UCL

1970 British Cohort Study

Age 51 Survey

User Guide (Version 1)

February 2025

CENTRE FOR LONGITUDINAL STUDIES

UKRI Economic and Social Research Council

Centre for Longitudinal Studies

> Centre for Longitudinal Studies (CLS)
>
> UCL Social Research Institute
>
> University College London
>
> 20 Bedford Way, London WC1H 0AL
>
> cls.ucl.ac.uk

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre. It is home to a unique series of UK national cohort studies. It is part of the UCL Social Research Institute, based at the IOE, UCL's Faculty of Education and Society.

This document is available in alternative formats. Please email the Centre for Longitudinal Studies at clsdata@uck.ac.uk

Contents

About the 1970 British Cohort Study

The 1970 British Cohort Study (BCS70) is a longitudinal birth cohort study, following a nationally representative sample of over 17,000 people born in Britain in a single week in 1970.

We have surveyed cohort members throughout their childhood and adult lives, mapping their individual trajectories and creating a unique resource for researchers. It is one of very few longitudinal studies following people of this generation anywhere in the world.

Featuring a range of objective measures and rich self-reported data, BCS70 covers an incredible amount of ground and can be used in research on many topics

Evidence from BCS70 has illuminated important issues for our society across five decades. Key findings include how reading for pleasure matters for children's cognitive development, why grammar schools have not reduced social inequalities, and how childhood experiences can impact on mental health in mid-life.

Everyday researchers from across the scientific community are using this important study to make new connections and discoveries.

Important note about figures in this document

Figures that are presented in this document may vary in comparison with the deposited data and figures reported in the BCS70 Technical Report. This happens for various reasons: requests for data deletion, resolution of duplicate cases, data editing and quality checking which can result in the removal of cases.

1. Introduction

The BCS70 Age 51 Survey, (or 'Life in Your Early 50s' Survey as known to study members) was conducted between 2021 and 2024 when participants were aged 51-53 years. It was the 11th sweep of BCS70.

This sweep was designed and managed by the Centre for Longitudinal Studies (CLS) at the UCL Faculty of Education and Society (IOE) and fieldwork was conducted by NatCen and Verian (formerly Kantar).

The main aim of the Age 51 Survey was to gather information to understand midlife outcomes across various life domains and their determinants over the life course, building on the wealth of data collected in previous sweeps. The study was initially planned and designed to be conducted face-to-face with fieldwork commencing in 2020. However, the COVID-19 pandemic meant that face-to-face interviewing was not feasible from May 2020 until early 2022. As such the protocol was changed so that interviews could be conducted by video-call. Fieldwork was launched using a video-only approach in November 2021. Face-to-face interviewing began in April 2022, though participants could still opt to participate via video if they preferred. Once mainstage fieldwork was complete, those who had not participated were invited to complete a short version of the questionnaire via web (known as the 'mop-up' survey). Emigrants were also invited to take part in this short web-survey.

A total of 12,041 cohort members were initially issued for fieldwork, followed by an additional 474 cohort members living abroad, issued to the 'mop-up' only phase. A total of 8,016[1] cohort members were interviewed and have data deposited, representing a response rate of 65%[2]. Overall, 3,812 study members took part in-person, 3,498 by video call, 29 by telephone and 677 as part of the web mop-up survey (227 of these cases were cohort members living abroad).

---

[1] 9 achieved interviews were removed prior to the data deposit following data checking and processing

[2] Response rate base is 12,412 cases (12041 cases issued to the pilots and mainstage fieldwork plus 474 emigrant cases with 103 confirmed ineligibles removed).

A full account of the survey development and fieldwork procedures can be found in the 1970 British Cohort Study technical report and appendices produced by NatCen Social Research, which accompanies this data deposit.

This user guide provides information about the data arising from the 1970 British Cohort Study Age 51 Survey and accompanies the deposit of the data at the UK Data Service. In addition to this user guide the Age 51 Survey data deposit includes:

- BCS70 - Age 51 Survey: Questionnaire
- 1970 British Cohort Study: Life in Your Early 50s – Self-completion Questionnaire
- BCS70 - Age 51 Survey: Technical Report

Data, questionnaires and user guides for all previous sweeps are also available at UKDS. All datasets use a common ID – BCSID.

2. History and background

The 1970 British Cohort Study (BCS70) is a renowned national longitudinal birth cohort study run by the Centre for Longitudinal Studies (CLS) at the UCL Social Research Institute.

BCS70 started by collecting data on babies born in the UK during one week in 1970. Since then, there have been ten further surveys, collecting information from participants in England, Scotland, and Wales. The focus of these surveys has expanded over time. Childhood sweeps focussed on physical, educational and social development (Birth, Age 5, 10, 16 sweeps), moving onto to capture their economic development along with family and relationships as they move into adulthood and beyond (Age 26, 30, 34, 38), through to mid-life circumstances, health and planning for older age (Age 42, 45, 51 sweeps).

In addition to these core surveys, three web surveys were conducted between 2020 and 2021 to explore the impact of the COVID-19 pandemic on participants.

The Centre for Longitudinal Studies (CLS) at the UCL Social Research Institute has managed the study since 1991, with funding from the Economic and Social Research Council (ESRC).

The chart below shows the number of interviews completed in each BCS70 survey. Most surveys were conducted in-person, except:

- Age 26: Postal survey
- Age 38: Telephone survey
- Age 51: Mostly by video call or in-person, with some completing a short web survey

Figure 1: Number of interviews per sweep of BCS70

Sweep	Number of interviews
1970 (Birth)	16,589
1975 (Age 5)	13,135
1980 (Age 10)	14,870
1986 (Age 16)	11,615
1996 (Age 26)	9,003
2000 (Age 30)	11,261
2004 (Age 34)	9,665
2008 (Age 38)	8,874
2012 (Age 42)	9,841
2016 (Age 46)	8,581
2021 (Age 51)	8,016
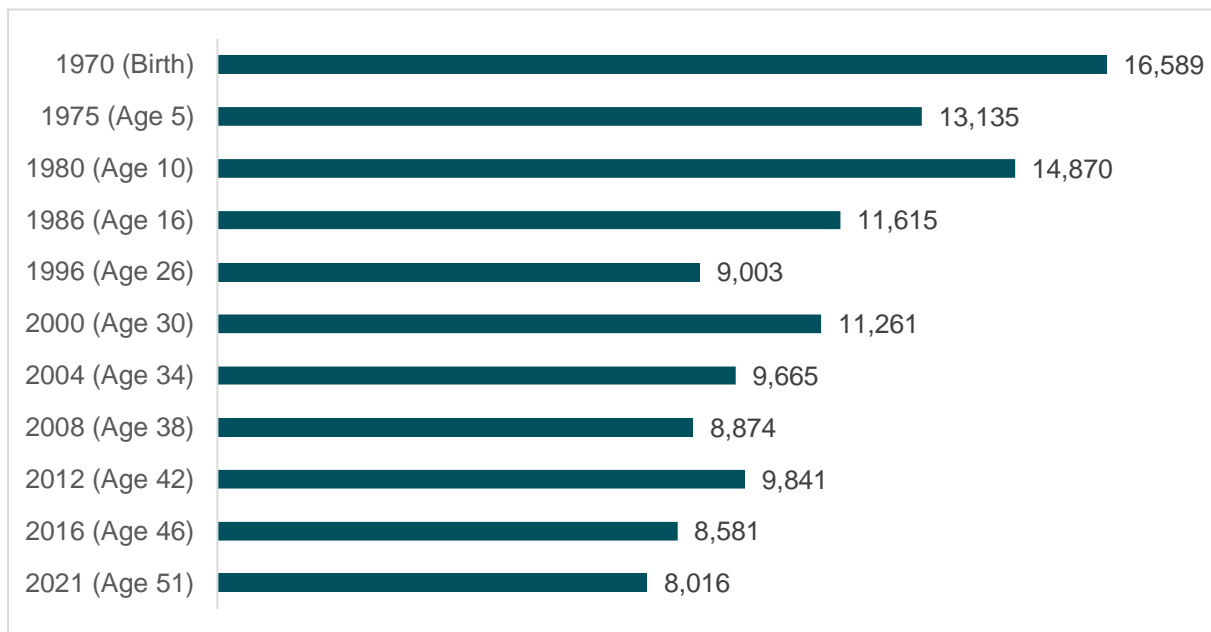
3. The Age 51 Survey

The main aim of the Age 51 Survey was to gather information to understand midlife outcomes across various life domains and their determinants over the life course. This survey builds on the extensive data collected since the participants' birth, allowing comparisons with other generations, such as the 1958 National Child Development Study (NCDS) cohort at age 50 and the 1946 National Survey of Health and Development (NSHD) cohort at age 52.

The Age 51 Survey collects information on many aspects of cohort members' lives including employment, relationships, health and wellbeing, cognition, politics, attitudes, caring responsibilities, as well as capturing pensions and finances.

The Age 51 Survey involved asking study members to complete an interview (mostly either in-person, or by video call with non-responders and emigrants being asked to complete a shortened web version). The main interview included a self-completion section containing the most sensitive questions, cognitive assessments and consent to data linkage. In addition to the main interview, participants were asked to complete a paper-self-completion questionnaire and an Online Dietary Questionnaire (ODQ).

This sweep was funded by the Economic and Social Research Council (ESRC). The lead fieldwork contractor was NatCen, who subcontracted Verian (formerly Kantar) to jointly conduct fieldwork.

Ethical approval was provided by London – Brighton & Sussex Research Ethics Committee. REC reference – 19/LO/1757.

4. Fieldwork

4.1 Issued sample

The BCS70 cohort includes all babies born in Great Britain during one week in 1970. During the childhood sweeps (at ages 5, 10 and 16), additional children born outside Great Britain during that week but who later moved to Britain were included. Information was collected in the birth sweep about children born in Northern Ireland, but they were not followed up in later sweeps (unless they moved to England, Scotland or Wales during childhood).

The initial issued sample for the Age 51 survey included 12,041 cohort members[3]. All participants who had taken part in previous sweeps were issued except 1) those who had previously indicated they did not wish to be contacted, 2) those who were known to have died, 3) those who were permanently untraced, and 4) those who were in prison or on probation.

A further sample of 474 cohort members known to be living abroad (also known as 'emigrant' cases) were only issued to the 'mop-up' phase. These cases were not eligible for the pilots and main stages of the survey.

4.2 Fieldwork period

Fieldwork for the main stage of the Age 51 Survey ran from 29th June 2021 until the 7th November 2023.

4.3 Fieldwork stages

The Age 51 Survey was conducted in several stages. The main stage fieldwork was preceded by two pilot stages. The Age 51 Survey was initially planned to be

---

[3] 29 cases from this issued sample were reviewed following fieldwork and re-classified as not being valid cohort members (ineligible).

conducted using a purely face-to-face approach. As such, the first pilot study, which took place in early 2020, was conducted in this way.

Main stage fieldwork was due to commence in June 2020, but the COVID-19 pandemic meant that this was not possible as face-to-face interviewing was suspended. There was considerable uncertainty as to when face-to-face interviewing could be resumed and as such, it was decided to adapt the survey so that it could be conducted by video-call. A second pilot study was conducted which sought to test the feasibility of conducting the survey by video.

Mainstage interviewing was initially launched using a video-only approach until it was feasible to re-launch face-to-face interviewing. Finally, once mainstage interviewing was complete, an online 'mop-up' survey was conducted where non-respondents (and emigrants) were invited to complete an abbreviated web version of the survey.

The two pilot stages were both conducted with BCS70 study members. Data collected during the pilots has been included in the deposited data (with the exception of any questions which were removed or significantly changed after the pilots).

Each of the stages is described below. Variable '**b11survey_stage'** in the dataset denotes the stage a case was completed in.

4.3.1   In-person pilot

The in-person pilot ran from January to mid-March 2020.

177 BCS70 cohort members were issued who were purposively selected to mirror the mainstage sample in terms of gender and past participation. The pilot sample oversampled cohort members who had previously declined data linkage in order to test a new procedure for recording consent in the CAPI programme instead of using paper consent forms as was used in previous sweeps. 116 full interviews were completed.

The pilot included all the main survey components as well as two additional cognitive tests, 'serial 7s' and 'counting backwards', which were dropped after the pilot. It was also agreed not to continue the module on Screentime Usage, which involved cohort members with Apple and Android phones retrieving usage data from their phones.

The average interview length of the pilot interview was 110 minutes, therefore a number of questions were removed after the pilot to reduce the length of the survey by 25 minutes. There was also some re-organisation of questions and sections to improve the overall flow of the interview.

4.3.2 Video Pilot

Fieldwork for the video pilot took place from the 9th September until the 7th October 2020. It was conducted to assess the feasibility of conducting the survey via video-call. Objectives included testing technology suitability, interview length, self-completion sections, and cognitive assessments.

The video pilot included 60 cohort members. Given that the key aim was to assess the feasibility of conducting a video interview, the issued sample was comprised of participants who had previously expressed willingness to be interviewed in this mode. A total of 44 interviews were completed. Video-interviews were conducted via MS Teams. The video interview mirrored the CAPI interview, with adjustments for the video format. These adjustments included modifying respondent-facing materials and administering the self-completion section as a web survey during the video call.

The letter cancellation test was removed from the video pilot. The letter cancellation ordinarily comes after the Animal Naming task, so between the initial Word Recall test and the Delayed Recall Test. This means that the gap between Word Recall and Delayed Word Recall was shorter in this video pilot than it is typically.

The video pilot confirmed that video interviews were feasible, which was further confirmed via a subsequent larger video pilot conducted with participants from the National Child Development Study (NCDS).

4.3.3 Soft Launch Conducted at Wave 1

Following the success of the video pilots, a soft launch was conducted between 29th June and 14th September 2021, also referred to as Wave 1 of the mainstage survey. The soft-launch was conducted with a video-only approach. A total of 1,067 cases were issued and 557 interviews completed. Participants with no email or telephone number were excluded from this stage as interviewers were only able to make contact using these methods.

Some adjustments were made after the video-pilot which included sending an "interview pack" with necessary materials before the interview (after an appointment had been made). This "interview pack" included the document required for the letter cancellation test which was re-introduced to the survey, along with the Paper Self Completion Questionnaire.

As per the pilot, the sensitive questions were asked via a web-survey. Participants were sent a link to the web-survey using the MS Teams chat function and asked to complete it while the interviewer remained on the call. If they were unable or unwilling to do this, they could complete the web survey at a later point.

Some adaptions were made to the questionnaire which included the addition of a module covering experience of COVID-19.

4.3.4 Main Stage

Waves 2-4 were also conducted using a video-only approach. Fieldwork for these waves took place between the 28th October 2021 and the 21st June 2022.

One key adaption made from Wave 2 onwards was to give study members the option to complete the sensitive questions with assistance from the interviewer. If a study member was unwilling to complete the sensitive questions web survey during the interview, the interviewer would then provide the option to share their screen for the participant to read the questions and provide the interviewer with the number of their response, for the interviewer to then enter into the CAPI programme. A final option remained for the study members to complete the web survey after the interview (with an e-mail being sent the day after the interview with a link to the web survey).

All contact and tracing in these video call only waves took place remotely (via email and telephone).

Waves 5-7 were primarily conducted face-to-face, as COVID restrictions had been lifted by this time. Fieldwork in these waves took place between 17th May 2022 and 25th May 2023. Participants who preferred not to be visited by an in-person interviewer could still opt to be interviewed by video. In addition, participants living in areas where no interviewer was available were still invited to take part by video.

Moving to in-person interviews meant the following changes were made to the design and administration of the survey:

- Updated training for in-person interviews.

- Sensitive questions were completed privately by participants on a laptop during the interview – Computer Assisted Personal Interviewing (CASI) (they could request to be assisted by the interviewer if necessary).

- All paper documents, including the paper questionnaire and letter cancellation document could be provided in person and collected by the interviewer (an option was still available to post these back).

- Interviewers could conduct in-person visits for contact and tracing.

In total 10,824 cases were issued to the main stage, 2,897 interviews were conducted by video and 3,696 took place in-person. 29 interviews were conducted by telephone. Interviews were not intended to be conducted by telephone. In the main these were interviews allocated to an in-person or video mode but due to technical problems, a telephone interview was permitted and the interview data has been retained.

4.3.5   Re-issues and reallocations

Between 8th August 2022 and the end of fieldwork on the 7th November 2023, cases that were unproductive at the main and pilot stages were re-issued or reallocated. Re-allocated cases were those that were initially approached by a video interviewer and had requested an in-person interview. A total of 4,355 (36%) of cases were reissued (or reallocated) across three batches. A total of 1,324 productive interviews were achieved during this stage.

Cases who were approached by an in-person interviewer and requested a video call interview were reallocated through-out the in-person fieldwork period.

4.3.6   Mop-up Survey

After the mainstage fieldwork, a web-based "Mop-up Survey" was conducted between December 2023 and January 2024. This stage included non-responders from the pilot and main stage as well as emigrants who were not eligible to take part in the main

stage survey. A total of 3,773 non-responders were invited to take part and 450 responded. A total of 474 emigrant cases were invited to take part and 227 responded.

This survey was a shortened version of the main survey, lasting around 25 minutes, with adaptions made to allow the questions to be administered as a web survey. This phase excluded cognitive assessments and data linkage consent. Participants were also not provided the paper self-completion questionnaires to complete.

4.3.7   Fieldwork timescales and mode

Table 1 below provides a summary of the fieldwork start and end dates along with main mode of issue during this time (video interviews were also conducted on request/when an in-person interviewer was not available during all stages, except for the mop-up which was web only).

Table 1. All stages and fieldwork dates for first issue and reissue cases

	Fieldwork mode	Start Date	End Date
Soft Launch (Wave 1)	Video-only	29/06/2021	18/09/2021
Wave 2	Video-only	28/10/2021	31/01/2022
Wave 3	Video-only	09/12/2021	30/03/2022
Wave 4	Video-only	01/03/2022	21/06/2022
Wave 5	In-person	17/05/2022	30/09/2022
Wave 6	In-person	06/09/2022	31/07/2023
Wave 7	In-person	25/10/2022	25/05/2023
Reissue / Reallocation (Batch 1)	In-person	08/08/2022	07/11/2023
Reissue / Reallocation (Batch 2)	In-person	27/04/2023	07/11/2023
Reissue / Reallocation (Batch 3)	CAPI In-person	26/05/2023	07/11/2023
Mop-up	Web	07/12/2023	14/01/2024

The mode the interview was conducted in can be identified in the data by variable **'b11survey_mode'**. The data also provides information on the date of interview across three variables – **'b11intd'**(day), **b11intm**' (month) and '**b11inty**' (year).

