+	1.22	0.95, 1.56	1.01	0.74, 1.38
Educational qualifications				
None attempted	1.00		1.00	
Up to GCE 'O' Level	2.14	1.61, 2.85	1.00	0.70, 1.42
GCE 'A' Level	2.44	1.88, 3.16	1.37	0.97, 1.94
First or higher degree	3.11	1.92, 5.05	1.02	0.59, 1.76
Economic activity				
Still in main occupation	1.00		1.00	
Retired but still earning	1.06	0.78, 1.43	1.13	0.74, 1.75
Fully retired/unemployed/housewife	0.88	0.67, 1.14	0.88	0.63, 1.24
Partnership status				
Single & never married	1.00		1.00	
Married	2.97	1.70, 5.19	2.57	1.34, 4.94
Separated/divorced/widowed	2.41	1.29, 4.52	1.90	0.97, 3.72
Smoking status				
Current Smoker	1.00		1.00	
Ex-smoker	2.33	1.63, 3.35	1.76	1.14, 2.74

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Never smoked	1.92	1.33, 2.78	1.48	0.89, 2.45
Social capital/social support: Frequency of meeting family and friends				
Never/almost never	1.00		1.00	
Fairly frequently	1.18	0.93, 1.50	0.86	0.63, 1.19
Very frequently	1.22	0.93, 1.59	0.86	0.61, 1.21
Number of persons per room (per person)	0.90	0.78, 1.03	0.76	0.64, 0.91
Cognitive ability	1.47	1.29, 1.68	1.15	0.96, 1.39
Psychological distress	0.98	0.95, 1.02	0.98	0.93, 1.02
Body mass index (kg/m ²)	1.01	0.98, 1.03	0.98	0.96, 1.00
Number of non-responses across all previous sweeps	0.85	0.82, 0.87	0.87	0.84, 0.89
Response to COVID-19 Wave 1 Survey				
Non-respondent	-	-	1.00	
Respondent	-	-	49.54	37.58, 65.30

Table A2. Estimated COVID-19 Wave 1 and 2 survey response models in NCDS (n = 15,291).

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Sex				
Male	1.00		1.00	
Female	1.12	1.03, 1.22	1.31	1.19, 1.46
Voting				
Didn't vote	1.00		1.00	
Voted	1.07	0.97, 1.19	1.12	1.00, 1.26
Membership in organisations				
No	1.00		1.00	
Yes	1.25	1.13, 1.38	1.10	0.98, 1.23
Membership in unions				
No	1.00		1.00	
Yes	1.10	0.99, 1.23	1.16	1.03, 1.31
Internet access prior to web survey				
Yes	1.00		1.00	
No	0.35	0.30, 0.40	0.61	0.53, 0.7
Consent for biomarkers				
Yes	1.00		1.00	
No	0.42	0.14, 1.21	0.70	0.28, 1.71
Economic activity				
Currently employed	1.00		1.00	
Not currently employed	0.83	0.71, 0.97	0.82	0.68, 0.99
Self-rated health				
Excellent/very good	1.00		1.00	
Good	0.88	0.80, 0.98	0.88	0.77, 1.01
Fair/poor	0.78	0.66, 0.91	0.85	0.73, 0.99
Income quintile				
1	1.00		1.00	

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
2	1.06	0.90, 1.24	0.96	0.79, 1.17
3	1.19	1.01, 1.40	1.06	0.77, 1.44
4	1.25	1.08, 1.46	1.22	0.99, 1.50
5	1.36	1.11, 1.66	1.39	1.06, 1.82
Parental social class				
Professional/managerial	1.00		1.00	
Intermediate	0.94	0.84, 1.04	0.91	0.81, 1.04
Partly-/unskilled	0.87	0.76, 1.00	0.94	0.80, 1.10
Educational qualifications				
None	1.00		1.00	
NQV Level 1-3	1.13	0.96, 1.33	1.20	1.01, 1.43
NVQ Level 4-5	1.52	1.27, 1.83	1.68	1.38, 2.06
Partnership status				
Single & never married	1.00		1.00	
Married/civil partner	1.29	1.11, 1.50	1.19	0.99, 1.42
Separated/divorced/widowed	1.09	0.91, 1.30	1.11	0.92, 1.34
Smoking status				
Never	1.00		1.00	
Former	1.01	0.91, 1.12	0.93	0.83, 1.05
Current	0.78	0.69, 0.89	0.73	0.64, 0.85
Social capital/social support: How often visit friends/have friends visit				
Never	1.00		1.00	
Fairly frequently	0.94	0.83, 1.06	0.93	0.78, 1.10
Very frequently	0.84	0.73, 0.95	0.93	0.78, 1.10
Social capital/social support: Have people around to listen to problems and feelings				
A little/not at all	1.00		1.00	
Somewhat	1.01	0.84, 1.21	0.99	0.80, 1.24