4.4   Contact Strategy

At the start of each wave (including the pilots and soft launch), NatCen sent advance letters and emails to all cohort members informing them of the latest survey. The

wording of the advance letter varied slightly depending on whether the cohort member had taken part in the previous sweep of the survey or prior to 2008. The letter for waves 1 to 4 invited cohort members to have a video interview although noted if they would prefer an in-person interview this could happen later. The letters used in waves 5 to 7 outlined that the interview could take part in-person or by video interview.

Included with the advance letter was an eight page booklet which provided more detailed information about the survey. Emails included a link to an electronic version of the leaflet.

Following the advance mailing, interviewers would then contact the study members. How contact was made would vary depending on whether the case was allocated to a video call only wave or in-person wave. In the video call waves interviewers first contacted cohort members by telephone, making at least three attempts before then sending a text or email.

In the in-person waves, approximately 80% of cases were first contacted by telephone. This group included those who had participated in previous surveys and had provided a telephone number. If telephone contact was unsuccessful, interviewers made personal visits.

For the remaining 20% of cases, initial contact was made through personal visits. If this failed, interviewers could attempt telephone contact if a number was available. Interviewers left calling cards at unattended homes, offering a Freephone number for rescheduling or opting out.

For reissue waves, no new advance letters or emails were sent initially, as cases were allocated gradually. However, from spring 2023, interviewers were provided with advance letters to send before making contact, following feedback from study members.

For the mop-up, study members were also sent an advance letter and e-mail, including a booklet/link to a booklet in advance of the survey. Those living abroad were only sent an e-mail invite. This correspondence was tailored to different respondent types (as 'emigrants' had not been invited to take part in the main stage). Study members were provided with a link to the survey and could also access a

short video (via a link) to the Study Director explaining more about the importance of participating.

A thank you letter, and e-mail was sent to those who had completed the survey. This included confirmation of data linkage consents provided during the interview.

4.4.1   Tracing Strategy

Whenever interviewers identified that cohort members had moved from the issued address, they carried out tracing activities, which primarily consisted of making calls to partners and stable contacts or asking current occupiers of the cohort member's address if they had a forwarding address (see section 6.7 of the BCS70 Technical Report: Life in Your Early 50s Survey). For video call only waves, all tracing had to be conducted remotely (no visits could be made at addresses). Participants that interviewers were unable to trace were sent to CLS who attempted to find new contact details via office-tracing.

All tracing activities were completed before marking cases as non-contact.

4.4.2   Reminder Strategy

Once an interview was arranged an appointment letter was sent (or provided) by interviewers to the study member (this letter also contained any further documents needed for the interview which varied depending on whether it was a video call or in-person interview).

In addition to this paper appointment letter an appointment reminder was also sent the day before the interview by e-mail and text (depending on contact information available). When an appointment was made for a video call interview, the interviewer would telephone the study member the day before the interview to confirm they had received the documentation needed and had all they needed to access the MS teams call.

Reminders were also sent following the interview by e-mail and text to remind the study member to:

- Ask their partner to complete the data linkage questionnaire (reminders to complete this were also sent directly to the partner)

- To complete the sensitive questions web survey (if not already done so during the interview – video only cases)

- To complete the Online Dietary Questionnaire

- To return the Paper Self Completion Questionnaire and Letter Cancellation document (if not collected by the interviewer)

For the mop-up survey – three electronic reminders were sent (by e-mail and text simultaneously) during the fieldwork period as well as one reminder letter.

Further detail on contact procedures can be found in the BCS70 Age 51 Survey: Technical Report.

5.   Questionnaire

5.1   Overview

The Age 51 Survey questionnaire contained a number of components including cognitive testing, a self-completion section consisting of sensitive questions, consent to data linkage and consent to take part in an online dietary questionnaire following the interview. Cohort members were also asked to complete a paper self-completion questionnaire before or after the interview, which was collected by the interviewer (at the end of the interview) or posted back.

The in-person and video interviews used the same questionnaire and Blaise 4 programme.

Slight adaptions were made to accommodate video interviewing, including:

- Adaptions to interviewer instructions – including use of 'show screen' when required.

- New section for video interviewers to record if and how the web survey containing sensitive questions would be administered e.g. independently during the interview, with assistance from the interviewer (showing their screen) or sent in an email to be completed after the interview.

- Instructions to use the letter cancellation document which was sent in advance

The interview lasted around 80 minutes on average.

The mop-up web survey was a shortened version of the main stage questionnaire, taking an average of 25 minutes to complete and did not include any of the additional elements (cognitive assessments, consent to data linkage or consent to complete the Online Dietary Questionnaire). A selection of the 'sensitive' questions were included in this survey.

5.2   Proxy, interpreter and partial interviews

The final data includes both partial and fully completed interviews. An interview was classed as partial if the respondent did not complete the interview but answered the last question in the household grid section. Partial interviews can be identified using the variable **'b11outcome'** [4]. 'Full' interviews are contingent on the completion of the core interview but do not take into account completion of the various 'Special elements' noted in section 5.4.

In cases where the cohort member was unable to understand survey questions or communicate the answers for themselves, a proxy interview could be conducted with a carer. This took them through a shorter route of the questionnaire lasting approximately 45 minutes. The survey was tailored to account for someone else participating on behalf of the cohort member, and did not include the modules on sensitive questions, cognitive testing, data linkage, the online dietary questionnaire, and the self-completion paper questionnaire.

If a respondent was able to participate themselves but just needed some assistance with communication (e.g. due to hearing or speech disability), an interpreter could be used, and the full interview would be conducted (unlike in a proxy case).

Proxy and interpreter interviews can be identified in the data using variable 'b11intwho'.

---

[4] To select partial interviews - b11outcome=2 (partially productive)

5.3   Questionnaire Content

5.3.1   Computer Assisted Interview questionnaire content

Outlined below is a summary of the questionnaire content.

Table 2. Questionnaire Modules

Questionnaire Module	Content	Included in the Mop up
Household grid	Cohabiting partnerships, children (including those not in the household) and any other household members.	This section was largely the same as the mainstage but included a small additional section for emigrants about when and why they left Great Britain
Family	Non-cohabiting relationships, grandchildren, parents, and social contact.	No
Housing	Cohort member's housing history and current accommodation.	A short section collecting information on cohort member's current accommodation (housing history was not collected)
Employment	Cohort member's current and previous economic activity (including economic activity history, education or training, unemployment, retirement, sickness or disability, and looking after the home or family) hours and pay and any changes in working practices due to the pandemic. Cohabiting partner's current economic activity income from employment, and any changes to working practices due to the pandemic.	Collected information on cohort member's current economic activity (including employment, education or training, unemployment, retirement, sickness or disability, and looking after the home or family) Cohabiting partner's current economic activity was also recorded. (Previous economic activity was not collected)
Income	Income from benefits, tax credits/allowances, pensions, investments and savings, inheritances and gifts received and debt for both the cohort member and their cohabiting partner.	Collected details on total income of the household from earnings, benefits, and any other form of earnings. One question replaced the detailed financial questions asked in the main stage
Cognitive Function	Five short memory, concentration and knowledge tasks (see section 5.4)	No
Lifelong Learning	Any new academic or vocational qualifications gained by the cohort member since the last interview (or 1st Jan 2012). Partner's highest qualification and age left education.	No
Health	Physical and mental health, wellbeing, dental health, hospital	Asked general questions around physical and mental health and

Questionnaire Module	Content	Included in the Mop up
	visits, exercise, drinking, smoking self-reported height and weight	wellbeing and collected a self-reported weight measurement
COVID-19	Whether experienced symptoms of COVID-19 and long COVID, if they had had a positive test, and if they had been vaccinated.	Shortened version covering same core areas, excluding more detailed information such as how tested for COVID_19/when had positive test/type of vaccine received
Sensitive Questions	Political attitudes and voting, relationships, share of household tasks, mental health, children who have died, unsuccessful pregnancies, gynaecological problems, menstruation, experience of domestic violence and life satisfaction.	Included questions on mental health and life satisfaction only
Data linkage consent	Consent was sought to link survey data to government health and economic records where not obtained at the Age 42 survey. Permission from both the cohort member and their cohabiting partner was sought.	No
Online Dietary Questionnaire (ODQ)	Recruitment to take-part in an online dietary questionnaire for two days following the interview.	No
Contact Information	Updating contact details for the cohort member, partner and stable contacts. New contact details were also collected if the cohort member was planning to move. If the interview was carried out by proxy, the contact details of the person who acted as proxy were collected.	This section was largely the same as the section included in the mainstage survey

5.3.2 Event Histories

There were three event histories included in the interview: a relationship history, a housing history, and an economic activity history.

- Cohort members that had been interviewed in at least one of the last three sweeps (Age 38, 42 or 46 sweeps) were asked to update their cohabiting relationship history. Cohort members that had not been interviewed in the last three sweeps were asked to update their situation from 1st January 2008.

- Cohort members that had been interviewed in one of the last two sweeps (at either the Age 42 or Age 46 sweeps) were asked to update their housing and economic situation from the date of their last interview. Cohort members that

had not been interviewed in the last two sweeps were asked to update their situation from 1st January 2012.

- Those taking part in the mop-up web survey were not asked their housing, or economic history and were only asked to provide information about their current circumstances.

5.4   Special elements

The special elements noted below were only asked of those taking part in the main and pilot in-person and video interviews and were not part of the mop-up web survey.

5.4.1   Cognitive function tasks

Cohort members were asked to undertake five different cognitive assessments. The interviewer asked for consent from the study member prior to each assessment being administered.

The tasks were designed to measure different aspects of cognition and have been included in various other studies such as the National Child Development Study (NCDS) and the English Longitudinal Study of Ageing (ELSA).

Immediate word-list recall

This tested verbal learning and recall. Cohort members listened to a list of 10 words. They were then asked to recall the words immediately. In most cases, the list was read out by the computer using a recorded voice.

In video interviews the interviewer shared their screen with sound over Microsoft Teams so the cohort member could hear the recording of the 10 words they needed to recall.

In some cases, where the cohort member could not hear the recorded voice, the interviewer read out the list.

Variable name	Variable label
b11cflisn	Immediate word recall number *Number of words recalled (0-10)*

Animal naming

This tested how quickly cohort members could think of words from a particular category. Cohort members were asked to name as many different animals as they could think of in one minute. The timing was controlled by the computer. Interviewers recorded the number of animals the cohort member said, not counting any repetitions.

Variable name	Variable label
b11cfani	Animal naming: – number of animals named *Number named within one minute*

Letter cancellation

This tested attention, mental speed, and visual scanning. Cohort members were given a page of random letters of the alphabet arranged in a grid and were asked to cross out as many "P's and "W's as possible in one minute. They were then scored on both how accurately they completed the task, and how far along the grid they managed to get within one minute.

In video interviews, the letter cancellation sheet was placed in a sealed envelope and posted out before the interview. On the envelope were clear instructions not to open the envelope before being asked to by the interviewer during the interview. The cohort member then posted back the sheet.

Scoring of this assessment was conducted in the office, rather than by interviewers.

Variable name	Variable label
b11cfrc	Letter cancellation (Computed) – speed score (number of letters scanned)
b11cflot	Letter cancellation: (Computed) Total number CORRECT + MISSED by respondent
b11cfmis	Letter cancellation (Computed) – accuracy score (number of Ps and Ws missed)

Delayed word-list recall

This tested short term memory. Cohort members had 2 minutes to recall as many words as they could remember from the list they heard during the first word recall test. They were not permitted to listen to the list again. Interviewers made a note of each word correctly recalled and entered the total number into the questionnaire programme.

Variable name	Variable label
b11cflisd	Delayed recall words: number *Number of words recalled (0-10)*

In the video pilot stage, the delayed word recall period was slightly shorter as this stage did not include the Letter Cancellation test prior to this cognitive test. Study members who opted out of taking part in the animal naming or letter cancellation test would also have a shorter time period between the immediate and delayed recall tests.

National Adult Reading Test (NART)

This tested knowledge of vocabulary. The words chosen for the test are intentionally challenging and unphonetic to test an individual's vocabulary rather than their ability to apply regular pronunciation rules. Cohort members were shown 25 unphonetic words and were asked to read them out loud. For each word, interviewers assessed whether the cohort member had pronounced the word correctly or not, they could also code if the cohort member did not wish to continue with the test. Interviewers were provided with training on the correct pronunciation of each of the words.

Noted below are the variable names and labels for each of these words.

Variable name	Variable label
b11narta	NART25 - CHORD
b11nartb	NART25 - AISLE
b11nartc	NART25 - DEBT
b11nartd	NART25 - NAIVE
b11narte	NART25 - BOUQUET
b11nartf	NART25 - PLACEBO
b11nartg	NART25 - SUBTLE
b11narth	NART25 - GOUGE
b11narti	NART25 - HIATUS

Variable name	Variable label
b11nartj	NART25 - HEIR
b11nartk	NART25 - EQUIVOCAL
b11nartl	NART25 - RAREFY
b11nartm	NART25 - FACADE
b11nartn	NART25 - ZEALOT
b11narto	NART25 - SUPERFLUOUS
b11nartp	NART25 - CELLIST
b11nartq	NART25 - QUADRUPED
b11nartr	NART25 - LEVIATHAN
b11narts	NART25 - ABSTEMIOUS
b11nartt	NART25 - BEATIFY
b11nartu	NART25 - SIDEREAL
b11nartv	NART25 - GAUCHE
b11nartw	NART25 - DETENTE
b11nartx	NART25 - SYNCOPE
b11narty	NART25 - DEMESNE

A variable was derived (noted below) which only included cohort members who completed this test – providing a response for all 25 words (those who requested not to continue at any point are excluded).

Variable name	Variable label
bd11nart	**(Derived) NART – number of words correctly pronounced**

5.4.2 Sensitive questions

During the main interview, the cohort member was asked to complete a self-completion section which lasted for approximately 10 minutes and covered sensitive questions. This was administered differently in the in-person and in video interviews.

During in-person interviews the interviewer passed their laptop over to the respondent so that they could answer the sensitive questions themselves. The option was available for the interviewer to read out the questions should the cohort member request this.

In video interviews, the sensitive questions were programmed in a web survey. During the interview, the interviewer pasted the cohort member's unique link to the web survey in the chat function on Microsoft Teams. The interviewer stayed on the video call while the cohort member completed the web survey.

If the cohort member could not access or complete the web survey themselves, the interviewer had the self-completion section on their interview programme and could share their screen so that the participant could read the questions and tell the interviewer the number of the response option they wished to choose. This option was only available after the soft launch.

The link to the web survey could also be emailed out following the survey if necessary. The web survey was programmed to be as similar as possible to the self-completion section used during in-person interviews to reduce any mode effects.

A 'don't know' or 'prefer not to say' answer option was not available at first. It was made clear that the cohort member could skip questions they did not want to answer by leaving them blank. On doing so the options of 'don't know' and 'prefer not to say' would then appear on their screen.

All variables asked as part this element have 'SC' at the beginning of the variable label. The responses completed as part of the web survey have been merged in the data with those completed in-person or with the interviewer. Variable '**bd11sc_mode'** provides further information on the mode of completion for this element.

5.4.3 Paper Self Completion questionnaire

Cohort members were asked to complete a paper self-completion questionnaire before their interview. The questionnaire was posted or given to the cohort member by the interviewer when making an appointment to conduct the interview. The questionnaire was either collected by the interviewer at the end of the interview (if an in-person interview) or posted back directly to the fieldwork agency. While it was mainly those taking part by video who posted back the questionnaire, this was an option in some cases where the questionnaire had not been completed in advance of an in-person visit.

The paper self-completion questionnaire included the following sections:

- Leisure activities

- Personality, views and attitudes

- Health

- Physical activities

- News

- Alcohol consumption

- Screen time and reading

- Expectations for the future

All variables in the data which are part of the paper self-completion have the variable labels prefixed with '(PSC)', for paper-self-completion.

Mop-up respondents were not asked to complete a paper-self-completion questionnaire. 32 mop-up cases did complete this and are included in the data. These cases would have received (and completed) the paper questionnaire when they were first invited to take part in the main survey (but did not take part at this stage and instead took part when issued to the 'mop-up').

5.4.4 Online dietary questionnaire

Cohort members were asked to complete an online dietary questionnaire about one randomly allocated weekday and one weekend day from the seven-day period following their interview. They were provided with a leaflet containing a link to the questionnaire and a unique login code. The questionnaire (Oxford WebQ) used was developed and hosted by the Cancer Epidemiology Unit at the University of Oxford.

The Oxford WebQ was developed for repeated implementation in large prospective studies, e.g., the UK Biobank and the Million Women Study. It asks about consumption of about 200 commonly consumed food and beverage items during the previous 24 hours. The quantity of each food or drink consumed during the reference period is calculated by multiplying the assigned portion of each food or beverage by the amount consumed. The nutrient intake is calculated by multiplying the quantity

consumed by the nutrient composition of the food or beverage (using McCance and Widdowson nutrient database). For further information please visit: www.ceu.ox.ac.uk/research/oxford-webq

Data from the OxfordWebQ is not included in this first deposit. The data will be made available later in 2025.

5.4.5 Data linkage consents

BCS70 collects consent from cohort members and their cohabiting partners to link data collected in the study with records held by the National Health Service (NHS), His Majesty's Revenue and Customs (HMRC), and the Department for Work and Pensions (DWP). The information contained in the health records provides details of hospital visits, long-lasting health conditions, treatments received, and medications prescribed. The economic records from DWP and HMRC includes details of benefits being received, national insurance and tax payments, and a full employment history.

Consent to data linkage was first asked about in the Age 42 survey. In the Age 51 Survey cohort members were asked to give consent if they had not taken part in the Age 42 survey or if they had taken part but refused one or more of the consents. If they had consented to link their data to some but not all of their records previously, they were only asked about those records they had refused.

Similarly, the cohort member's partner was asked for consent if the cohort member had not taken part in the Age 42 Survey, they were a new partner or they were the same partner, but they had refused consent to one or more of their records before.

Consent to data linkage was recorded in the interview programme. Unlike the Age 42 survey, there was no paper consent form that the cohort member had to sign. Partners who were not present during the interview were sent a web survey asking them to provide their consents.

All cohort members and partners who were asked for consent were later sent an email or a letter which detailed what consents they had agreed or refused to and the process to follow if they changed their minds.

A full description of the consent process and consent rates obtained is provided in the BCS70 - Age 51 Survey: Technical Report.

For up-to-date information about the availability of linked data for research visit:
www.cls.ucl.ac.uk/data-access-training/linked-data.

5.4.6   Occupation coding

Participants were asked to provide details about their current job, as well as any previous jobs they had since the last interview (or since 1st January 2012 if they had not taken part in the last two sweeps). Those with a cohabiting partner were also asked to provide details about their partner's job.