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
A great deal	1.00	0.85, 1.18	1.04	0.86, 1.27
Social capital/social support: Whether most people can be trusted				
Most people can be trusted	1.00		1.00	
Can't be too careful	0.88	0.80, 0.96	0.96	0.85, 1.09
Other/depends	0.80	0.68, 0.95	1.03	0.85, 1.25
Number of persons per room (per person)	0.91	0.86, 0.96	0.95	0.89, 1.00
Cognitive ability	1.43	1.34, 1.52	1.09	1.01, 1.18
Early life mental health (int)	0.92	0.85, 0.98	0.96	0.89, 1.03
Early life mental health (ext)	1.19	0.97, 1.46	1.03	0.81, 1.30
Psychological distress	1.02	1.00, 1.05	0.98	0.95, 1.01
Body mass index (kg/m ²)	1.00	0.99, 1.01	1.00	0.99, 1.01
Number of non-responses across all previous sweeps	0.62	0.61, 0.64	0.67	0.66, 0.69
Response to COVID-19 Wave 1 Survey				
Non-respondent	-	-	1.00	
Respondent	-	-	14.99	13.46, 16.69

Table A3. Estimated COVID-19 Wave 1 and 2 survey response models in BCS70 (n = 17,486).

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Sex				
Male	1.00		1.00	
Female	1.69	1.55, 1.85	1.37	1.25, 1.51
Voting				
Didn't vote	1.00		1.00	
Voted	1.30	1.12, 1.50	1.30	1.15, 1.47
Consent for biomarkers				
No to one/both	1.00		1.00	
Yes to both	1.17	1.00, 1.36	1.17	0.88, 1.54
Economic activity				
Currently employed	1.00		1.00	
Not currently employed	0.83	0.71, 0.97	0.96	0.79, 1.18
Self-rated health				
Excellent/very good	1.00		1.00	
Good	0.87	0.77, 0.99	0.98	0.87, 1.10
Fair/poor	0.81	0.71, 0.93	0.96	0.83, 1.12
Income quintile				
1	1.00		1.00	
2	1.16	0.99, 1.36	1.20	1.00, 1.45
3	1.30	1.12, 1.50	1.21	0.95, 1.54
4	1.45	1.21, 1.75	1.32	1.11, 1.57
5	1.43	1.13, 1.80	1.32	1.08, 1.61
Parental social class				
Professional/managerial	1.00		1.00	
Intermediate	0.95	0.84, 1.06	0.90	0.81, 1.00
Partly-/unskilled	0.96	0.85, 1.10	1.02	0.89, 1.17
Membership in organisations				

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
No organisations	1.00		1.00	
1 organisation	1.23	1.08, 1.40	1.08	0.95, 1.23
2+ organisations	1.21	1.02, 1.44	1.20	1.02, 1.41
Internet access prior to web survey				
None/little	1.00		1.00	
Medium	1.22	1.09, 1.38	0.99	0.86, 1.14
Lots	1.30	1.14, 1.48	1.04	0.91, 1.18
Educational qualifications				
None	1.00		1.00	
NQV Level 1-3	1.30	1.09, 1.55	1.00	0.85, 1.18
NVQ Level 4-5	1.44	1.19, 1.74	1.03	0.86, 1.23
Partnership status				
Never married/in CP	1.00		1.00	
Married/CP	1.07	0.95, 1.21	1.18	1.03, 1.35
Separated/divorced/widowed	0.96	0.81, 1.13	1.04	0.88, 1.22
Smoking status				
Never	1.00		1.00	
Former	0.93	0.84, 1.03	1.00	0.88, 1.13
Current	0.78	0.66, 0.93	0.85	0.70, 1.04
Social capital/social support: Frequency of meeting family and friends				
Never/rarely	1.00		1.00	
Fairly frequently	0.88	0.79, 0.99	1.05	0.92, 1.21
Very frequently	0.77	0.68, 0.86	0.87	0.76, 0.99
Social capital/social support: Have people around to listen to problems				
A little/not at all	1.00		1.00	
Somewhat	1.09	0.87, 1.38	1.00	0.78, 1.27
A great deal	1.04	0.81, 1.34	1.12	0.90, 1.38
Number of rooms at home (per room)	1.01	0.98, 1.04	1.01	0.98, 1.03