All occupations were coded to the four-digit standard occupation coding frame (SOC 2020). The SOC2020 codes are deposited in the data. To minimise disclosure risk, 3-digit SOC codes are included in EUL deposit. The 4-digit SOC codes are available under Secure Access (see Section 7.1.2).

5.5   Scales

The BCS70 Age 51 Survey included several established scales which are listed below. Overall scores for each scale have been derived and included within the data deposit (data set name bcs11_age51_main). Further details regarding the derivation of the scores can be found in Appendix 1 – 'Derived Variables'. Original wording used to establish the scales can be found in the 1970 British Cohort Study (Age 51) Questionnaire.

5.5.1   Health module: ONS long lasting health conditions and illnesses: Impairments and Disability (ONS, 2015)

The Age 51 Survey included a sub-set of the ONS harmonised set of questions on Long-lasting Health Conditions and Illnesses including Impairments and Disability. The three items listed below are used to derive variables indicating whether cohort members are disabled using the Equality Act 2010 definition (B11DDISEA) and whether they have a long-standing illness or condition using the European Union's Statistics on Income and Living Conditions (EU-SILC) definition (B11DDISEU) (ONS, 2015). B11DDISEA identifies individuals as disabled or not, B11DDISEU identifies individuals as having no long-standing health condition, having a condition which

hampers daily activities to an extent or having a condition which severely hampers daily activities.

Variable name	Variable label
b11loil	Any physical or mental health conditions or illnesses lasting or expected to last 12 months
b11lolm	Whether illness/conditions reduce ability to carry out day to day activities
b11lolp	Length of time ability to carry out day to day activities has been reduced
bd11diseq	**(Derived) Disability classification Equality act (2010)**
bd11disls	**(Derived) Disability classification EU-SILC**

According to the Equality Act 2010 definition, a cohort member is considered to be disabled if they report a longstanding illness (b11loil) and have a reduced ability to carry out day-to-day activities as a result of their illness (b11lolm).

According to the EU-SILC definition, a cohort member is considered to be disabled if they report a longstanding illness (b11loil), have a reduced ability to carry out day to-day activities as a result of their illness (b11lolm), and this reduced ability has lasted for more than 6 months (b11lolp). This variable also distinguishes between those that are disabled to some extent, and those that are severely hampered (from b11lolm).

5.5.2   Self-completion module: Social provisions

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). These were included in the soft launch and main stages only and were part of the section covering 'sensitive questions'.

The Social Provisions Scale measures the availability of social support. 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: a) I have family and friends who help me feel safe, secure and happy b) There is someone I trust whom I would turn to for advice if I were having problems c) There is no one I feel close to…. described their current relationship with other people from the following responses:

1. Very true

2. Partly true
3. Not true at all

Variable Name	Variable label
b11socprova	(SC) Have family and friends who help feel safe, secure and happy
b11socprovb	(SC) Have someone to turn to for advice
b11socprovc	(SC) No one close to

These questions were asked as part of the sensitive questions section in the soft launch and main stages only (they were not included in the pilots and mop-up).

5.5.3 Self-completion module: GAD2 (Generalised Anxiety Disorder 2-item)

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 was 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 how often they have been bothered by problems over the last 2 weeks: a) "Feeling nervous, anxious or on edge"; and b) "Not being able to stop or control worrying", 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	Variable label
b11gad2phq2a	(SC) Whether nervous, anxious or on edge over the last 2 weeks
b11gad2phq2b	(SC) Whether not able to stop or control worrying in the last 2 weeks
bd11gad2	(Derived) Generalised Anxiety Disorder 2-item

These questions were asked as part of the sensitive questions section in the soft launch and main stage as well as being included in the mop-up. They were not included in the pilots.

5.5.4 Self-completion module: PHQ2 (Patient Health Questionnaire 2-item)

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 how often they have been bothered by problems over the last 2 weeks: c) "Little interest or pleasure in doing things"; and d) "feeling down, depressed or hopeless", 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	Variable label
b11gad2phq2c	(SC) Whether had little interest or pleasure in doing things in the last 2 weeks
b11gad2phq2d	(SC) Whether feeling down, depressed or hopeless in the last 2 weeks
bd11phq2	DV: Patient Health Questionnaire 2-item

These questions were asked as part of the sensitive questions section in the soft launch, main stage and mop-up. They were not included in the pilots.

5.5.5 Self-completion module: UCLA loneliness 3 item

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

Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. Res Aging. 2004;26(6):655-672. doi: 10.1177/0164027504268574. PMID: 18504506; PMCID: PMC2394670.

Three items from the 20-item UCLA loneliness scale were asked of all cohort members in the in-person pilot, main stage and mop-up stages. These questions were not included in the video pilot. 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

A combined score is obtained by adding the score for each question (Total points). The score for each question is:

1 = Hardly ever

2 = Some of the time

3 = Often

In addition, a fourth item (How often do you feel lonely?) was included in the BCS70 paper questionnaire and as part of the mop-up survey. This is not part of the UCLA scale.

Variable name	Variable label
b11lonela	(SC) Feeling lack of companionship
b11lonelb	(SC) Feeling left out
b11lonelc	(SC) Feeling isolated from others
bd11loneliness	(Derived) UCLA loneliness 3 item

5.5.6   Self-completion module – MALAISE Inventory

Rutter, M., Tizard, J., & Whitmore, K. (1970). Education, health, and behaviour. London: Longman.McGee, R., Williams, S., and Silva, P. A. (1986) 'An evaluation of the Malaise Inventory', Journal of Psychosomatic Research, 30(2), pp.147-152.

Earlier sweeps of the study have included a set of 24 self-completion questions which combine to measure levels of psychological distress, or depression (Rutter et al, 1970). As per the Age 46 and Age 42 Survey, the Age 51 Survey used 9 of the original 24 items.

Variable name	Variable label
b11mal02	(SC) Whether feel tired
b11mal03	(SC) Whether feel depressed
b11mal05	(SC) Whether worried
b11mal09	(SC) Whether gets enraged
b11mal12	(SC) Whether gets scared
b11mal14	(SC) Whether gets easily upset
b11mal16	(SC) Nervousness
b11mal20	(SC) Whether annoyed and worn out
b11mal21	(SC) Whether heart races
BD11MAL	**(Derived) Total Malaise score (9 questions)**
BD11MALG	**(Derived) Total Malaise score – grouped**

These questions were asked as part of the sensitive questions section in all stages.

5.5.7 Paper Self Completion: – SF-36

Ware J.E., Snow K.K., Kosinsk,i M., & Gandek, B. (1993), SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: New England Medical Center, The Health Institute.

SF-36 is a widely used multi-purpose health survey comprised of 36 questions. It yields a 9-scale profile of functional health and well-being scores as well as psychometrically-based physical and mental health summary measures and a preference-based health utility index (Ware et al, 1993).

Each of the 9 scales are scored between 0 and 100 with higher scores indicating higher levels of health.

Variable name	Variable label
b11q16a	(PSC) Whether health limits activities – vigorous activities
b11q16b	(PSC) Whether health limits activities – moderate activities
b11q16c	(PSC) Whether health limits activities – carrying groceries
b11q16d	(PSC) Whether health limits activities – climbing several flights of stairs
B11q16e	(PSC) Whether health limits activities – one flight of stairs
b11q16f	(PSC) Whether health limits activities – Bending kneeling or stooping
b11q16g	(PSC) Whether health limits activities – walking more than a mile
b11q16h	(PSC) Whether health limits activities – walking half a mile
b11q16i	(PSC) Whether health limits activities – walking 100 yards
b11q16j	(PSC) Whether health limits activities – bathing or dressing self
BD11PHHE	**(Derived) SF-36 Physical functioning score**
b11q17a	(PSC) Past 4 weeks physical health problems limiting – time spent on other activities
b11q17b	(PSC) Past 4 weeks physical health problems limiting – ability to work on other activities
b11q17c	(PSC) Past 4 weeks physical health problems limiting – accomplished less than you would like
b11q17d	(PSC) Past 4 weeks physical health problems limiting – difficulty performing other activities
BD11RLMP	**(Derived) SF-36 Role-limitations due to physical health**
b11q18a	Past 4 weeks physical health problems limiting - time spent on other activities
b11q18b	Past 4 weeks physical health problems limiting -accomplished less than you would like
b11q18c	Past 4 weeks physical health problems limiting – not done your work or other activities as carefully as usual

Variable name	Variable label
BD11RLME	**(Derived) SF-36 Role-limitations due to emotional problems**
b11q22a	(PSC) Past 4 weeks energy and emotion – full of life
b11q22e	(PSC) Past 4 weeks energy and emotion - a lot of energy
b11q22g	(PSC) Past 4 weeks energy and emotion - worn out
b11q22i	(PSC) Past 4 weeks energy and emotion - tired
BD11ENFA	**(Derived) SF-36 Energy/fatigue score**
b11q22b	(PSC) Past 4 weeks energy and emotion -a very nervous person
b11q22c	(PSC) Past 4 weeks energy and emotion down in the dumps nothing could cheer CM up
b11q22d	(PSC) Past 4 weeks energy and emotion - calm and cheerful
b11q22f	(PSC) Past 4 weeks energy and emotion - downhearted and low
b11q22h	(PSC) Past 4 weeks energy and emotion -a happy person
BD10EMWB	**(Derived) SF-36 Emotional Well-Being score**
b11q19	(PSC) Past 4 weeks physical or emotional problem interfered with social activities
b11q22j	(PSC) Past 4 weeks health limited social activities
BD10SOCF	**(Derived) SF-36 Social Functioning score**
b11q20	(PSC) Past 4 weeks bodily pain
b11q21	(PSC) Past 4 weeks pain interferes with work
BD11PAIN	**(Derived) SF-36Pain score**
b11hlthgn	General state of health
b11q23a	(PSC) General health – ill easier than others
b11q23b	(PSC) General health - healthy as anyone I know
b11q23c	(PSC) General health – expects health to get worse
b11q23d	General health -health is excellent
BD11GENH	**(Derived) General health score**
b11khlstt	General health compared to one year ago
BD11RPHT	**(Derived) Reported health transition**

5.5.8 Paper Self Completion: Personality Traits – Mini International Personality Item Pool (IPIP)

> *Donnellan MB, Oswald FL, Baird BM, Lucas RE. The mini-IPIP scales: tiny-yet effective measures of the Big Five factors of personality. Psychol Assess. 2006 Jun;18(2):192-203. PubMed PMID: 16768595.*
>
> *Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological Assessment, 4, 26-42.*
>
> *Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. Journal of Research in Personality, 51, 78-89.*

A shortened version of the 50-item IPIP representation of the Goldberg (1992) markers for the Big-Five factor structure has been used. This contains 20 items, with four items per Big Five trait (Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness).

Cohort members were asked to use the rating scale (below) to describe how accurately each of 20 statements describes them:

Rating scale:

1. Very inaccurate
2. Moderately inaccurate
3. Neither inaccurate or accurate
4. Moderately accurate
5. Very accurate

The 20 statements included:

1. I am the life of the party.
2. I sympathize with others' feelings.
3. I get chores done right away.
4. I have frequent mood swings.
5. I have a vivid imagination.
6. I don't talk a lot.
7. I am not interested in other people's problems.
8. I often forget to put things back in their proper place.
9. I am relaxed most of the time.
10. I am not interested in abstract ideas.

11. I talk to a lot of different people at parties.

12. I feel others' emotions

13. I like order.

14. I get upset easily.

15. I have difficulty understanding abstract ideas.

16. I keep in the background.

17. I am not really interested in others.

18. I make a mess of things.

19. I seldom feel blue.

20. I do not have a good imagination.

Statements are then given a score of 1-5 (according to the rating scale), with statements 6, 7, 8, 9, 10, 15, 16, 17, 18, 19 and 20 being scored in the reverse order 5-1. Five personality traits are then combined using the mean scores as noted below:

Neuroticism: Mean of items 4, 9r, 14, 19r

Extraversion: Mean of items 1, 6r, 11, 16r

Openness: Mean of items 5, 10r, 15r, 20r

Agreeableness: Mean of items 2, 7r, 12, 17r

Conscientiousness: Mean of items 3, 8r, 13, 18r

Variable Name	Variable label
b11q25d	(PSC) CM self-assessment: I have frequent mood swings
b11q25i	(PSC) CM self-assessment: I am relaxed most of the time
b11q25n	(PSC) CM self-assessment: I get upset easily
b11q25s	(PSC) CM self-assessment: I seldom feel blue
BD11IPIP1	**(Derived) Personality IPIP-FFM: Neuroticism**
B11q25a	(PSC) CM self-assessment: I am the life of the party
B11q25f	(PSC) CM self-assessment: I don't talk a lot
B11q25k	(PSC) CM self-assessment: I talk to a lot of different people at parties
B11q25p	(PSC) CM self-assessment: I keep in the background
BD11IPIP2	**(Derived) Personality IPIP-FFM: Extraversion**
B11q25e	(PSC) CM self-assessment: I have a vivid imagination
B11q25j	(PSC) CM self-assessment: I am not interested in abstract ideas
B11q25o	(PSC) CM self-assessment: I have difficulty understanding abstract ideas

Variable Name	Variable label
B11q25t	(PSC) CM self-assessment: I do not have a good imagination
BD11IPIP3	**(Derived) Personality IPIP-FFM: Openness**
B11q25b	(PSC) CM self-assessment: I sympathise with others' feelings
B11q25g	(PSC) CM self-assessment: I am not interested in other people's problems
B11q25l	(PSC) CM self-assessment: I feel others' emotions
B11q25q	(PSC) CM self-assessment: I am not really interested in others
BD11IPIP4	**(Derived) Personality IPIP-FFM: Agreeableness**
B11q25c	(PSC) CM self-assessment: I get chores done right away
B11q25h	(PSC) CM self-assessment: I often forget to put things back in their proper place
B11q25m	(PSC) CM self-assessment: I like order
B11q25r	(PSC) CM self-assessment: I make a mess of things
BD11IPIP5	**(Derived) Personality IPIP-FFM: Conscientiousness**

5.5.9 Paper Self Completion - AUDIT-PC

F. Babor, T., C. Higgins-Biddle, J., B. Saunders, J., & G. Monteiro, M. (2001). The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care. W. H. Organisation.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1113176/Alcohol-use-disorders-identification-test-for-primary-care-AUDIT-PC_for-print.pdf

The AUDIT-PC consists of 5 questions covering alcohol consumption, problems and dependency. Responses to each question are scored from 0 to 4 giving a maximum score of 20 (BD11AUDIT). Scores of 5 or more are considered AUDIT-PC positive and associated with increasing or higher risk drinking (BD11AUDG). The AUDIT-PC was included in the BCS70 Age 46 Survey, BCS70 Age 42 Survey and the NCDS Age 61 Survey. It is an abbreviated version of the full AUDIT scale included in the Age 50 follow-up of NCDS (Babor et al., 2001).

Variable Name	Variable label
b11q11	(PSC) Drinking frequency
b11q12	(PSC) Drinking amount
b11q13a	(PSC) How often cannot stop drinking
b11q13b	(PSC) Whether drinking affects life responsibilities
b11q14	(PSC) Whether advised against drinking
BD11AUDIT	**(Derived) Total AUDIT-PC Score**
BD11AUDG	**(Derived) AUDIT-PC Group**

5.5.10 Paper Self Completion: Shortened version of Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS)

Michael T. McKay, James R. Andretta, Evidence for the Psychometric Validity, Internal Consistency and Measurement Invariance of Warwick Edinburgh Mental Well-being Scale Scores in Scottish and Irish Adolescents, Psychiatry Research, Volume 255, 2017,Pages 382-386, ISSN 0165-1781, https://doi.org/10.1016/j.psychres.2017.06.071.Hanzlová, R. and Lynn, P., 2023. Item response theory-based psychometric analysis of the Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) among adolescents in the UK. Health and Quality of Life Outcomes, 21(1), p.108.

Ringdal, R., Eilertsen, M.-E., Bjørnsen, H., Espnes, G., & Moksnes, U. (2018). Validation of two versions of the Warwick-Edinburgh Mental Well-Being Scale among Norwegian adolescents. Scandinavian Journal of Public Health, 46, 140349481773539. https://doi.org/10.1177/1403494817735391

Ng Fat, L., Scholes, S., Boniface, S., Mindell J., & Stewart-Brown S. (2017) Evaluating and establishing the national norms for mental well-being using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. Quality of Life Research, 26(5), 1129-1144.

SWEMWBS is a short version of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS). SWEMWBS uses 7 of the WEMWBS's 14 statements about thoughts and feelings, which relate more to functioning than feelings therefore offering a slightly different perspective on mental wellbeing.

The seven positively worded items with five response categories are outlined below.

The SWEMWBS is scored by first summing the scores for each of the seven items, which are scored from 1 to 5. The total raw scores are then transformed into metric scores using the SWEMWBS conversion table which can be found here:
www.warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/howto/swemwbs_raw_score_to_metric_score_conversion_table.pdf

Scores range between 7 to 35 and higher scores indicate higher levels of mental well-being.

Rating scale:

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

Statements:

1. I've been feeling optimistic about the future
2. I've been feeling useful
3. I've been feeling relaxed
4. I've been dealing with problems well
5. I've been thinking clearly
6. I've been feeling close to other people
7. I've been able to make up my own mind about things

Variable Name	Variable label
b11q35a	(PSC) WEMWB: Past 2 weeks how optimistic about the future
b11q35b	(PSC) WEMWB: Past 2 weeks how often CM felt useful
b11q35c	(PSC) WEMWB: Past 2 weeks how often CM felt relaxed
b11q35d	(PSC) WEMWB: Past 2 weeks how often CM been dealing with problems well
b11q35e	(PSC) WEMWB: Past 2 weeks how often CM been thinking clearly
b11q35f	(PSC) WEMWB: Past 2 weeks how often been feeling close to other people
b11q35g	(PSC) WEMWB: Past 2 weeks how often CM been able to make up own mind about things
BD11WEMWB	**(Derived) Warwick Edinburgh Mental Well-Being Scale**

6. Response

6.1 Overall response

Overall, 12,041 cohort members were initially invited to take part in this sweep of BCS70. An additional 474 cases known to be living outside of Great Britian (emigrant cases) were invited to take part in the mop-up web, boosting the total issued sample to 12,515. Emigrant cases were not considered as 'eligible' to take part in the main stage or pilots which were conducted in-person or by video as these cases are not formally considered part of the target population.

A total of 8,025 study members participated in either the pilot, main stage or mop-up web survey (including partial interviews, proxy cases and emigrants).