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Cognitive ability	1.36	1.26, 1.46	1.14	1.07, 1.20
Early life mental health	1.01	0.99, 1.02	1.01	0.99, 1.03
Psychological distress	0.97	0.94, 0.99	1.00	0.97, 1.03
Body mass index (kg/m ²)	1.01	1.01, 1.02	1.00	0.99, 1.01
Number of non-responses across all previous sweeps	0.66	0.64, 0.67	0.68	0.66, 0.69
Response to COVID-19 Wave 1 Survey				
Non-respondent	-	-	1.00	
Respondent	-	-	12.37	11.2, 13.65

Table A4. Estimated COVID-19 Wave 1 and 2 survey response models in Next Steps (n =15,770).

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Sex				
Male	1.00		1.00	
Female	2.12	1.90, 2.38	2.04	1.85, 2.25
Voting				
Didn't vote	1.00		1.00	
Voted	0.76	0.67, 0.86	0.80	0.71, 0.90
Membership in organisations				
Yes	1.00		1.00	
No	0.91	0.80, 1.03	0.95	0.85, 1.06
Economic activity				
Currently employed	1.00		1.00	
Not currently employed	0.79	0.67, 0.94	0.93	0.79, 1.09
Self-rated health				
Excellent/very good	1.00		1.00	
Good	0.88	0.77, 1.01	0.90	0.80, 1.01
Fair/poor	0.84	0.69, 1.04	0.87	0.70, 1.07
Income quintile				
1	1.00		1.00	
2	1.13	0.92, 1.38	1.04	0.88, 1.22
3	1.20	1.00, 1.45	0.99	0.84, 1.16
4	1.31	1.05, 1.64	1.07	0.90, 1.28
5	1.68	1.34, 2.10	1.06	0.88, 1.27
Parental social class				
Managerial	1.00		1.00	
Intermediate	0.90	0.78, 1.03	1.02	0.90, 1.15
Routine/semi-routine	0.77	0.66, 0.90	0.98	0.86, 1.11
Never worked	0.65	0.49, 0.86	0.92	0.76, 1.12

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Internet access prior to web survey				
None	1.00		1.00	
Little	1.07	0.86, 1.32	1.06	0.91, 1.23
Lot	1.29	1.04, 1.59	1.12	0.94, 1.33
Consent for linkages				
None	1.00		1.00	
Some	1.39	1.18, 1.63	1.15	1.01, 1.31
All	1.66	1.44, 1.92	1.08	0.94, 1.25
Educational qualifications				
None	1.00		1.00	
NQV Level 1-3	1.62	1.12, 2.34	1.13	0.84, 1.53
NVQ Level 4-5	2.10	1.46, 3.04	1.47	1.10, 1.96
Partnership status				
None	1.00		1.00	
Spouse/civil partner	1.01	0.83, 1.24	1.08	0.90, 1.28
Cohabiting partner	1.05	0.92, 1.21	1.05	0.91, 1.21
Smoking status				
Never	1.00		1.00	
Former	0.79	0.67, 0.93	0.83	0.7, 0.99
Current	0.73	0.63, 0.86	0.80	0.7, 0.91
Social capital/social support: How often meet up with family and friends				
Very frequently	1.00		1.00	
Fairly frequently	1.42	1.26, 1.60	1.24	1.10, 1.39
Rarely/never	1.44	1.17, 1.78	1.25	1.03, 1.51
Social capital/social support: Have people around to listen to problems				
A little/not at all	1.00		1.00	
Somewhat	0.84	0.63, 1.11	1.03	0.80, 1.33
A great deal	0.70	0.54, 0.90	0.86	0.68, 1.09

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Ethnicity				
White	1.00		1.00	
Indian/Pakistani/Bangladeshi	0.55	0.45, 0.68	0.88	0.74, 1.03
Black Caribbean/Black African	0.36	0.27, 0.48	0.64	0.51, 0.79
Mixed/Other	0.68	0.55, 0.85	0.86	0.71, 1.03
Early life mental health	1.02	1.00, 1.05	0.99	0.97, 1.01
Psychological distress	1.01	0.99, 1.04	1.02	1.00, 1.05
Body mass index (kg/m ²)	1.01	1.00, 1.03	1.01	1.00, 1.02
Social capital/social support: Trust scale	0.99	0.97, 1.02	1.02	0.99, 1.05
Number of non-responses across all previous sweeps	0.67	0.64, 0.70	0.67	0.65, 0.69
Response to COVID-19 Wave 1 Survey				
Non-respondent			1.00	
Respondent			13.92	12.16, 15.93

Table A5. Estimated COVID-19 Wave 1 and 2 survey response models in MCS (n = 19,243).

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Sex				
Male	1.00		1.00	
Female	2.93	2.66, 3.24	1.64	1.48, 1.81
Membership in organisations				
At least once a month	1.00		1.00	
Less than once a month	0.86	0.78, 0.95	0.98	0.88, 1.07
Economic activity				
Currently employed	1.00		1.00	
Not currently employed	1.01	0.88, 1.15	1.01	0.88, 1.15
Smoking status				
Never smoked	1.00		1.00	
Current/former/tried	0.61	0.53, 0.71	0.68	0.57, 0.80
Social capital/social support: Family and friends who help me feel safe, secure and happy				
Very true	1.00		1.00	
Partly true/not true at all	1.14	0.99, 1.32	0.83	0.70, 0.97
Social capital/social support: Someone I trust whom I would turn to if I had problems				
Very true	1.00		1.00	
Partly true/not true at all	1.02	0.89, 1.17	1.18	1.04, 1.35
Social capital/social support: No one I feel close to				
Very/partly true	1.00		1.00	
Not true at all	1.23	1.02, 1.47	0.94	0.79, 1.11
Self-rated health				
Excellent/very good	1.00		1.00	
Good	1.00	0.90, 1.11	1.01	0.90, 1.14

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Fair/poor	0.97	0.82, 1.13	1.08	0.92, 1.26
Income quintile				
1	1.00		1.00	
2	1.29	1.09, 1.54	1.10	0.89, 1.36
3	1.27	1.05, 1.55	1.04	0.83, 1.30
4	1.41	1.17, 1.71	1.05	0.87, 1.27
5	1.40	1.15, 1.69	1.06	0.87, 1.28
Parental social class (9 months)				
Managerial	1.00		1.00	
Intermediate	0.88	0.76, 1.02	0.83	0.71, 0.97
Routine/semi-routine	0.89	0.76, 1.03	0.72	0.61, 0.85
Parental social class (age 11)				
Managerial	1.00		1.00	
Intermediate	0.95	0.84, 1.07	1.00	0.87, 1.15
Routine/semi-routine	0.84	0.70, 1.00	1.05	0.84, 1.32
Internet access prior to web survey				
Little/none	1.00		1.00	
Medium	1.00	0.89, 1.14	0.94	0.84, 1.06
Lots	1.01	0.89, 1.15	0.88	0.77, 1.00
Educational qualifications				
None	1.00		1.00	
NQV Level 1-3	1.19	1.00, 1.42	1.20	1.01, 1.42
NVQ Level 4-5	1.38	1.13, 1.69	1.35	1.11, 1.64
Partnership status				
None	1.00		1.00	
Spouse/civil partner	1.18	1.01, 1.37	1.27	1.09, 1.47
Separated/divorced/widowed	1.26	1.05, 1.51	1.12	0.94, 1.34
Ethnicity				
White	1.00		1.00	

	Wave 1		Wave 2	
	OR	95% CI	OR	95% CI
Indian/Pakistani/Bangladeshi/Other Asian/Chinese	1.17	0.99, 1.39	1.08	0.90, 1.29
Black Caribbean/Black African/Other Black	1.05	0.79, 1.41	1.13	0.85, 1.51
Mixed/Other ethnic group	0.89	0.68, 1.16	0.98	0.75, 1.29
Number of rooms at home (per room)	0.97	0.94, 1.00	1.01	0.98, 1.05
Cognitive ability	1.37	1.27, 1.46	1.24	1.15, 1.35
Early life mental health	0.98	0.97, 0.99	1.00	0.99, 1.01
Psychological distress	1.01	1.00, 1.02	1.01	1.00, 1.02
Body mass index (kg/m ²)	1.00	0.99, 1.02	0.99	0.97, 1.00
Maternal mental health	0.98	0.95, 1.00	0.98	0.95, 1.00
Number of non-responses across all previous sweeps	0.42	0.39, 0.44	0.49	0.46, 0.52
Response to COVID-19 Wave 1 Survey				
Non-respondent	-	-	1.00	
Respondent	-	-	10.72	9.66, 11.90