Nine productive interviews were removed from the data prior to deposit, following the data cleaning and checking process, resulting in a total number of **8,016 deposited interviews available for analysis**. This response chapter is based on the 8,016 interviews deposited. The response figures in the technical report are based on the 8,025 achieved interviews.

Due to the addition of 'emigrant' cases, which were not part of the initial 'target' population, two separate response rates are outlined below:

Response rate A, <u>excludes</u> the 'emigrant' cases from interviews achieved and issued sample, as these cases were deemed as ineligible for the main and pilot stages.

Response rate B <u>includes</u> these 'emigrant' cases as eligible in both achieved interviews and issued sample.

A total of 111 confirmed ineligibles cases have been removed from both response rates (for example those found to have died/were in prison).

- Response rate A (emigrants ineligible) –7,789 productive interviews and a response rate of 65% [5]

- Response rate B (emigrants eligible) – 8,016 productive interviews and a response rate of 65% [6]

Out of the 8,016 productive interviews, 11 were partially productive in-person/video interviews and 49 were partially completed mop-up web surveys (13 of these cases were 'emigrant' cases).

Five cases took part via proxy and one with the aid of an 'interpreter' (see section 5.2 for further information).

6.2   Response by fieldwork stage

6.2.1   Main stage and pilots

A total of 7,179 cohort members were interviewed during mainstage fieldwork between June 2020 and early November 2023, including the soft launch. An additional 160 interviews were achieved during the two pilot studies. The overall response rate after completion of the mainstage and pilots was 62%.

6.2.2   Mop-Up Survey (excluding emigrants)

In the web-based Mop-Up Survey which took place from December 2023 to January 2024, a further 3,773 cohort members who had been invited to the main stage and pilots but had not participated were invited to take part and 450 interviews were achieved, giving a total of 7,789 interviews. The mop-up survey increased the response rate to 65%. Table 3 provides a breakdown of response rate A.

Table 3. Overview of response rates from pilots, mainstage survey and mop-up (excluding emigrants)

---

[5] Response = 100*(7789 productive interviews/(12041 original cases issued minus111 confirmed ineligibles))

[6] Response=100*(8016 productive interviews including emigrants/(12041 original cases issued plus 474 emigrant cases issued minus 111 confirmed ineligibles))

	N	%
Productive	7,789	65
by video interview	*3,498*	*45*
by in-person interview	*3,812*	*49*
..by telephone	*29*	*0*
by web in mop-up	*450*	*6*
Non-contact	656	5
Refusal	2620	22
Other unproductive	391	3
Unknown eligibility (no contact)	474	4
Ineligible	111	1
Total	*12,041*	*100*
Response rate[7]		**65**

6.2.3 Mop-Up Survey (including emigrants)

A total of 474 study members living outside of Great Britian were invited to take part in the mop-up survey and 227 participated (response 48%). Table 4 below provides further information on the response to the mop-up survey.

Table 4. Mop-up Survey response

		Productive completes		Productive partials		All productives	
	Sample size	*N*	*%*	*N*	*%*	*N*	*%*
Non-responders to main survey	3,773	414	11	36	1	450	12
Emigrants	474	214	45	13	3	227	48
Total	4,247	628	15	49	1	677	16
Base: All cases issued to mop-up survey (4,247)							

The tables 5 provides an overview of each stage and the response achieved.

Table 5. Survey response at each stage of fieldwork

---

[7] The response rate is the percentage of productive interviews from the issued sample, excluding those confirmed ineligible cohort members. The issued sample includes 29 cases which were re-classified as not being eligible cohort members after fieldwork.

		Mainstage (including soft launch)	Mainstage with pilots	Mainstage/ pilots mop-up cases (excluding emigrants)	Mainstage/ pilots mop-up cases (including emigrants)
Number of completed interviews	N	7,179	7,339	7,789	8,016
Response rate	%	61	62	65	65

Base: All known eligible cases issued to mainstage and mop-up survey (base 11,930 – excludes 'emigrant' cases and ineligibles, base 12,404 includes 'emigrant' cases, excludes ineligibles)

6.3   Mode of Completion

Of the 8,016 total number of interviews achieved in the pilot, video pilot, mainstage and mop-up survey (including emigrants), 48% were completed in-person, 44% were completed by video, 8% were completed by web and 29 cases, 0.4% were completed by telephone.

Conducting the interview by telephone was not an official option in this survey, however a small number of interviews were conducted by telephone rather than video call (n=29). This took place mainly due to technical difficulties experienced during video interviews.

Survey mode is denoted by '**b11survey_mode'** variable in the survey data set.

Table 6. Mode of response

Mode	Frequency	Percent
In-person interviews	3812	47.6%
Video call interviews	3498	43.6%
Telephone interviews	29	0.4%
Web interviews	677	8.4%
Total	8016	100%

6.4   Response by country of issue

Survey response (including the mop-up survey but excluding emigrants) varied by country. Scotland had the highest response rate (69%), then England (65%) and Wales had the lowest response rate (60%). See table 7.

Table 7. Response by country (including pilot, mainstage and Mop-up but excluding emigrant cases)

	ENGLAND		SCOTLAND		WALES		JERSEY/ GUERNSEY/ ISLE OF MAN		TOTAL	
	N	%	N	%	N	%	N	%	N	%
Total issued	10,340	100	966	100	699	100	30	100	12,035	100
Productive	6,686	65	664	69	416	60	22	73	7,7788	65
Non-contact	571	6	44	5	37	5	4	13	656	5
Refusal	2,255	22	177	18	180	26	3	10	2,615	22
Other unproductive	337	3	24	2	29	4	0	0	390	3
Unknown Eligibility (no-contact)	396	4	44	5	33	5	1	3	474	
Ineligible	87	1	12	1	4	1	0	0	103	1
Response rate	65%		70%		60%		73%		65%	
Base: All productive interviews (excluding emigrants, those without and address and data deletion cases)12035										

6.5   Response rates for each element

Further detailed information on the response to each element is available in the BCS70 - Age 51 Survey: Technical Report in section 9.8.

6.5.1   Paper self-completion

A total of 5,972 paper self-completion questionnaires are included in the data deposited. Those taking part in the mop-up stage were not asked to complete a paper-self completion questionnaire, but 32 cases did receive and complete this. These cases were initially invited to take part in an earlier stage (and were provided with a paper questionnaire then) but did not take part until the mop up. These cases are included in the data deposited but are excluded from table 8 along with proxy interviews who were not asked to complete a paper self-completion questionnaire.

Table 8. Completion of paper self-completion questionnaire by interview mode

	In-person		*Video*		Telephone		*Total*	
	N	%	N	%	N	%	N	%
Completed PSC	2918	76.7	3008	86.0	14	48.3	5940	81.0
Not Completed PSC	889	23.4	490	14.0	15	51.7	1394	19.0

TOTAL	3807	100	3498	100	29	100	7334	100

Base: all 7,344 fully productive and partially productive cases from mainstage, pilot and video pilot (proxy and mop-up cases excluded from this table)

6.5.2 Completion of sensitive questions

In total, 93% respondents completed the section with sensitive questions. Rate of completion varied by mode, with 97% of those interviewed in-person completing the self-completion module compared to 89% interviewed by video.

Table 9 provides further detail on the completion of this section.

Self completion mode is denoted by variable: '**bd11sc_mode'**.

Table 9: Completion of sensitive questions

	In-person		Video		Telephone		Total	
	N	%	N	N			N	%
Web completed during interview	-	-	2,479	70.9		-	2,479	33.8
Web completed after interview	1[8]	0	401	11.5	-	-	402	5.5
Completed with interviewer during video interview	-	-	232	6.6	-	-	232	3.2
Self completion during in-person/telephone interview	3,585	94.0	-	-	-	-	3,585	48.8
With interviewer during in-person interview	98	2.6			5	17.2	103	1.4
Not completed	128	3.4	386	11.0	24	82.8	538	7.3

Base: All cohort members who were asked to complete the section with sensitive questions, 7339

6.5.3 Cognitive assessments completion

The cognitive assessments include immediate word-list recall, animal naming, letter cancellation, delayed word recall and the National Adult Reading Test (NART), which were completed during an in-person or video interview.

The agreement rate for all five of the cognitive assessments was very high. Ninety nine percent or more of those completing the interview completed three of the tasks. Ninety-six per cent of cohort members completed the National Adult Reading Test. Rates of completion were similar across both in person and video call modes. The letter cancellation test had a lower response of 88%, with a lower completion rate in the video mode compared to the in-person mode (77% vs. 87% respectively).

The letter cancellation task had to be posted to the cohort member before the video interview which may explain the lower agreement rates in this mode – some participants may not have received the task or were unable to locate it when required.

---

[8] Completion of the sensitive questions via web was not an option for those taking part in person but in one case this was specifically requested and arranged.

Table 10: Completion of cognitive tests

	In-person		Video		Telephone		Total	
	N	%	N	%	N	%	N	%
Immediate word recall test								
Completed	3,758	98.6	3,476	99.4	27	93.1	7,261	98.9
Not completed	54	1.4	22	0.6	2	6.9	78	1.1
Animal naming								
Completed	3,752	98.4	3,477	99.4	28	96.6	7,257	98.9
Not completed	60	1.6	21	0.6	1	3.4	82	1.1
Letter cancellation								
Completed	3,322	87.1	2693	77.0	1	3.4	6,016	87.8
Not completed	490	12.9	805	23.0	28	96.7	1,323	12.2
Delayed word recall test								
Completed	3,758	98.6	3,476	99.4	27	93.1	7,261	98.9
Not completed	54	1.4	22	0.6	2	6.9	78	1.1
National Adult Reading Test (NART)								
Completed	3,640	95.5	3,399	97.2	2	6.9	7,041	95.9
Not completed	172	4.5	99	2.8	27	93.1	298	4.1

Base: all 7,339 productive cases. Proxy respondents were not asked this section of the questionnaire

7.   Survey Research Data

7.1   Licensing and data access

The BCS70 Age 51 survey research have been processed by CLS and supplied to the UK Data Service (UKDS).

All users of the data need to be registered with the UKDS and to sign the UKDS End User Licence. Details of how to do this are available at www.ukdataservice.ac.uk/get-data/how-to-access/registration.

Please refer to section 7.8 for information on how these data have been de-identified for sharing.

7.1.1 Safeguarded data (EUL)

The majority of the BCS age 51 survey data are available from the UK Data Service as safeguarded data, which can be downloaded once the End User Licence (EUL) has been signed by the user.

The safeguarded data exclude detailed information that presents a potential risk for disclosivity or is of sensitive nature, which is instead shared as controlled data.

7.1.2 Controlled data (Secure Access)

Some BCS70 survey must be accessed as controlled data from the UK Data Service SecureLab due to their potentially disclosive and/or sensitive nature. This applies to:

1. Uncommon health conditions (including age at diagnosis)
2. Full employment codes and income/finance details
3. Specific life circumstances (e.g. pregnancy details, year/age of emigration from GB)

Applicants wishing to access this data need to abide by the terms and conditions of the UK Data Service Secure Access licence. Before gaining access, researchers must make an application detailing the intended analysis and provide a justification as to why this data is requested. Application guidance can be found at

Data access will be granted once the form has been reviewed by UK Data Service and approved by the CLS Data Access Committee.

7.2 Datasets

The Age 51 survey research data consists of 16 datasets. Datasets are organised by format (wide or long) and topic.

Dietary diary data, additional derived variables and full derived geographical identifiers will be made available separately.

The majority of wide format (flat) data are included in the main survey file (bcs11_age51_main) i.e. where one record exists for each cohort member. Due to their specialised nature and structure, all unfolding bracket responses are contained in a separate dataset.

All long format (hierarchical) datasets contain multiple records for each cohort member, and are identifiable by the 'longf' suffix. These datasets consist of responses to questions where the respondent is asked a set of questions which are repeated until no more information is required. Each long format dataset contains a 'record number' variable, which is unique within each case. The main survey dataset will contain responses to questions that initiate entry of the loop and, if it suits the data, the current status of the respondent.

Table 11. List of safeguarded datasets (End User Licence)

Dataset name	Content summary
bcs11_age51_main	Main survey datafile including the majority of the core interview (both main stage and mop-up), paper self-completion, and derived variables
bcs11_age51_relationships_longf	Relationship histories looped data since last sweep / 2008, one row per cohabitation period with a partner
bcs11_age51_persongrid_longf	Details of persons living with respondent (past and present) and absent children, one row per person mentioned across all household member loops
bcs11_age51_housing_longf	Housing histories since last sweep / 2012, one row per address
bcs11_age51_employment_longf	Economic activity histories since last sweep / 2012, one row per change in economic activity circumstance

Dataset name	Content summary
	(excluding cases whose employment is unchanged since last survey)
bcs11_age51_qualifications_longf	Details of each qualification recorded by respondent since last sweep / 2012, one row per qualification
bcs11_age51_qualificationsp_longf	Details of each qualification recorded by respondent's partner, one row per qualification
bcs11_age51_regincome_longf	Sources of regular household income (excluding job income), one row per income type
bcs11_age51_savings_longf	Sources of savings and investments, one row per saving type/source
bcs11_age51_econshock_longf	Details of economic shocks since covid-19 outbreak (March 2020), one row per type of economic shock
bcs11_age51_econshockp_longf	Details of economic shocks experienced by partner since covid-19 outbreak (March 2020), one row per type of economic shock
bcs11_age51_covidvax_longf	Details of vaccinations for covid-19 received by the respondent, one row per vaccine recorded
bcs11_age51_covidbens_longf	Details of covid-19 state benefits received, one row per benefit type
bcs11_age51_benefits_longf	Details of state benefits received, one row per benefit type
bcs11_age51_debt_longf	Details of debts owed, one row per debt type
bcs11_age51_unfolding	Data relating to a series of questions which aim to get an approximate value for income or payments where participants are unable or unwilling to answer precisely

Table 12. List of controlled datasets (Secure Access)

Dataset name	Content summary
bcs11_age51_main_restricted	Secure access data in flat/wide format including derived variables
bcs11_age51_housing_longf_restricted	Secure access data from housing histories since last sweep / 2012, one row per address
bcs11_age51_employment_longf_restricted	Secure access data from economic activity histories since last sweep / 2012, one row per change in economic activity circumstance (excluding cases whose employment is unchanged since last survey)
bcs11_age51_regincome_longf_restricted	Secure access data from sources of regular household income (excluding job income), one row per income type
bcs11_age51_savings_longf_restricted	Secure access data from sources of savings and investments, one row per saving type/source
bcs11_age51_benefits_longf_restricted	Secure access data from details of state benefits received, one row per benefit type

Dataset name	Content summary
bcs11_age51_debt_longf_restricted	Secure access data from details of debts owed, one row per debt type
bcs11_age51_pregloop_longf_restricted	Details of pregnancy outcomes reported, one row per pregnancy

7.3   Data documentation

In addition to this User Guide, the following documentation accompanies the data deposited at the UKDS:

Table 13. BCS70 survey documents

Name of the document	Content summary
bcs70_sweep11_age51_variable_lookup_table.xlsx	Lookup between deposit variable name and questionnaire CAPI name/question number
bcs70_sweep11_age51_questionnaire.pdf	This document provides the questions asked in the Age 51 Survey. It includes details on routing as well as both mode and stage specific adjustments. Questions asked as part of the 'mop-up' stage are flagged.
bcs70_sweep11_age51_paper_questionnaire.pdf	This document is a copy of the self-completion questionnaire which was provided to study members to complete as part of their participation in the Age 51 Survey.
bcs70_sweep11_age51_technical_report.pdf	This document has been produced by the leading fieldwork agency – NatCen. It provides all technical details regarding the design and implementation of the survey. Some figures presented in the Technical Report may vary with the figures presented in this User Guide (and with the data deposited). Please see P6 for further information.

7.4   Identifiers

All datasets are pseudonymised by identifying them with the same research identifier (BCSID), used for all BCS70 cohort data available at the UKDS.

7.5 Variable description

7.5.1 Variable order

The order in which variables appear in the datasets broadly follows the order of sections, and of questions within sections, of the survey instruments. However, due to the repeating loop nature of some sections of the survey instruments the order is determined by the structure of the CAI program, which does not necessarily hold each question in the order in which they are put to the respondent. As a result, several variables have been re-ordered so similar variables are together.

The survey included several stages with questions occasionally asked in a different order, or in one stage and not another. Variable order primarily follows the main stage questionnaire structure, and mop-up survey only variables moved after the rest of the main survey questions rather than their original order.

7.5.2 Variable names

The variable names are all prefixed by 'b11', denoting the wave/sweep of the cohort study. For ease of tracking variables longitudinally, other than the prefix, variable names are consistent with those used the prior sweep (where the prefix was 'b10').

The variable names are based on those used in the CAI program and are documented in the questionnaire and self-completion questionnaire documentation, but do not match exactly in all cases.

To facilitate matching between dataset variables and CAPI questions, a variable lookup excel spreadsheet has been included in the deposit (bcs11_age51_variable_lookup_table.xlsx).

Variables from the paper self-completion questionnaire have names derived from the question numbers as they appear on the printed questionnaire (e.g. Q1A = 'b11q1a'). The mop-up has several survey specific variables which are identified by an 'mu' prefix following 'b11'

For multi-coded variables, where a single question produces more than one response, a suffix has been used to identify the iteration. 01, 02, 03…..been used to denote the 1st, 2nd, 3rd, ...iterations respectively. Any new codes which were added following the interview, during the coding process, have been allocated a suffix

starting at 50 e.g. ….b11cfwt50, b11cft51 are codes which were added following a review of what the cohort members detailed in their 'other specify' response.

Derived variables in the dataset 'bcs11_derived.sav' are given the prefix "bd11".

7.5.3 Variable labels

The variable labels included in the dataset are based on the question wording that can be found in the core interview and self-completion questionnaire documentation. Where necessary, labels have been modified in an effort to ensure they are comprehensible and accurate. Certain mode/stage information is added to the start of labels for ease of identification. These are as follows:

- '(SC)'- Variables in the self-completion section of the main survey

- '(PSC)'- Variables from the paper self-completion

- '(Mop-up)'- Variables only in the mop-up survey

- '(Pilot)', '(Video Pilot)', '(Soft-Launch)'- Variables only asked at specific described stages

7.5.4 Value labels

The value labels for valid responses are based on the question responses used in the CAI program as documented in the questionnaire documentation. Value labels have been individually reviewed and amended, where necessary.

7.6　Income and payment unfolding brackets

A feature of income or payment questions is the use of unfolding brackets for those cases where a respondent refuses or is unable to provide an exact answer. The unfolding brackets questions are designed to elicit a minimum and maximum value that define a range or "closed band" within which the actual value lies. On entering the unfolding brackets, respondents are asked to say whether they have more, less or about the same as a particular value. This question is repeated using different values (which will be a lower or higher value depending on the answer to the preceding question). The procedure stops at the point when either: an upper and lower bound is provided; the respondent refuses or says "don't know"; or the respondent places themselves in the top or bottom bracket. The unfolding bracket questions are randomly ordered for each respondent. This will average any possible 'anchoring' effects (i.e. where people use the suggested figure as a reference point and adjust it to reach their answer) from the procedure across the distribution. The bracket values are selected based on the density of the underlying financial variable.

7.7　Missing values

Missing values are consistently labelled as follows (unless otherwise stated):

-9 = Refusal

-8 = Don't Know

-3 = Not asked at case fieldwork stage

-2 = Not asked due to scripting/routing error

-1 = Item not applicable

In the Age 51 survey –3 has been reserved for questions not asked at a particular stage e.g. not asked in the pilots, soft launch or mop-up and -2 has been used to flag questions unanswered as a result of a routing error. No routing errors were found in the survey but -2 coding has been used in a small number of cases where a response was missing, caused by a respondent giving a particular answer at one question and going back and changing their answer, resulting in a subsequent question being missed.

The value –1 is used for missing responses to questions which study members would not have been asked if they only partially completed the survey (or they were not routed to this question). For derived variables -8 is typically reserved for 'Not codeable' values, where there is insufficient data for the variable to be derived.

7.8   Data de-identification

In addition to the pseudo-anonymisation (i.e. use of the identifier BCSID), all variables that contained potentially identifiable information provided by the respondents have been removed from the research dataset. This includes job titles, job descriptions, 'Other (please specify)' inputs, town name, postcodes and open-ended questions. These potentially identifiable CLS data can be accessed securely by applying directly to the CLS Data Access Committee: https://cls.ucl.ac.uk/data-access-training/data-access/accessing-data-directly-from-cls/.

For certain potentially disclosive multi-coded data, including health conditions, low-count responses have been combined into a new variable labelled 'Other answer(s) from code frame (Derived)'. This variable combines all coded and back-coded categories not deposited as safeguarded data under EUL, while the full breakdown is available as controlled data under Secure Access.

Occupation coding provides details of CM and partner occupations, and derived variables have been created for geodata.

7.9   Data cleaning of back-coded variables ('other')

Where possible, 'Other' variables have been back-coded to provide categorical data from these open-text responses. Questions that include 'Other (please specify)' categories allow the respondent to give open text responses that are back coded after the interview is completed. Some of these variables are used in filtering cases to subsequent questions. Where backcoding has occurred after the interview, the value will not be used for filtering.

7.10 Weights variables

The variables containing the calculated weights are as follows:

Variable name	Dataset	Variable Description
bd11weight_main	bcs11_age51_main	(Derived) Non-response weight for main survey
bd11weight_psc	bcs11_age51_main	(Derived) Non-response weight for paper self-completion
bd11weight_odq	bcs11_age51_main	(Derived) Non-response weight for online dietary questionnaire

7.11 Output Disclosure Control

The two UK Data Service Secure Lab rules of thumb that will be applied to all research outputs (summary tables, graphs, etc) are:

- Threshold rule: No cells should contain less than 10 observations
- Dominance rule: No observation should dominate the data to a huge extent

The controlled data included in Table 12 is only available via the UKDS Secure Lab. The UK Data Service will always perform a certain level of disclosure control on the outputs generated by researchers, as outlined in their SDC Handbook which can be downloaded from: www.securedatagroup.org/sdc-handbook/

8. Derivation and Implementation of Non-Response Weights

8.1 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 (www.cls.ucl.ac.uk/data-access-training/handling-missing-data/). The approaches we recommend to researchers capitalise on the rich data cohort members provided over the years before their nonresponse. These approaches 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 Age 51 BCS70 survey, non-response weights are provided with the data, so that IPW analysis can be undertaken, either in isolation or in combination with MI or FIML. This section of the User Guide describes the derivation and implementation of these non-response weights. The weights were created and documented by Liam Wright and Richard Silverwood and closely followed the procedures used to derive non-response weights for CLS' COVID-19 Surveys (www.cls.ucl.ac.uk/wp-content/uploads/2017/02/UCL-Cohorts-COVID-19-Survey-user-guide.pdf).

8.2 Target population and response

For the purposes of weighting, we have defined the target population as individuals born in 1970 in Great Britain and who, at the start of fieldwork, were alive and residing in the UK (n = 15,780). We constructed three weights – one each for response to the main interview (CAPI, CAVI or web mop-up), paper self-completion questionnaire, and diet questionnaire.

The Age 51 main interview was issued to a relatively small number of cohort members who had emigrated from the UK. These individuals answered the survey via web 'mop-up' questionnaire. To make the most use of respondents' data, we created survey weights for all respondents, emigrant or otherwise, but for the purpose of constructing the weight, emigrant non-respondents were not considered part of the target sample. Respondents to the paper self-completion and diet questionnaires were subsamples of those who responded to the main interview. Emigrant responders and web-mop-up participants were not invited to complete the paper self-completion questionnaire or the diet questionnaire[9]. Consequently, we did not include emigrants in the target population for the self-completion and diet questionnaire weights. Response rates, based on these target samples were 50.8% (8,016 of 15,780) for the main interview, 38.1% (5,938 of 15,585) for the paper self-completion and 32.4% (5,055 of 15,585) for the diet questionnaire, respectively.[10]

8.3   Derivation of non-response weights

The derivation of each Age 51 Survey non-response weight proceeded as follows:

1. Within the sample corresponding to the target population (broadly, those alive and living in Britain, for each questionnaire type, model Age 51 survey response conditional on a set of covariates using logistic regression. The selection of covariates was informed from results of the CLS Missing Data Strategy (www.cls.ucl.ac.uk/data-access-training/handling-missing-data).

---

[9] A small number of web mop-up respondents completed the paper self-completion questionnaire. We do not create a paper self-completion response weight for these individuals.

[10] The weights were constructed prior to a data deletion request from one cohort member. We removed the individual from the deposited weight data but did not rederive the weights excluding data from this cohort member. In addition, updated response information on a further two cohort members was delivered after the derivation of the weights. This explains the discrepancy in sample size for the paper self-completion questionnaire in the current section (n = 5,938) and other sections (n = 5,940) of this User Guide. Weights will be rederived in a future data release, but given only three cohort members are implicated, we anticipate this will have only a very small impact on analyses using the non-response weights.

2. Among respondents to the relevant Age 51 survey, calculate the probability of response implied by each model.

3. Calculate each Age 51 survey non-response weight as the inverse of the probability of response.

4. Examine the distribution of derived non-response weights to decide whether truncation may be desirable, applying truncation, if so.

5. Finally, calibrate the Age 51 survey non-response weights so that they sum to the number of respondents for each questionnaire type (n = 8,016, 5,938, and 5,055 for main, self-complete and diet questionnaires, respectively).

The variables included in the response model in stage 1 are listed in Table 14. We used the same set of variables as used to derive weights for the BCS70 in CLS's COVID-19 Surveys, with the exception that we also included a variable for number of COVID-19 sweeps a participant responded to (continuous; 0-3). Further, when creating a weight for responses to the Age 51 diet questionnaire, we additionally included a variable for completing the Age 46 diet questionnaire, given its direct relevance. The code used to clean the data and create the weights can be viewed at www.osf.io/wfkm5/.

Table 14: Variables used as predictors in models used to create non-response weights

Variable	Description
Sex	Categories: Male; Female.
Highest Education	Categories: None; Level 1; Level 2; Level 3; Level 4; Level 5.
Number of Rooms @ Age 0y	Mean: 4.6, Range: 1 - 88
Social Class @ Age 10y	Categories: Professional or Managerial; Intermediate; Partly-Skilled or Unskilled.
Cognitive Ability @ Age 10y	Mean: -0.1, Range: -3.7 - 2.9
Malaise @ Age 16y	Mean: 9.4, Range: 0 - 44
Vote @ Age 42y	Categories: Didn't vote; Voted.
Number of Organisations @ Age 42y	Mean: 0.8, Range: 0 - 15
Activity @ Age 46y	Categories: Employed; Not Employed.
Household Income @ Age 46y	Categories: Quintile 1; Quintile 2; Quintile 3; Quintile 4; Quintile 5.
Internet Use @ Age 46y	Categories: None/Little; Medium; Lots.
Marital Status @ Age 46y	Categories: Never Married; Married; Separated / Widowed.

Variable	Description
Social Visits Freq. @ Age 46y	Categories: Very Frequently; Fairly Frequently; Never / Rarerly.
Social Support @ Age 46y	Categories: A Little/ Not at All; Somewhat; A Great Deal.
Malaise @ Age 46y	Mean: 2.1, Range: 0 - 9
BMI @ Age 46y	Mean: 28.6, Range: 12.7 - 80.2
Smoking Status @ Age 46y	Categories: Never Smoked; Ex-Smoker; Current Smoker.
Self-rated Health @ Age 46y	Categories: Excellent / Very Good; Good; Fair / Poor.
Consent to Biological Samples @ Age 46y	Categories: Consented; Did Not Consent.
# Main Sweeps Non-Response	Mean: 3.4, Range: 0 - 10
# COVID-19 Sweeps Non-Response	Mean: 2.1, Range: 0 - 3

Missing values in the above variables were handled using multiple imputation (MI). The imputation models included these variables plus response for the relevant section of the Age 51 Survey. Separate imputation models were conducted for each survey weight (twenty imputed datasets created using chained equations). Twenty was deemed sufficient as only a point estimate (probability of Age 51 survey response) was to be estimated from the MI analysis; more imputations may be required for inference.

Models for Age 51 survey response were fitted in each imputed dataset. Pooled models are reported in Appendix 3. From these models, the probability of Age 51 survey response was predicted for each respondent, with the non-response weight calculated as the inverse of the response probability. These weights were then averaged to get a single weight per individual.

Test analyses were conducted in each cohort at different levels of weight truncation (max = 10, 20, and 50). This suggested that truncation at a maximum of 20 could provide some improvement in precision without undue introduction of bias. The non-response weight was therefore truncated to 20 and was then calibrated so that it summed to the number of Age 51 survey respondents in each cohort by multiplying them by the ratio of the number of responses to the total of the uncalibrated nonresponse weights. The distributions of the resultant calibrated non-response weights are presented in Table 15 along with the distribution prior to truncation.

Table 15: Distributions of the non-response weight (prior and post truncation and calibration)

	Main Interview		Self-Completion		Diet Questionnaire	
	Prior	Post	Prior	Post	Prior	Post
Mean	1.79	1.00	2.27	1.00	2.59	1.00
SD	2.34	1.15	3.03	1.13	3.62	1.22
0%	1.00	0.57	1.04	0.47	1.06	0.42
5%	1.01	0.57	1.08	0.49	1.11	0.44
25%	1.03	0.58	1.15	0.52	1.19	0.47
50%	1.12	0.63	1.37	0.62	1.42	0.57
75%	1.56	0.89	2.12	0.96	2.24	0.89
95%	4.52	2.55	5.93	2.68	8.02	3.19
100%	52.47	11.32	71.76	9.03	51.49	7.96

8.4   Weight 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 4). We considered the sample means of two variables – birthweight and vocabulary at age 5y (English Picture Vocabulary Test) – in the full sample and in the (weighted and unweighted) sample of Age 51 respondents. These variables were collected early in cohort members lives so were observed for a very high proportion of participants. Further, the variables were 'unseen' in the sense that they did not appear as predictors in the response models, so are not balanced mechanistically. A good performing weight is indicated by having a similar sample mean to the initial sample. The results of this exercise are presented in Figure 2. We observed that there is considerable bias when using the unweighted Age 51 survey data but this is markedly reduced – though not completely (especially for the diet questionnaire) – when applying the non-response weights.

Figure 2: **Means for two variables that were not used to create non-response weights in the target population and in weighted and unweighted Age 51 responding sample. Variables standardised (mean = 0, SD = 1) using the target population mean and SD**
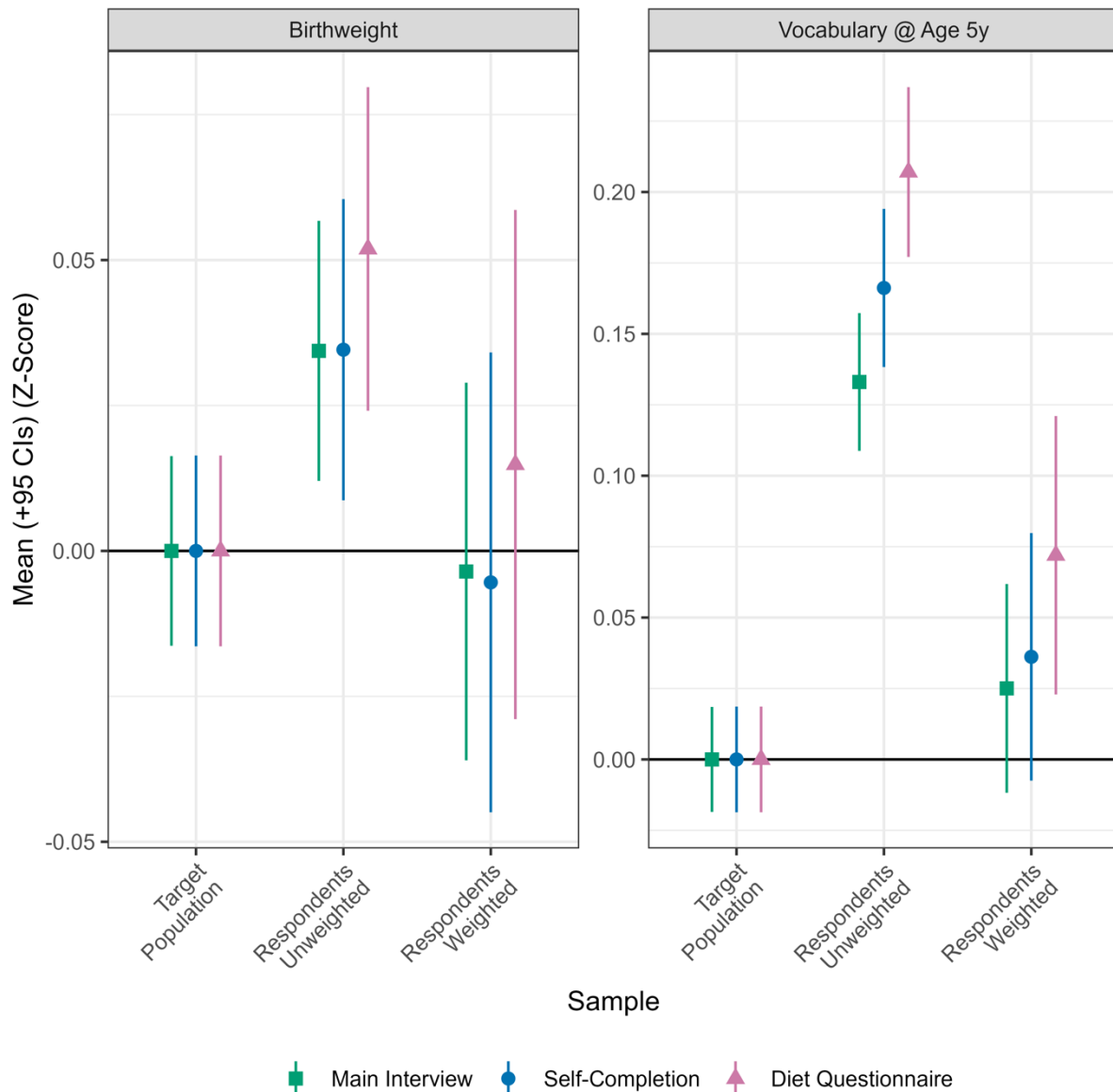
Appendix 4 shows the results of a similar exercise, this time looking at the ability of the weights to recover the distribution (means for continuous variables, proportions for categorical variables) of variables that appeared in the response predictor models (e.g., sex). Weighting reduces bias in almost all cases, and in many cases recovering sample means and proportions close to the initial target population. Bias remains in several variables however, including education and self-rated health, indicating that either observed information in BCS70 does not fully account for

attrition or that to fully restore representativeness, it may be necessary to adjust the weighting of these variables or apply a targeted multiple imputation approach, both of which would incorporate additional information from BCS70.

8.5   Implementation of non-response weights

The choice of non-response weight to use depends on the analysis being conducted. As the paper self-completion questionnaire and diet questionnaire were completed by subsamples of the main interview respondents, users of BCS70 Age 51 sweep data who want to use variables from either questionnaire should consider using the relevant non-response weight. Note, however, that different non-response patterns may warrant the creation of extra weights *de novo*. For instance, data users who want to use data from both the self-complete and diet questionnaires should consider creating their own weights for this specific sample, or apply analysis specific MI or FIML. Further, data users wanting to use variables from other sweeps should consider creating weights to reflect the particular response pattern their analysis implies – non-response patterns in the BCS70 are not uniformly monotonic. For instance, a large number of participants did not participate at the Age 16 sweep due to a teacher strike but participated at later sweeps. In this instance a combination of IPW and MI, or just MI may be more suitable.

The Age 51 mop-up questionnaire was shorter than the main interview. Users of BCS70 Age 51 Sweep data may find that their particular combination of variables exclude participants in the mop-up interview. We tested creating separate non-response weights for 'main interview' and 'main interview or mop-up questionnaire' participants. However, these were very highly correlated (r > 0.98) so we did not include the former with the data release.