APPENDIX 2 – Restoring sample representativeness – further examples

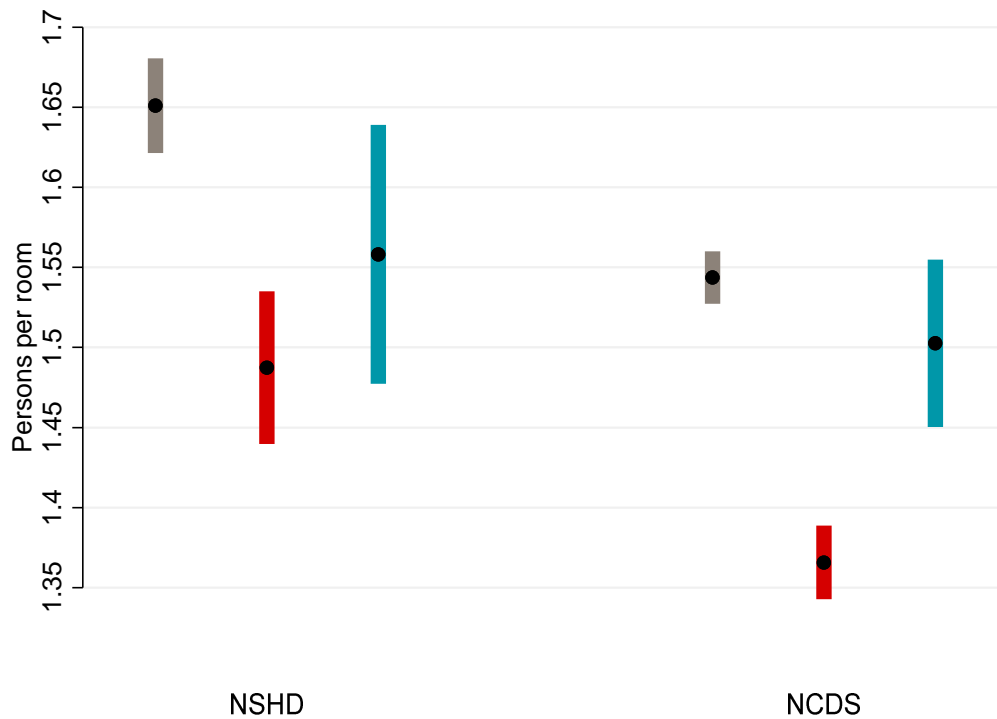


Fig. A2-1. Percentage of persons per room in NSHD and NCDS under different estimation approaches. **Grey:** using observed baseline data from the whole cohort; **red:** using observed baseline data from COVID-19 Wave 1 survey respondents only – unweighted (NCDS) or using design weight only (NSHD); **blue:** using observed baseline data from COVID-19 Wave 1 survey respondents only – weighted using non-response weights (in addition to design weights as appropriate).

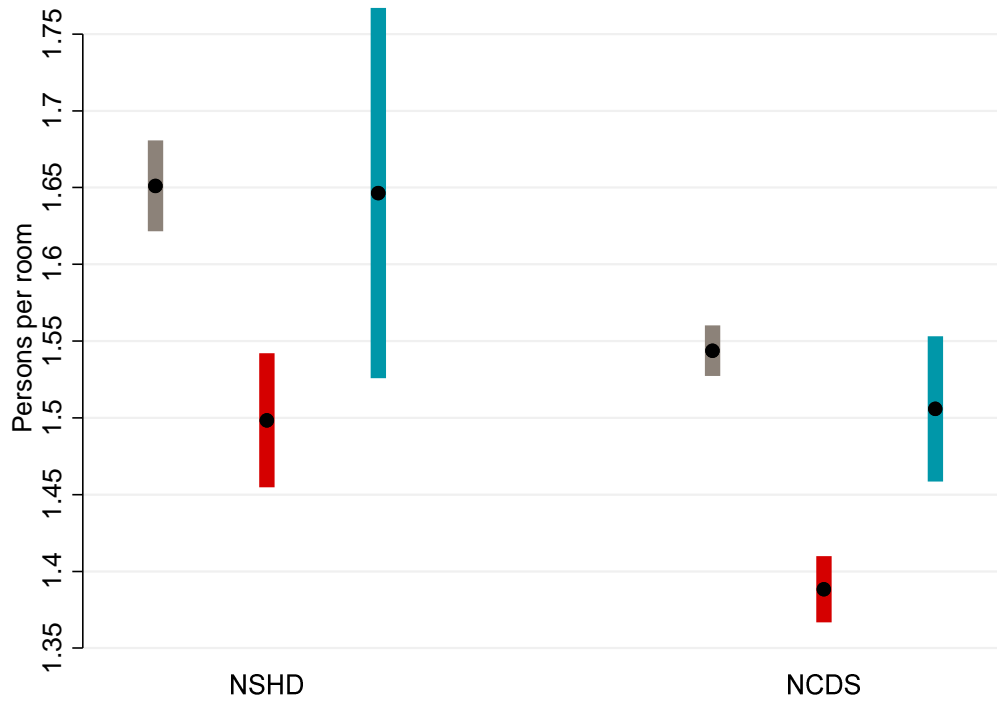


Fig. A2-2. Persons per room in NSHD and NCDS under different estimation approaches. **Grey:** using observed baseline data from the whole cohort; **red:** using observed baseline data from COVID-19 Wave 2 survey respondents only – unweighted (NCDS) or using design weight only (NSHD); **blue:** using observed baseline data from COVID-19 Wave 2 survey respondents only – weighted using non-response weights (in addition to design weights as appropriate).

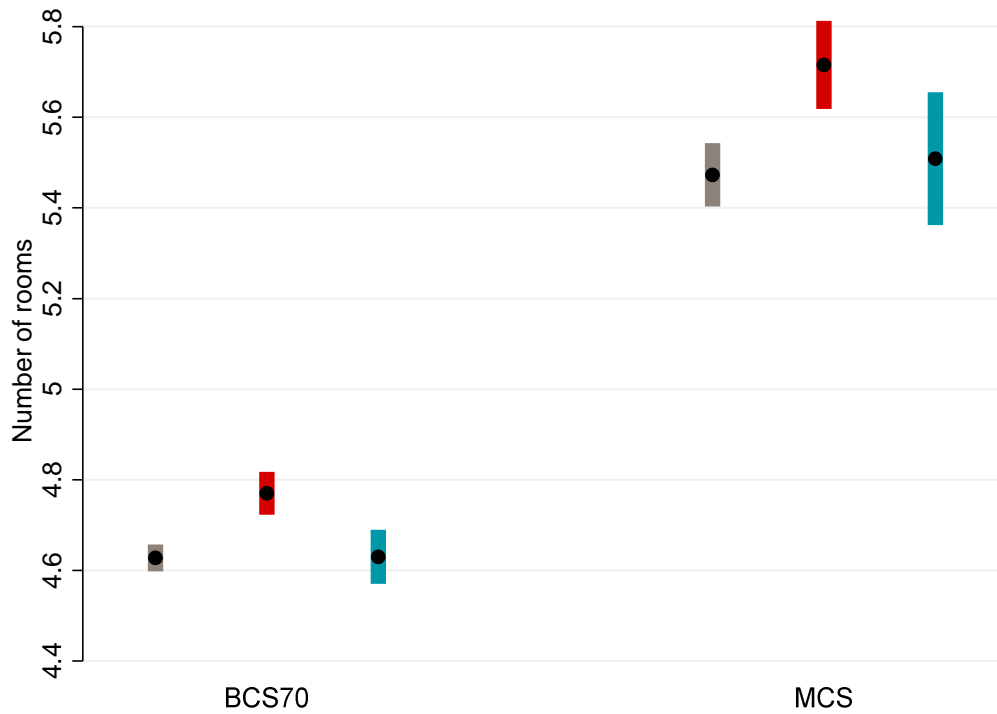


Fig. A2-3. Percentage of number of rooms in BCS70 and MCS under different estimation approaches. **Grey:** using observed baseline data from the whole cohort; **red:** using observed baseline data from COVID-19 Wave 1 survey respondents only – unweighted (BCS70) or using design weight only (MCS); **blue:** using observed baseline data from COVID-19 Wave 1 survey respondents only – weighted using non-response weights (in addition to design weights as appropriate).