9. Mode effects

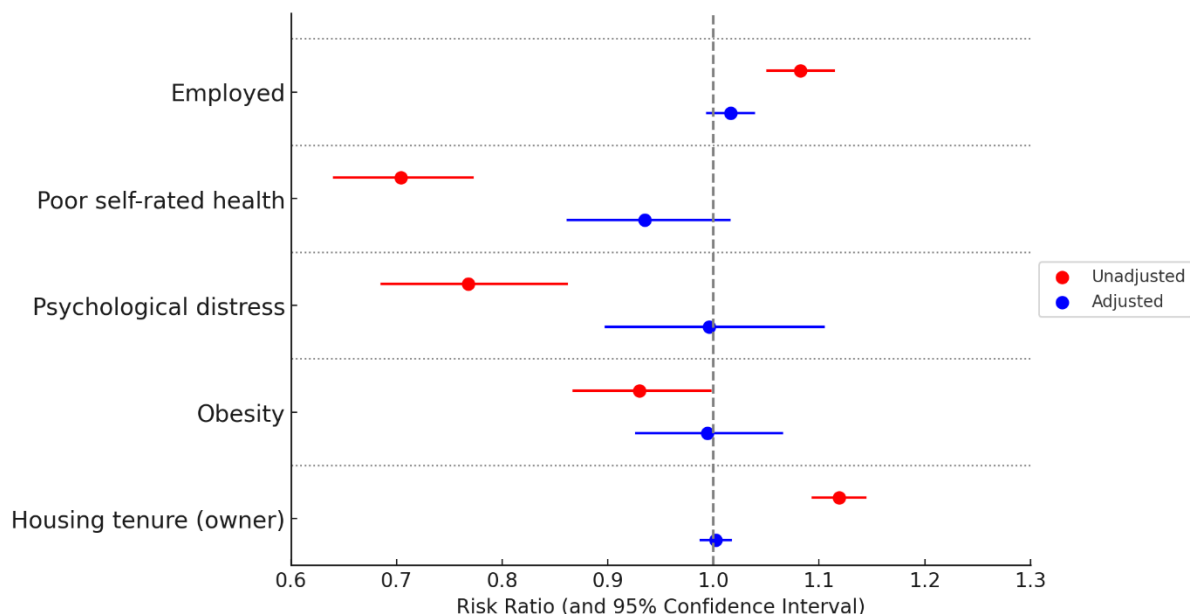The BCS70 Age 51 Survey data collection was carried out with a mixed mode design, where the vast majority of interviews were either competed in person (48%) or using video (44%). Differences in responses between survey modes can arise due to mode effects (i.e., measurement error caused by differences in how questions are presented) or mode selection (i.e., differences in the types of respondents who choose each mode). As outlined in Section 4 of the Handling Mode Effects in the CLS Studies User Guide, accounting for survey mode is only necessary when response differences are driven by mode effects (measurement error), rather than selection effects alone. Section 4.7 of the Handling Mode Effects User Guide presents an example where mode differences are driven by selection only. Since bias arising from missing data related to mode selection mechanisms will be addressed using missing data handling approaches, this section focuses on the issue of non-random allocation to survey mode. This occurs when respondents self-select into a particular mode based on specific characteristics, potentially influencing differences between in-person and video interviews. In Table 16 and Figure 3, we present results from models in which we regressed six key outcomes observed at age 51 on the mode indicator. This allows us to first examine whether mode differences exist and then, leveraging the extensive data available in BCS70, adjust for these differences using variables collected across the life course that are expected to influence mode selection to examine whether such mode differences may be due to mode effects or mode selection. With the exception of psychological distress that was part of the self-completion questionnaire, where participants either responded on the interviewer's laptop during in person interviews or via web during video interviews, all other outcomes were assessed by the interviewer either in person or via video.

Table 16: Regression coefficient, risk ratios and 95% confidence Intervals of the association between survey mode (in person vs video) and six key outcomes.

	Unadjusted Beta[2]/Risk Ratio	95% Confidence Interval	Adjusted[1] Beta/Risk Ratio	95% Confidence Interval
Income	0.106	0.070 to 0.142	-0.002	-0.029 to 0.025
Employed	1.082	1.051 to 1.114	1.016	0.994 to 1.038
Poor self-rated health	0.704	0.641 to 0.772	0.935	0.862 to 1.015
Psychological distress	0.768	0.686 to 0.861	0.993	0.882 to 1.096
Obesity	0.930	0.868 to 0.997	0.994	0.927 to 1.065
Housing tenure (owner)	1.119	1.094 to 1.144	1.002	0.988 to 1.016

1. Adjusted Risk Ratios from models including parental sex at birth, parental social class at birth number of rooms at home/persons per room at birth, cognitive ability at age 10, early life mental health at age 16, voting at age 42, membership in organisations at age 42, internet access at age 46, consent for biomarkers at age 46, educational qualifications at age 42, economic activity at age 46, partnership status at age 46; psychological distress at age 46, Body Mass Index at age 46, self rated health at age 46; smoking status at age 46, social support at age 46, income at age 46, number of non-responses across all previous sweeps, response to COVID-19 wave 1 web survey, response to COVID-19 wave 2 web survey, financial circumstances during COVID-19, relationship satisfaction at age 46; self-efficacy at age 46, whether had savings at age 46, longstanding illness at age 46, debt at age 46, housing tenure at age 46; whether lives in London of Sout East at age 46, trust in people at age 46, whether voted in May 2015 General Election, earnings at age 46, total household income at age 46, employment status at age 46, whether had fixed place of work at age, index of multiple deprivation at age 46, BCS 46 partnership status at age 46, who cohort member lived with during COVID-19.
2. Log income was modelled with a linear regression, values >0 represent higher likelihood of video interview. For all other outcomes a modified Poisson regression was employed a risk ratio > 1 represents higher likelihood of video interview, while a risk ratio <1 represents a higher likelihood for in person interview.

Figure 3: Risk ratios and 95% confidence intervals of the association between survey mode (in person vs video) and six key outcomes.

1. Adjusted Risk Ratios from models including parental sex at birth, parental social class at birth number of rooms at home/persons per room at birth, cognitive ability at age 10, early life mental health at age 16, voting at age 42, membership in organisations at age 42, internet access at age 46, consent for biomarkers at age 46, educational qualifications at age 42, economic activity at age 46, partnership status at age 46; psychological distress at age 46, Body Mass Index at age 46, self rated health at age 46; smoking status at age 46, social support at age 46, income at age 46, number of non-responses across all previous sweeps, response to COVID-19 wave 1 web survey, response to COVID-19 wave 2 web survey, financial circumstances during COVID-19, relationship satisfaction at age 46; self-efficacy at age 46, whether had savings at age 46, longstanding illness at age 46 debt at age 46, housing tenure at age 46; whether lives in London of Sout East at age 46, trust in people at age 46, whether voted in May 2015 General Election, earnings at age 46, total household income at age 46, employment status at age 46, whether had fixed place of work at age, index of multiple deprivation at age 46, BCS 46 partnership status at age 46, who cohort member lived with during COVID-19.

2. All outcomes were modelled with a modified Poisson regression was employed a risk ratio > 1 represents higher likelihood of video interview, while a risk ratio <1 represents a higher likelihood for in person interview.

The unadjusted results in Table 16 and Figure 3 show differences between video and in person interviews. Those who were employed, homeowners, or had higher incomes were more likely to do video interviews. In contrast, individuals with poor self-rated health, psychological distress, or obesity were less likely to opt for video interviews, suggesting that health and well-being may have influenced mode preference. Overall,

these differences could be due to those from advantaged socioeconomic backgrounds, who also tended to be healthier, being more likely to prefer video interviews, or mode induced measurement error (mode effects). To distinguish between mode effects and selection requires comparison with adjusted results.

However, the adjusted results, which accounted for drivers of selection into mode, provided insights on differences in responses that arose from the interview mode itself in the form of measurement error ("mode effects") rather than differences in the types of participants selecting each mode. Compared to the unadjusted results, the strong attenuation observed across all six outcomes suggested that mode differences were primarily driven by selection effects rather than mode effects. For example, in the unadjusted results, employed participants were more likely to have chosen video interviews (RR = 1.08, 95%CI: 1.05 to 1.11), but after adjustment, this was greatly reduced (Adjusted RR = 1.02, 95%CI: 0.99 to 1.04), indicating that the initial difference was due to selection, reflecting that employed individuals had different characteristics that influenced their preference for video interviews. Similarly, homeowners showed a higher likelihood of selecting video interviews in the unadjusted model (RR = 1.12, 95%CI: 1.09 to 1.14), but this disappeared in the adjusted results (RR = 1.00, 95%CI: 0.99 to 1.02). A similar pattern of attenuation was observed for health-related variables. In the unadjusted model, participants with poor self-rated health were less likely to have selected video interviews (RR = 0.70, 95%CI: 0.64 to 0.77), but this difference weakened substantially after adjustment (RR = 0.94, 95%CI: 0.86 to 1.02). Likewise, for psychological distress, the initial preference for in-person interviews observed in the unadjusted model (RR = 0.77, 95%CI: 0.69 to 0.86) was no longer evident after adjustment (RR = 1.00, 95%CI: 0.90 to 1.10). These findings indicated that health-related mode differences were largely driven by selection effects rather than mode effects, meaning that individuals with poorer health or psychological distress may have had characteristics that led them to preferring in-person interviews, rather than the interview mode itself influencing their responses. For obesity, the attenuation was less pronounced, with a small reduction in effect size from the unadjusted estimate (RR = 0.93, 95%CI: 0.87 to 1.00) to the adjusted estimate (RR = 0.99, 95%CI: 0.93 to 1.07).

Overall, the substantial attenuation of effects in the adjusted models confirmed that observed mode differences were primarily due to selection effects rather than mode

effects. The strongest selection effects were seen for income, self-rated health, and psychological distress, but even then initial differences between video and in-person interviews diminished after adjustment. The finding that mode effects do not drive mode differences in the six outcomes considered here suggests that data analysis can proceed as it would when data are collected from a single survey mode. We note that since we have adjusted for a somewhat limited set of imperfectly measured characteristics, it is likely that adjustment for a further and/or better measured characteristics would likely attenuate estimates further - i.e. selection into mode is likely even greater and therefore mode effects even less pronounced. However, this may not apply to outcomes from other domains available in the age 51 sweep. If mode effects are suspected, or substantial mode differences persist after preliminary analysis similar to that presented here, it is advisable to follow the guidance in the Handling Survey Mode in CLS Cohorts User Guide (Wright, 2024).

References

F. Babor, T., C. Higgins-Biddle, J., B. Saunders, J., & G. Monteiro, M. (2001). The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care. W. H. Organisation.

Brown, M., Fitzsimons, E., Goodman, A., Peters, A., Ploubidis, G.B., Sanchez, A., Silverwood, R., Smith, K. (2021) COVID-19 Survey in Five National Longitudinal Studies: Waves 1, 2 and 3 User Guide (Version 4). London: UCL Centre for Longitudinal Studies and MRC Unit for Lifelong Health and Ageing. https://cls.ucl.ac.uk/wp-content/uploads/2017/02/UCL-Cohorts-COVID-19-Surveyuser-guide.pdf

Brown, M., Peters, A. (2023) 1970 British Cohort Study: Age 46 User Guide (Version 2). UCL Centre for Longitudinal Studies Cutrona CE, Russell DW. The provisions of social support and adaptation to stress. Advance in Personal Relationships. 1987;1:37–67

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

Donnellan MB, Oswald FL, Baird BM, Lucas RE. The mini-IPIP scales: tiny-yet effective measures of the Big Five factors of personality. Psychol Assess. 2006 Jun;18(2):192-203. PubMed PMID: 16768595.

Ng Fat, L., Scholes, S., Boniface, S., Mindell J., & Stewart-Brown S. (2017) Evaluating and establishing the national norms for mental well-being using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. Quality of Life Research, 26(5), 1129-1144.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological Assessment, 4, 26-42.

Hanzlová, R. and Lynn, P., 2023. Item response theory-based psychometric analysis of the Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) among adolescents in the UK. Health and Quality of Life Outcomes, 21(1), p.108.

Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. Res Aging. 2004;26(6):655-672. doi: 10.1177/0164027504268574. PMID: 18504506; PMCID: PMC2394670.

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. Journal of Research in Personality, 51, 78-89.

Katsoulis, M., Narayanan, M., Dodgeon, B., Ploubidis, G., & Silverwood, R. (2024). A data driven approach to address missing data in the 1970 British birth cohort. *medRxiv*, 2024.2002. 2001.24302101.

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.

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

Michael T. McKay, James R. Andretta, Evidence for the Psychometric Validity, Internal Consistency and Measurement Invariance of Warwick Edinburgh Mental Well-being Scale Scores in Scottish and Irish Adolescents, Psychiatry Research, Volume 255, 2017,Pages 382-386, ISSN 0165-1781, https://doi.org/10.1016/j.psychres.2017.06.071.

Mostafa, T., Narayanan, M., Pongiglione, B., Dodgeon, B., Goodman, A., Silverwood, R. J., & Ploubidis, G. B. (2021). Missing at random assumption made more plausible: evidence from the 1958 British birth cohort. J Clin Epidemiol, 136, 44-54.

Ringdal, R., Eilertsen, M.-E., Bjørnsen, H., Espnes, G., & Moksnes, U. (2018). Validation of two versions of the Warwick-Edinburgh Mental Well-Being Scale among Norwegian adolescents. Scandinavian Journal of Public Health, 46, 140349481773539. https://doi.org/10.1177/1403494817735391

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470316696

Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, health, and behaviour.* London: Longman.McGee, R., Williams, S., and Silva, P. A. (1986) 'An evaluation of the Malaise Inventory', Journal of Psychosomatic Research, 30(2), pp.147-152.

Silverwood, R., Narayanan, M., Dodgeon, B., Katsoulis, M., Ploubidis, G. (2024) Handling missing data in the CLS cohort studies: User guide. London: UCL Centre for Longitudinal Studies.

Ware J.E., Snow K.K., Kosinsk,i M., & Gandek, B. (1993), SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: New England Medical Center, The Health Institute.

Wright, L., Ploubidis, G., Silverwood, R. (2024) Handling Mode Effects in the CLS Cohort Studies: Executive Summary. London: UCL Centre for Longitudinal Studies.

Appendix 1: Derived Variables

A1.1 Paradata

The following variables are derived from data surrounding the interview circumstances rather than data derived from within the interview itself.

BD11AGEINT - "(Derived) Age in months at interview"

Description: Cohort member's age in months at interview.

Population: All cohort members

Value Labels:

(-8) "Not enough information"

Derivation description: (Interview year (b11inty) – birth year) * 12 + (Interview month (b11intm) – birth month)

BD11AGEINTY – "(Derived) Age at interview (year part)" and **BD11AGEINTM – "(Derived) Age at interview (month part)"**

Description: This is a more readable format of the above variable separated into years and months (to be used in combination).

Value labels:

(-8) "Not enough information"

BD11SC_MODE – "(Derived) Mode for self-completion"

Description: How cohort members completed the section with sensitive questions

Value labels:

(-8) "Not enough information"

(-1) "Not applicable

(1) "Web – during video interview"

(2) "Web – after video interview"

(3) "Web – with interviewer during video interview"

(4) "Self-completion – during in person interview"

(5) "With interviewer – during in person interview"

(6) "Not completed"

A1.2 Geographical variables

A series of geographical variables have been derived from the addresses at which participants were interviewed (see below). Country of interview, Government Office Region, Index of Multiple Deprivation Rank and Urban/Rural indicator are available in the main file.

The remaining variables are provided in 'bcs_age51_geographical_identifiers' which is available via secure access due to the potentially disclosive nature of the information.

Variable name	Derived variable label	Dataset
bd11cntry	Country of interview	bcs_age51_main
bd11rgn	2020 Region of interview	bcs_age51_main
bd11imd	2015 Index of Multiple Deprivation rank decile	bcs_age51_main
bd11ruc11	2011 Urban/rural indicator	bcs_age51_main
b11imd	2015 Index of Multiple Deprivation rank	bcs_age51_geographical_identifiers
b11osward	2023 Ward	bcs_age51_geographical_identifiers
b11casward	2003 CAS ward	bcs_age51_geographical_identifiers
b11oa21	2021 OA Code	bcs_age51_geographical_identifiers

b11lsoa11	2011 Lower Super Output Area	bcs_age51_geographical_identifiers
b11msoa11	2011 Middle Super Output Area	bcs_age51_geographical_identifiers
b11oslaua	April 2023 Local Authority	bcs_age51_geographical_identifiers
b11pcon	2014 Westminster Parliamentary Constituency	bcs_age51_geographical_identifiers
b11wz11	2011 Workplace Zone	bcs_age51_geographical_identifiers
b11lsoa21	2021 Lower Super Output Area	bcs_age51_geographical_identifiers
b11msoa21	2021 Middle Super Output Area	bcs_age51_geographical_identifiers

A1.3 Household and family variables

Household derived variables were taken from the five person loops in the CAPI household grid: partner grid, two child grids (children reported at last sweep and additional children not previously mentioned), two other household members grids (other household members reported at last sweep and additional household members not previously mentioned). These cover all possible household members at the time of interview. The five loops have subsequently been combined into the person grid dataset.

BD11HSIZE – "(Derived) HH Size"

Description: Number of people currently living in same household as cohort member (includes cohort member)

Population: All cohort members

Value Labels:

Greater than zero

Derivation description: Household size includes: the cohort member, partner, own children in household and all other household members (b11gslive=1).

BD11NUMCH – "(Derived) Number of children in HH"

Description: The total number of children in the household according to relationship and not age, including children of other household members.

Population: All cohort members

Value labels:

(-8) "No information"

Derivation description: Number of children (b11grtok=4,5,6,7,8,28) in household (b11gslive=1).

BD11NOCHH – "(Derived) Number of own children in HH"

Description: How many of the respondent's natural children live in the household at the time of interview.

Population: All cohort members

Value labels:

(-8) "No information"

Derivation description: Number of own (biological) children (b11grtok=4) in household (b11gslive=1).

BD11NPCHH – "(Derived) Number of children of current or previous partner in HH"

Description: How many children of the respondent's current or previous partner live in the household at the time of interview.

Population: All cohort members

Value labels:

(-8) "No information"

Derivation description: Number of current partner's (b11grtok=6), or previous partner's (b11grtok=7) children in household (b11gslive=1).

Number of natural children by age

Description: There are 7 variables that provide the counts of natural children of the respondent according to age group.

BD11NOC2A – "(Derived) Number of own children aged 0 to 2 in HH or absent"

BD11NOC4A – "(Derived) Number of own children aged 3 to 4 in HH or absent"

BD11NOC11A – (Derived) Number of own children aged 5 to 11 in HH or absent

BD11NOC15A – (Derived) Number of own children aged 12 to 15 in HH or absent

BD11NOC20A – (Derived) Number of own children aged 16 to 20 in HH or absent

BD11NOC30A – (Derived) Number of own children aged 21 to 30 in HH or absent

BD11NOC31A – (Derived) Number of own children aged 31 and over in HH or absent

Population: All cohort members

Value labels:

Zero and above

Derivation description: The number of own children (b11grtok=4) reported within each age category (b11gage).

BD10NACAB – "(Derived) Number of absent children (including step-children etc)"

Description: Reported children of the respondent or other hh members who are not living in the household at the time of interview.

Population: All cohort members

Value labels:

Zero and above

Derivation description: The number of children (b11grtok=4,5,6,7,8,28) absent from household at interview (b11gslive=0).

BD11GCHLD – "(Derived) Whether has grandchildren (own or consider themselves to be)"

Description: Reported grandchildren of the respondent.