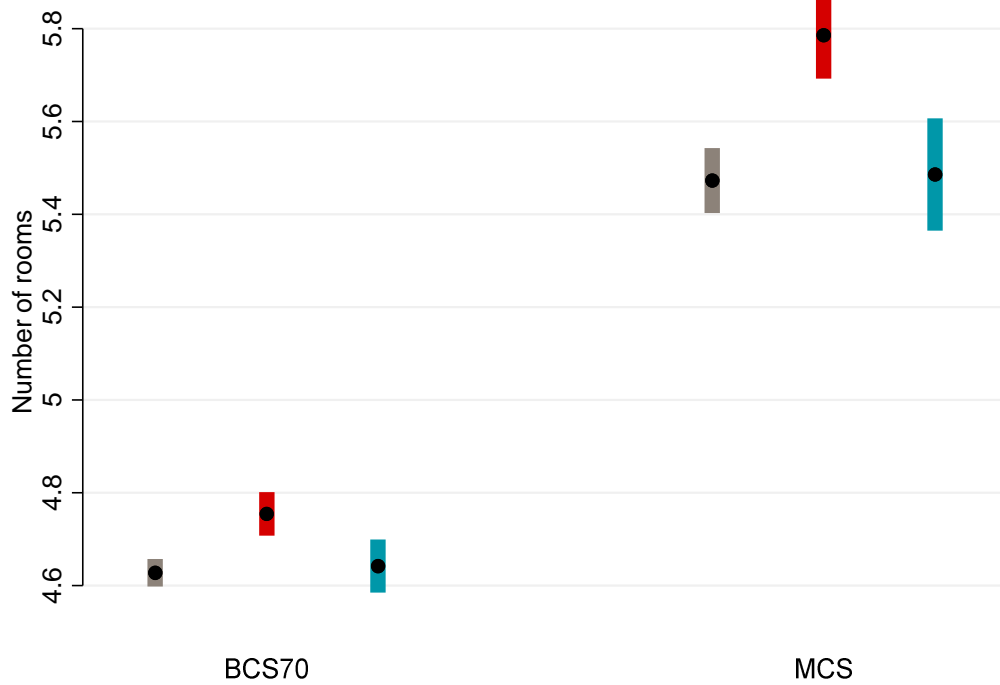


Fig. A3-4. Number of rooms in BCS70 and MCS under different estimation approaches. **Grey:** using observed baseline data from the whole cohort; **red:** using observed baseline data from COVID-19 Wave 2 survey respondents only – unweighted (BCS70) or using design weight only (MCS); **blue:** using observed baseline data from COVID-19 Wave 2 survey respondents only – weighted using non-response weights (in addition to design weights as appropriate).

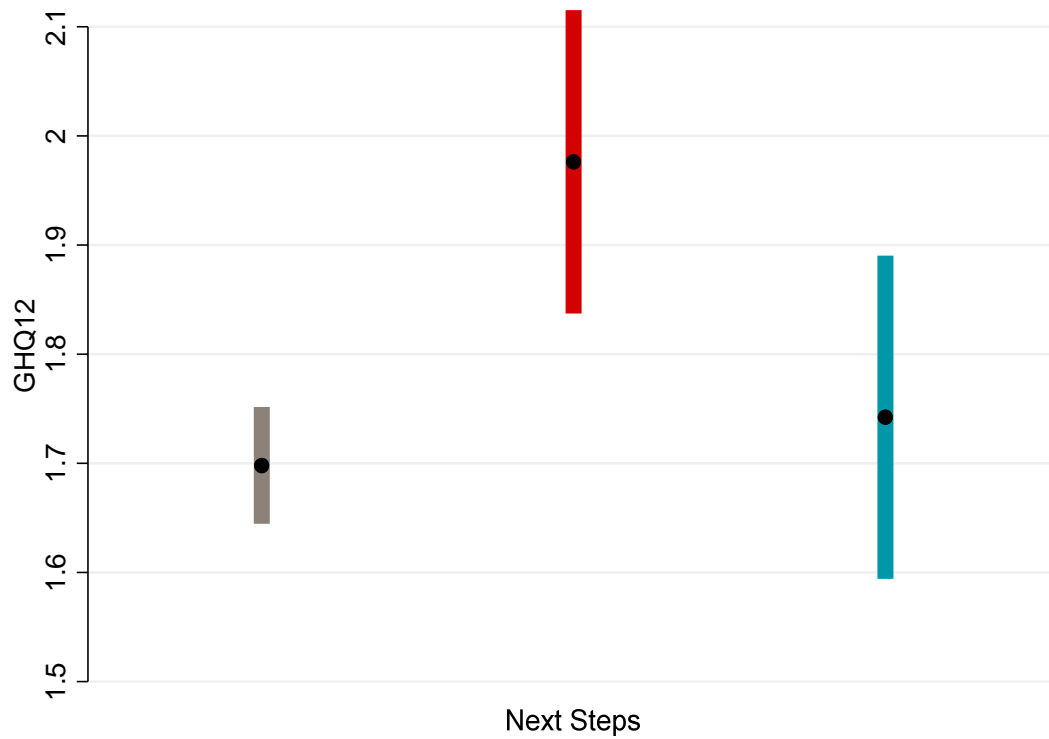


Fig. A3-5. GHQ12 psychological distress score in Next Steps under different estimation approaches. **Grey:** using observed baseline data from the whole cohort; **red:** using observed baseline data from COVID-19 Wave 1 survey respondents using design weight only; **blue:** using observed baseline data from COVID-19 Wave 1 survey respondents only – weighted using non-response weights in addition to design weights.

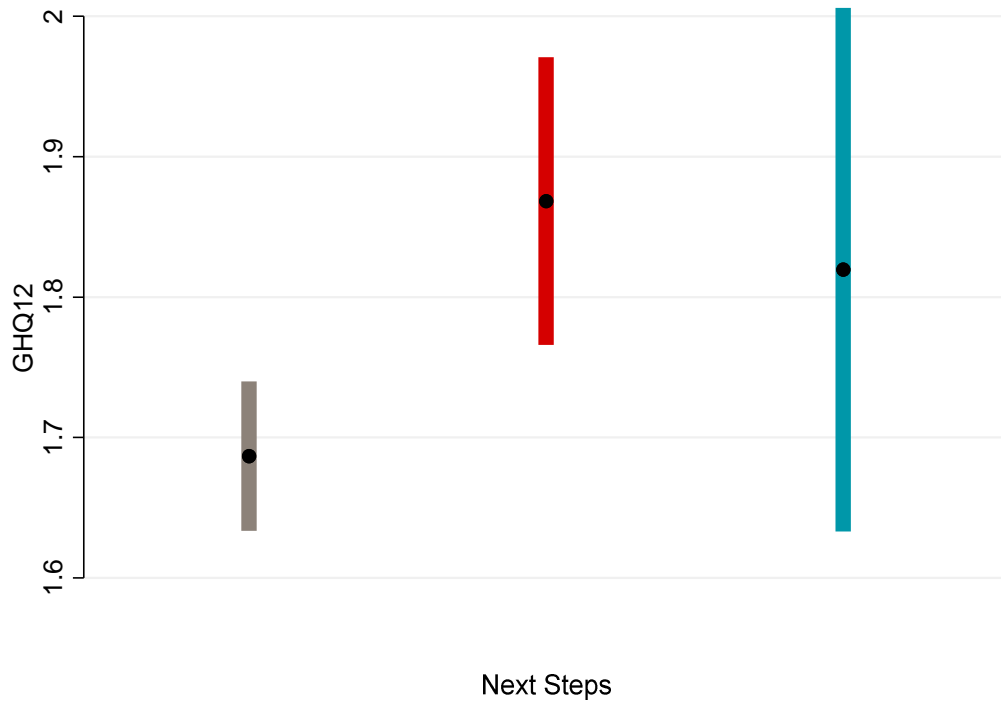


Fig. A3-6. GHQ12 psychological distress score in Next Steps under different estimation approaches. **Grey:** using observed baseline data from the whole cohort; **red:** using observed baseline data from COVID-19 Wave 2 survey respondents using design weight only; **blue:** using observed baseline data from COVID-19 Wave 2 survey respondents only – weighted using non-response weights in addition to design weights.