Population: All cohort members

Value labels:

Zero and above

Derivation description: Any grandchildren in household (b11grtok=20) or absent (b11gcnum > 0).

BD11NGCHLD – "(Derived) Total number of grandchildren (own or consider themselves to be)"

Description: Number of reported grandchildren of the respondent

Population: All cohort members

Value labels:

   Zero and above

Derivation description: The number of grandchildren in household (b11grtok=20) or absent (b11gcnum).

BD11AYCHH – "(Derived) Age of youngest child (under 19) in household"
Description: The youngest child living in the respondent's household under the age of 19. 'Child' is defined in the same manner as bd11numch

Population: All cohort members

Value labels:

   (-1) "Not applicable"

Derivation description: The youngest reported child (b11grtok=4,5,6,7,8,28) in household (b11gslive=1) if age is under 19 (b11gage<19).

BD11AOCHH – "(Derived) Age of oldest child in household"

Description: The oldest child (defined by bd11numch) living in the household.

Population: All cohort members

Value labels:

(-1) "Not applicable"

Derivation description: The oldest reported child (b11grtok=4,5,6,7,8,28) in household (b11gslive=1).

A1.4 Relationships

BD11MS – "(Derived) CM's legal marital status"

Description: The legal marital status of the respondent

Population: All cohort members

Value labels:

(-8) "No information"

(1) "Legally separated"

(2) "Married"

(3) "Divorced"

(4) "Widowed"

(5) "A Civil Partner"

(6) "A former Civil Partner"

(7) "A surviving Civil Partner"

(8) "Never married or in a Civil Partnership"

Derivation description: Combined updated martial status (b11hms) and existing marital status if unchanged from previous sweep (b11marchk, b11divchk).

BD11COHAB – "(Derived) Whether CM cohabiting as a couple"

Description: Whether the cohort member has a co-resident spouse or partner

Population: All cohort members

Value labels:

(-8) "No information"

(-1) "Not applicable"

(0) "No"

(1) "Yes"

Derivation description: Partners in relationship histories grid recorded as continuously living with CM.

A1.5 Housing

BD11TENURE – "(Derived) Housing Tenure"

Description: Tenure of cohort member's accommodation

Population: All cohort members

Value labels

(-8) "No information"

(1) "Own outright"

(2) "Own, buying with help of mortgage/loan"

(3) "Part rent, part mortgage (shared equity)"

(4) "Rent it"

(5) "Live rent-free, incl. relatives/friends"

(6) "Squatting"

(7) "Other"

Derivation description: bd11ten (Current tenure) if tenure has changed or previous reported tenure if same as last sweep (bd11tenck=1). Category 6 has been removed and cases added to "Other" for EUL, and is only available to SA users.

BD11RENTFROM – "(Derived) Who rents from"

Description: Owner of cohort member's rented housing

Population: Cohort members in rented housing

Value labels:

(-8) "No information"

(-1) "Not applicable"

(1) "A Local Authority"

(2) "A Housing Association"

(3) "A Private landlord"

(4) "A Parent"

(5) "Someone else"

Derivation description: b11rentfrom (current rental provider) if tenure has changed, or previous reported rental provider if tenure is same as last sweep.

BD11WHOTEN – "(Derived) Whose name is accommodation held in"

Description: Name cohort member's housing held in

Population: Homeowners and/or rent payers

Value labels:

(-8) "No information"

(-1) "Not applicable"

(1) "CM's name only"

(2) "Both CM and partner's names"

(3) "Partner's name only"

(4) "CM and someone else's name's"

(5) "Someone else's name (including parents)"

Derivation description: Based on b11wten, provided bd11tenure is between 1-5 so that the current accommodation can be in someone's name.

BD11TIMAD - "(Derived) Time at current address (months)"

Description: Number of months cohort member has lived at current address

Population: All cohort members

Value Labels:

(-9) "Refused"

(-8) "Insufficient information"

(-1) "Not applicable"

Derivation description: The date the cohort member moved into their current address (01, b11movinm, b11moviny) subtracted from the date of interview (b1intm, b11inty).

A1.6 Education

BD11ACHQ1 – "(Derived) Highest Academic Qualification CM reported in Age 51 Survey"

Description: This is highest academic qualification obtained since the 2016 survey or 2012 if they were not interviewed last sweep.

Population: All cohort members

Value labels:

(-8) "Not enough information"

(-1) "Not applicable"

(0) "no academic qualification"

(1) "gcse d-e"

(2) "cses2-5, other Scottish quals"

(3) "gcse a-c, good o levels scot standards"

(4) "as levels or 1 a level"

(5) "2+ a levels, scot higher/6th"

(6) "diploma"

(7) "degree level"

(8) "higher degree"

Derivation description: All loops of b11qualtp in the education grid. The qualification from b11qualtp attributed to the highest value of those listed in value labels.

BD11ANVQ1 – "(Derived) Highest NVQ level from an academic qualification reported in Age 51 Survey"

Description: Highest academic qualification at the current sweep (as defined in bd11achq1) but categorised by their equivalent NVQ level rather than qualification type.

NVQ format (levels 1-5) for consistency with previous BCS70 data.

Population: All cohort members

Value labels:

(-1) "not codeable"

(0) "none"

(1) "nvq1 level"

(2) "nvq2 level"

(3) "nvq3 level"

(4) "nvq4 level"

(5) "nvq5 level"

Derivation description: All loops of b11qualtp in the education grid. The academic qualification from b11qualtp attributed to the highest value of those listed in value labels

BD11ALVL1 - "(Derived): Highest qualification level from an academic qualification reported in the Age 51 Survey– 8 level version"

Description: Highest academic qualification at the current sweep (as defined in bd11achq1) but categorised by level rather than qualification type (9 categories – 0-8).

Population: All cohort members

Value Labels:

(-8) "Insufficient information"

(-1) "Not applicable"

(0) "Entry Level"

(1) "Level 1"

(2) "Level 2"

(3) "Level 3"

(4) Level 4"

(5) "Level 5"

(6) "Level 6"

(7) "Level 7"

(8) "Level 8"

(95) "Other academic qualification"

(96) "No academic qualification"

Derivation description: All loops of b11qualtp in the education grid. The academic qualification from b11qualtp attributed to the highest value of those listed in value labels.

Nine category qualification levels for England, Wales and Northern Ireland What qualification levels mean: Overview - GOV.UK with equivalent levels for Scottish qualifications www.sqa.org.uk/sqa/64561.html

BD11VNVQ1 – "(Derived) Highest NVQ level from a Vocational Qualification reported in Age 51 Survey"

Description: Vocational qualification with the highest associated NVQ level obtained since the 2016 survey or 2012 if they were not interviewed then.

Used NVQ format (levels 1-5) for consistency with previous BCS70 data.

Population: All cohort members

Value **labels:**

(-1) "not codeable"

(0) "none"

(1) "nvq1 level"

(2) "nvq2 level"

(3) "nvq3 level"

(4) "nvq4 level"

(5) "nvq5 level"

Derivation description: All loops of b11qualtp in the education grid. The vocational qualification from b11qualtp attributed to the highest value of those listed in value labels

BD11VLVL1 - "(Derived): Highest qualification level from a vocational qualification reported in the Age 51 Survey– 8 level version"

Description: Vocational qualification with the highest associated education level obtained since the 2016 survey or 2012 if they were not interviewed then. (9 categories, 0-8)

Population: All cohort members

Value Labels:

(-8) "Insufficient information"

(-1) "Not applicable"

(0) "Entry Level"

(1) "Level 1"

(2) "Level 2"

(3) "Level 3"

(4) Level 4"

(5) "Level 5"

(6) "Level 6"

(7) "Level 7"

(8) "Level 8"

(95) "Other vocational qualification"

(96) "No vocational qualification"

Derivation description: All loops of b11qualtp in the education grid. The vocational qualification from b11qualtp attributed to the highest value of those listed in value labels

Nine category qualification levels for England, Wales and Northern Ireland What qualification levels mean: Overview - GOV.UK with equivalent levels for Scottish qualifications www.sqa.org.uk/sqa/64561.html

BD11NVQ1 – "(Derived) Highest NVQ level from an Academic or Vocational Qual reported in Age 51 Survey"

Description: The associated NVQ level of the highest overall qualification obtained since the 2016 survey or 2012 if they were not interviewed then.

NVQ format (levels 1-5) for consistency with previous BCS70 data.

Population: All cohort members

Value labels:

(-1) "not codeable"

(0) "none"

(1) "nvq1 level"

(2) "nvq2 level"

(3) "nvq3 level"

(4) "nvq4 level"

(5) "nvq5 level"

Derivation description: The NVQ level of either the highest academic qualification (bd11anvq1) and vocational qualification (bd11vnvq1) depending on which is higher.

BD11LVL1- "(Derived): Highest qualification level from an academic or vocational qualification reported in the Age 51 Survey– 8 level version"

Description: Cohort member's highest qualification level achieved from an academic or vocational qualification – self reported (9 categories, 0-8)

Population: All cohort members

Value Labels:

(-8) "Insufficient information"

(-1) "Not applicable"

(0) "Entry Level"

(1) "Level 1"

(2) "Level 2"

(3) "Level 3"

(4) Level 4"

(5) "Level 5"

(6) "Level 6"

(7) "Level 7"

(8) "Level 8"

(95) "Other qualification"

(96) "No qualifications"

Derivation description: Level of either the highest academic qualification (bd11anvq1) and vocational qualification (bd11vnvq1) depending on which is higher.

Nine category qualification levels for England, Wales and Northern Ireland What qualification levels mean: Overview - GOV.UK with equivalent levels for Scottish qualifications www.sqa.org.uk/sqa/64561.html

BD11HACHQ – "(Derived) Highest Academic Qualification up to age 51"

Description: The highest overall academic qualification of the cohort member.

Population: All cohort members

Value labels:

(-8) "Not enough information"

(-1) "Not applicable"

(0) "no academic qualification"

(1) "gcse d-e"

(2) "cses2-5, other scottish quals"

(3) "gcse a-c, good o levels scot standards"

(4) "as levels or 1 a level"

(5) "2+ a levels, scot higher/6th"

(6) "diploma"

(7) "degree level"

(8) "higher degree"

Derivation description: The highest academic qualification at the current sweep (bd11achq1) compared to the highest level recorded from all previous sweeps

BD11HANVQ – "(Derived) Highest NVQ level from an academic qualification up to age 51"

Description: The highest overall NVQ level from an academic qualification of the cohort member.

Population: All cohort members

Value labels:

(-1) "not codeable"

(0) "none"

(1) "nvq1 level"

(2) "nvq2 level"

(3) "nvq3 level"

(4) "nvq4 level"

(5) "nvq5 level"

Derivation description: The highest NVQ level from an academic qualification (bd11anvq1) was compared to the highest academic qualification from all previous sweeps

BD11HVNVQ – "(Derived) Highest NVQ Level from a Vocational Qualification up to age 51"

Description: The highest overall NVQ level from a vocational qualification of the cohort member.

Population: All cohort members

Value labels:

(-1) "not codeable"

(0) "none"

(1) "nvq1 level"

(2) "nvq2 level"

(3) "nvq3 level"

(4) "nvq4 level"

(5) "nvq5 level"

Derivation description: The highest NVQ level from a vocational qualification (bd11vnvq1) compared to the highest vocational qualification from all previous sweeps

BD11HNVQ – "(Derived) Highest NVQ Level from an Academic or Vocational Qual up to age 51"

Description: The highest qualification of any type the cohort member has ever obtained

Population: All cohort members

Value labels:

(-1) "not codeable"

(0) "none"

(1) "nvq1 level"

(2) "nvq2 level"

(3) "nvq3 level"

(4) "nvq4 level"

(5) "nvq5 level"

Derivation Description: Either the highest academic qualification (bd11hanvq) or the highest vocational qualification (bd11hvnvq) ever obtained represented as an NVQ level.

BD11DEGP - "(Derived) Whether achieved first degree or higher"

Description: Whether cohort member has achieved a degree level qualification or higher

Population: All cohort members

Value labels:

(-8) "Insufficient information"

(0) "No degree"

(1) "First or higher degree"

Derivation Description: Based on derived academic qualifications recorded, if highest recorded level is degree or higher

8.7 Health

BD11DISEQ – "(Derived) Disability classification Equality act (2010)"

Description: The classification of whether or not the respondent is disabled according to the Equality act 2010

Population: All cohort members

Value labels:

(-8) "Not enough information"

(0) "Not disabled (Equality act)"

(1) "Disabled (Equality Act)"

Derivation description: Classified disabled if any physical/mental health conditions lasting or expected to last 12 months (bd11loil=1) and illnesses/conditions reduce ability to carry out day to day activities (b11lolm=1,2).

BD11DISLS – "(Derived) Disability classification EU-SILC"

Description: Further classification of disability to ascertain the severity of the respondent's illness/condition.

Population: All cohort members

Value labels:

(-8) "Not enough information"

(0) "No EU-SILC long-standing health condition"

(1) "EU-SILC classification to some extent"

(2) "EU-SILC classification severely hampered"

Derivation description: Longstanding illness and ability to carry out day to day activities is reduced a little for 6 months or more (b11loil=1 AND (b11lolp=2,3) AND (b11lolm=2)), CM is disabled/hampered to some extent. Longstanding illness and the ability to carry out day to day activities is reduced a lot for 6 months or more (b11loil=1 AND (b11lolp=2,3) AND (b11lolm=1)) CM is categorised as being disabled, severely hampered.

For more information on disability classification see ONS document 'Long-lasting Health Conditions and Illnesses; Impairments and Disability': www.ons.gov.uk/ons/guide-method/harmonisation/primary-set-of-harmonised-concepts-and-questions/index.html

BD11HGHTM – "(Derived) Self-reported height in metres"

Description: Most recent self-reported height measurement in metres and centimetres

Population: All cohort members

Value labels:

(-8) "No information"

Derivation description: Self-reported height provided as metres and centimetres (b11htmees, b11htcms) combined with height in feet and inches converted using the

formula (b11htfeet*12)+b11htines)/39.370. When any height measurements are missing, the last self-reported height recorded from sweeps 9/10 are used.

BD11WGHTK – "(Derived) Self-reported weight in kilograms"

Description: Current self-reported weight in kilograms

Population: All cohort members

Value labels:

(-8) "No information"

Derivation description: Self-reported weight in kilograms (b11wtkis) combined with weight in stones and pounds, converted to kg with the formula (b11wtste*14+b11wtpod)*0.453592.

BD11BMI – "(Derived) Body mass index (based on self-reported data)"

Description: Body mass index of cohort member, based on self-reported height and weight

Population: All cohort members

Value labels:

(-8) "Not enough information"

Derivation description: Calculated with the derived self-reported height and weight measurements using the formula bd11wghtk/(bd11hghtm* bd11hghtm).

BD11BMIC – "(Derived) Body mass index – classification (based on self-reported data)"

Description: Self-reported BMI values separated into 5 weight range categories defined by the same classification as that of the NHS.

Population: All cohort members

Value labels:

(-8) "Not enough information"

(0) "Underweight (below 18.5)"

(1) "Healthy weight range (18.5-24.9)"

(2) "Overweight (25-29.9)"

(3) "Obese (30-39.9)"

(4) "Morbidly obese (Over 40)"

BD11SMOKE – "(Derived) Smoking habits"

Description: Typical number of cigarettes smoked daily

Population: All cohort members

Value labels:

(-8) "Not enough information"

(-1) "Not applicable"

(0) "Never smoked"

(1) "Ex smoker"

(2) "Occasional smoker"

(3) "Up to 10 a day"

(4) "11 to 20 a day"

(5) "More than 20 a day"

(6) "Daily but frequency not stated"

Derivation description: For current smokers uses the number of cigarettes smoked daily (b11nfcigs) as a categorical range. Non-smokers and ex smokers derived from b11smokig

BD11AUDIT – "(Derived) Total AUDIT-PC score"

Description: Cohort member's score on the Alcohol Use Disorders Identification Test

Population: Cohort members that drink alcohol who completed the paper self-completion

Value labels:

(-8) "Not enough information"

(-1) "NA – does not drink alcohol"

Derivation description: Total score of 5 questions (b11q11, b11q12, b11q13a, b11q13b, b11q14), each on a 0-4 point scale. Scores of 5 or above indicate higher risk drinking

BD11AUDG – "(Derived) AUDIT-PC Group"

Description: A grouped version of bd11audit, according to the definition of low and high-risk drinking

Population: Cohort members that drink alcohol who completed the paper self-completion

Value labels:

(-8) "Not enough information"

(-1) "NA – does not drink alcohol"

(1) "Unproblematic drinking (0-4)"

(2) "Increasing or high risk drinking (5+)"

A1.8 Mental Health/Wellbeing

BD11WEMWB – "(Derived) Warwick Edinburgh Mental Well-Being Scale"

Description: A short (7-item) version of the scale designed to represent the respondent's mental well-being. Score ranges from 7-35 where higher scores indicate higher positive mental wellbeing

Population: All cohort members who completed the paper self-completion

Value labels:

    (-8) "Not enough information"

Derivation description: The sum of 7 items, each on a 5-point scale (b11q35a+ b11q35b+ b11q35c+ b11q35d+ b11q35e+ b11q35f+ b11q35g)

BD11MAL – "(Derived) Total Malaise score (9 questions)"

Description: Total malaise score from a 9-item short version of the malaise inventory. Items cover negative emotions and physical response and are coded so high malaise scores always relate to affirmative responses

Population: All cohort members who completed the self-completion

Value labels:

    (-8) "Not enough information"

Derivation description: The score is the total number of 'yes' responses (b11mal02, b11mal03, b11mal05, b11mal09, b11mal12, b11mal14, b11mal16, b11mal20, b11mal21) with higher scores corresponding to higher malaise. Cases were excluded if the number of items without a response accounted for scores being under 4 (the prerequisite score for high malaise).

BD11MALG – "(Derived) Total Malaise score – grouped"

Description: A grouped version of bd11mal, according to the definition of low and high malaise. High malaise requires at least four affirmative responses.

Population: All cohort members who completed the self-completion

Value labels:

(-8) "Not enough information"

(1) "Low malaise (0-3)"

(2) "High malaise (4+)"

Short-form health survey

The 36 item short-form health survey (which was integrated into the self-completion questionnaire) covers eight different aspects of health (4 mental and 4 physical), as well as comparing general health to one year ago. There are 9 derived scores (one for each aspect), all ranging between 0 and 100, meaning original variables are scaled accordingly to calculate this. For all derived variables higher scores indicate better health in the area described. One response for the associated questions is sufficient for a score to be created for these variables.

BD11PHHE – "(Derived) SF-36 Physical functioning score"

Description: Physical health score coded from 10 items. Higher scores represent better functioning across a range of physical activities (e.g. running, dressing, climbing stairs).

Population: All cohort members who completed the paper self-completion

Value labels:

(-8) "Not enough information"

Derivation description: Recoded 10 items in scale (b11q16a, b11q16b, b11q16c, b11q16d, b11q16e, b11q16f, b11q16g, b11q16h, b11q16i, b11q16j) to 1=0, 2=50, 3=100. Physical functioning scores are the mean of the total number of questions answered (between 1 and 10 responses).

BD11RLMP – "(Derived) SF-36 Role-limitations due to physical health"

Description: Physical health score coded from 4 items. Lower scores indicate greater lifestyle limitations as a result of physical health problems in the four weeks prior to interview.

Population: All cohort members who completed the paper self-completion

Value labels:

(-8) "Not enough information"

Derivation description: Recoded 4 items in scale (b11q17a, b11q17b, b11q17c, b11q17d) to 1=0, 2=100. Scores are the mean of the total number of questions answered (between 1 and 4 responses).

BD11RLME – "(Derived) SF-36 Role-limitations due to emotional problems"

Description: Mental health score coded from 3 items. Lower scores indicate greater lifestyle limitations as a result of emotional problems in the four weeks prior to interview.

Population: All cohort members who completed the paper self-completion

Value labels:

(-8) "Not enough information"

Derivation description: Recoded 3 items in scale (B11q18a, b11q18b, b11q18c) to 1=0, 2=100. Scores are the mean of the total number of questions answered (between 1 and 3 responses).

BD11ENFA – "(Derived) SF-36 Energy/fatigue score"

Description: Mental health score coded from 4 items. Lower scores represent increased tiredness and lack of energy in the four weeks prior to interview

Population: All cohort members who completed the paper self-completion

Value labels:

(-8) "Not enough information"

Derivation description: Two of the items (b11q22a, b11q22e) are recoded to 1=100, 2=80, 3=60, 4=40, 5=20, 6=0, while the remaining two (b11q22g, b11q22i) are reverse-coded to 1=0, 2=20, 3=40, 4=60, 5=80, 6=100. Energy/fatigue scores are the mean of the total number of questions answered (between 1 and 4 responses).

BD11EMWB – "(Derived) SF-36 Emotional Well-Being score"

Description: Mental health score coded from 5 items. Higher scores result from an overall more positive emotional outlook during the four weeks prior to interview.

Population: All cohort members who completed the paper self-completion

Value labels:

(-8) "Not enough information"

Derivation description: Three items (b11q22b, b11q22c, b11q22f) are coded so 1=0, 2=20, 3=40, 4=60, 5=80, 6=100, while the other two items (b11q22d, b11q22h) are coded so 1=100, 2=80, 3=60, 4=40, 5=20, 6=0. Emotional well-being scores are the mean of the total number of questions answered (between 1 and 5 responses).

BD11SOCF – "(Derived) SF-36 Social Functioning score"

Description: Mental health score coded from 2 items. Lower scores indicate greater impact of emotional and physical problems on social life.

Population: All cohort members who completed the paper self-completion

Value labels:

(-8) "Not enough information"

Derivation description: b11q19 is a 5-point scale recoded to 1=100, 2=75, 3=50, 4=25, 5=0 and b11q22j is a 6-point scale recoded so 1=0, 2=20, 3=40, 4=60, 5=80,

6=100. Social functioning scores are the mean of the total number of questions answered (1 or 2 responses).

BD11PAIN – "(Derived) SF-36 Pain score"

Description: Physical health score coded from 2 items. Lower scores show an increase in pain and its impact on normal life.

Population: All cohort members who completed the paper self-completion

Value labels:

(-8) "Not enough information"

Derivation description: b11q20 is a 6-point scale recoded so 1=100, 2=80, 3=60, 4=40, 5=20, 6=0. b11q21 is on a 5-point scale recoded so 1=100, 2=75, 3=50, 4=25, 5=0. Pain scores are the mean of the total number of questions answered (1 or 2 responses).

BD11GENH – "(Derived) SF-36 General health score"

Description: Physical health score coded from 5 items. Higher scores indicate better overall health.

Population: All cohort members who completed the paper self-completion

Value labels:

(-8) "Not enough information"

Derivation description: Three of the items (b11hlthgn, b11q23b, b11q23d) are recoded so that 1=100, 2=75, 3=50, 4=25, 5=0, while b11q23a and b11q23c are recoded to 1 =0, 2=25, 3=50, 4=75, 5=100). General health scores are the mean of the total number of questions answered (between 1 and 5 responses).

BD11RPHT – "(Derived) SF-36 Reported health transition"

Description: General health score coded from one item from the core interview. Scores close to 50 indicate similar health compared to one year ago, while higher scores are better health and scores closer to 0 worse.

Population: All cohort members

Value labels:

   (-8) "Not enough information"

Derivation description: b11khlstt is the only item recoded so health transition score translates as 1=100, 2=75, 3=50, 4=25, 5=0.

BD11LONELINESS – "(Derived) UCLA Loneliness 3-item"

Description: Measurement of loneliness

Population: All cohort members who completed the self-completion

Value labels:

   (-8) "Not enough information"

Derivation description: Addition of the scores of three questions (b11lonela, b11lonelb, b11lonelc). Only computed if all questions have valid responses.

BD11GAD2 – "(Derived) Generalised Anxiety Disorder 2-item"

Description: Measurement of the frequency of feeling nervous, anxious or on edge.

Population: All cohort members who completed the self-completion

Value labels:

   (-8) "Not enough information"

Derivation description: Addition of the scores of two questions (b11gad2phq2a, b11gad2phq2b). Only computed if all questions have valid responses.

BD11PHQ2 – "(Derived) Patient Health Questionnaire 2-item"

Description: Measurement of the frequency of depressed mood over the past 2 weeks

Population: All cohort members who completed the paper self-completion

Value labels:

(-8) "Not enough information"

Derivation description: Addition of the scores of two questions (b11gad2phq2c, b11gad2phq2d). Only computed if all questions have valid responses.

BD11IPIP1 – "(Derived) Personality IPIP-FFM: Neuroticism"

Description: International personality item pool measure of neuroticism

Population: All cohort members who completed the self-completion

Value labels:

(-8) "Not enough information"

Derivation description: Mean of four questions; two as originally coded (b11q25d, b11q25n) and two reverse-coded (b11q25i, b11q25s).

BD11IPIP2 – "(Derived) Personality IPIP-FFM: Extraversion"

Description: International personality item pool measure of extraversion

Population: All cohort members who completed the self-completion

Value labels:

(-8) "Not enough information"

Derivation description: Mean of four questions; two as originally coded (b11q25a, b11q25k) and two reverse-coded (b11q25f, b11q25p).

BD11IPIP3 – "(Derived) Personality IPIP-FFM: Openness"

Description: International personality item pool measure of openness

Population: All cohort members who completed the self-completion

Value labels:

 (-8) "Not enough information"

Derivation description: Mean of four questions; one as originally coded (b11q25e) and three reverse-coded (b11q25j, b11q25o, b11q25t).

BD11IPIP4 – "(Derived) Personality IPIP-FFM: Agreeableness"

Description: International personality item pool measure of agreeableness

Population: All cohort members who completed the self-completion

Value labels:

 (-8) "Not enough information"

Derivation description: Mean of four questions; two as originally coded (b11q25b, b11q25l) and two reverse-coded (b11q25g, b11q25q).

BD11IPIP5 – "(Derived) Personality IPIP-FFM: Conscientiousness"

Description: International personality item pool measure of conscientiousness

Population: All cohort members who completed the self-completion

Value labels:

 (-8) "Not enough information"

Derivation description: Mean of four questions; two as originally coded (b11q25c, b11q25m) and two reverse-coded (b11q25h, b11q25r).

A1.9 Finance

BD11BENE - "(Derived) Whether cohort member or partner receives any benefits"

Description: Whether cohort member or partner receives any benefits

Population: All cohort members

Value labels:

(-9) "Refused"

(-8) "Not known"

(1) "Receives benefits"

(2) "Does not receive benefits"

Derivation description: Cohort members are categorised as receiving benefits if they reported receiving universal credit (b11uncr) or any of the benefits mentioned in b11bent01-b11bent11.

A1.10   Activities and Employment

BD11NSSEC7 – "(Derived) NS-SEC 7 analytic classes (CMs current job)"

Description: Cohort member's current job in seven category NS-SEC format

Population: All cohort members

Value labels:

(-9) "Not classified"

(-8) "Never worked and long-term unemployed"

(-1) "Not applicable"

(1) "Higher managerial and professional occupations"

(2) "Lower managerial and professional occupations"

(3) "Intermediate occupations"

(4) "Small employers and own account workers"

(5) "Lower supervisory and technical occupations"

(6) "Semi-routine occupations"

(7) "Routine occupations"

Derivation description: Derived from full NS-SEC analytic sub-classes (b11cjnss8) into 7 classes

BD11NSSEC5 – "(Derived) NS-SEC 5 analytic classes (CMs current job)"

Description: Cohort member's current job in five category NS-SEC format

Population: All cohort members

Value labels:

(-9) "Not classified"

(-8) "Never worked and long-term unemployed"

(-1) "Not applicable"

(1) "Managerial and professional occupations"

(2) "Intermediate occupations"

(3) "Small employers and own account workers"

(4) "Lower supervisory and technical occupations"

(5) "Semi-routine and routine occupations"

Derivation description: Further simplified from NS-SEC analytic sub-classes (b11cjnss8) into 5 classes.

A1.11   Cognitive Skills

BD11NART – "(Derived) NART – Number of words correctly pronounced"

Description: Total number of words out of the 25 in the NART pronounced correctly by the cohort member

Population: All cohort members who completed the NART

Value labels:

    (-8) "Not enough information"

Derivation description: Sum of correct answers from the NART (bd11narta-bd11narty) provided the entire test was completed to provide a score between 0 and 25

Appendix 2: Deriving total net family income in BCS70 age 51

A2.1 Introduction

Total family income is an outcome and/or covariate of key interest for many disciplines, but the estimation of it is not straightforward. This section explains the procedure that was used to derive total net family income for BCS70 at the age 51 sweep. We estimate net, rather than gross, income as it is a common measure of living standards. Our definition of family is consistent with a benefit unit, where dependent children have no income, and includes the main cohort member (CM) and their partner, if applicable. Payments received for dependent children, such as child benefit or child maintenance, are also incorporated into the calculation.

A2.2 Components of income

Total net family income is created by summing the four components, which are:

a) CM's net earnings from employment or self-employment, including second job and any income from occasional work, if applicable.
b) CM's partner's net earnings from employment or self-employment, if applicable.
c) CM's and their partner's benefit income (such as jobseekers' allowance, child benefit, income support etc.).
d) CM's and their partner's other regular income (for example, rent payment, maintenance grant, pension etc.).

A2.2.1 Cohort members' net earnings

The CM's net earnings encompass earnings from their main job, if CM is in paid work (i.e. employed or self-employed) as well as earnings from second job and/or occasional work, if applicable.

Earnings from main job are derived separately for employees and the self-employed, as described below. Figure 3 shows how the sample in the respective groups was broken down by how the estimates for earnings were derived. The total sample for which the income was derived in N = 7789, which excludes emigrants, but includes participants living in Great Britain who took part in the mop-up survey (n=450)[11] .
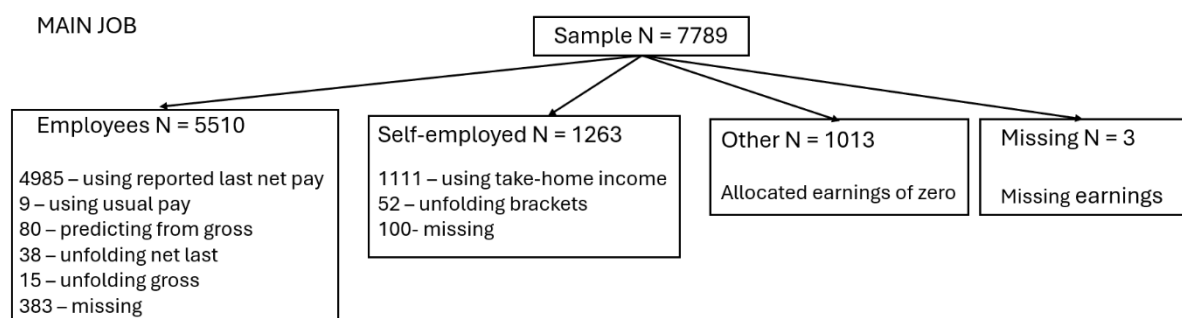

Figure 3: Sample sizes and derivation strategy for earnings from main job

Employees

Employees are defined as those CMs whose current main economic activity is employed, either part or full time. Employees are asked about their pay from main job in three ways:

1. Gross pay (before any deductions for tax, National Insurance, pension, union dues etc.)
2. Net (i.e. take-home) pay the last time cohort member was paid (after any deductions were made for tax, National Insurance, pension, union dues etc.)
3. If the last take-home pay is not equal to the usual take home pay, usual pay is also asked

Each of these questions are reported for the period that best suits respondents, but all are automatically converted to weekly, and soft checks are conducted for implausible values (for soft checks details see questionnaire).

---

[11] Mop-up survey participants were asked their current economic activity but were not asked questions on earnings or other income, and therefore this group constitutes a large proportion of the missing cases, whose income is estimated by multiple imputations (for more details on this approach see section A2.3)

We use reported net earnings from last pay in the first instance. In cases where employees report weekly earnings of zero, we re-coded them into missing, based on the assumption that if they are in paid employment their earnings must be greater than zero. This approach relies on the richness of the BCS70 life course data included in the imputation to estimate earnings (see section A2.3 for more details). In cases where CMs do not report their net pay but provide their usual pay, usual pay is used instead.

In cases where CMs do not report net or usual pay but report gross pay we convert these to net using the following procedure:

1) we exclude cases where reported gross income is lower than net or the ratio of net to gross is less than 0.5;
2) we use natural logarithm transformations on both net and gross earnings;
3) using the cases where both types of earnings are reported, we fit a linear quadratic regression to quantify the relationship between net and gross earnings;
4) for cases where only gross earnings are reported, we predict net earnings using the equation estimated in 3, which are then exponentiated.

If CMs do not provide values for earnings they are asked a series of unfolding brackets questions, that aim to estimate the minimum and maximum value of earnings, rather than their exact value. In cases where both minimum and maximum values are provided, or where the upper bracket is the minimum value allowing the lower bracket to be estimated as zero, we estimate earnings as the mid-point between the minimum and maximum value. This information is used in cases where CM did not provide value for net, usual or gross pay. When only gross pay is reported in unfolding brackets questions, we apply the conversion to net pay as described above.

Self-employed

The self-employed are defined as those cohort members whose current, main economic activity as self-employed, either full or part time. These CMs are asked about their net (i.e. take-home) income in the last 12 months, that is, the amount they personally took out of the business after all taxes and costs. We convert these values to weekly income, by dividing them by 52 – the same procedure as used in the automatic conversion described above. This information is used in the first instance, and where not available estimates obtained from unfolding brackets are used instead.

Other economic activity

CMs whose main activity is unemployed and seeking work, in education, on a government training scheme, temporarily or long-term sick, looking after family, or retired are coded as "other" and allocated earnings of zero. CMs who do not fall into any of the above categories are coded as "something else" and in these cases follow-up questions are asked. Although the follow up answers are not provided for around 50% of these cases, the inspection of these answers indicates that this group is unlikely to be in paid work[12], and therefore they are also allocated earnings of zero. The earnings of CMs who do not know their economic activity or refuse to answer are coded as missing.

In several cases, the values of the main respondent's derived earnings variable were extremely high or extremely low. Therefore, the values above the 99th percentile of the distribution are replaced with the value of the 99th percentile. Similarly, and values below 1st percentile of the distribution are replaced with the value of the 1st percentile. Those who did not report to be employed, and therefore whose earnings are estimated as zero, were not included in the computation of these percentiles. This procedure is referred to as top and bottom coding thereafter.

---

[12] the most common answers include awaiting to start employment or on career break/garden leave, caring for family member or doing voluntary work

Earnings from second job

Questions about second job are asked to those in paid work. Figure 4 shows how the sample the sample in the respective groups was broken down by how the estimates for earnings from second job were derived.
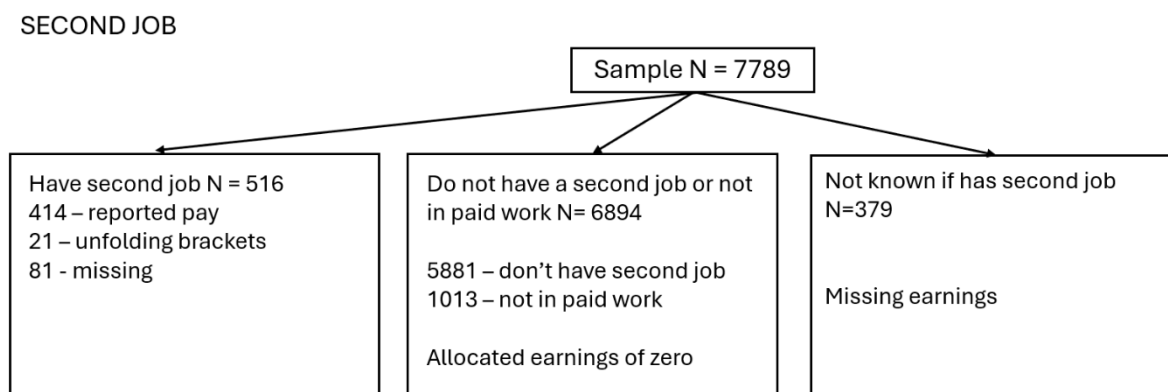
SECOND JOB

```
                        ┌─────────────────────┐
                        │  Sample N = 7789    │
                        └─────────────────────┘
```

Have second job N = 516	Do not have a second job or not in paid work N= 6894	Not known if has second job N=379
414 – reported pay		
21 – unfolding brackets	5881 – don't have second job	Missing earnings
81 - missing	1013 – not in paid work	
	Allocated earnings of zero	

Figure 4: Sample sizes for earnings from second job

The reported weekly earnings are used in first instance, where available. If missing, estimates obtained from unfolding brackets are used. Those who report not to have a second job or are not in paid work and were therefore not asked the questions are allocated earnings of zero. Cases where we do not know whether CM has a second job, or they do not report the amount earned in their second job, are treated as missing. The reported values are also top and bottom coded as described above.

Income from occasional work

All CMs are asked if they receive regular income from other regular jobs. Figure 5 shows how the sample in the respective groups was broken down by how the estimates for earnings from occasional work were derived.
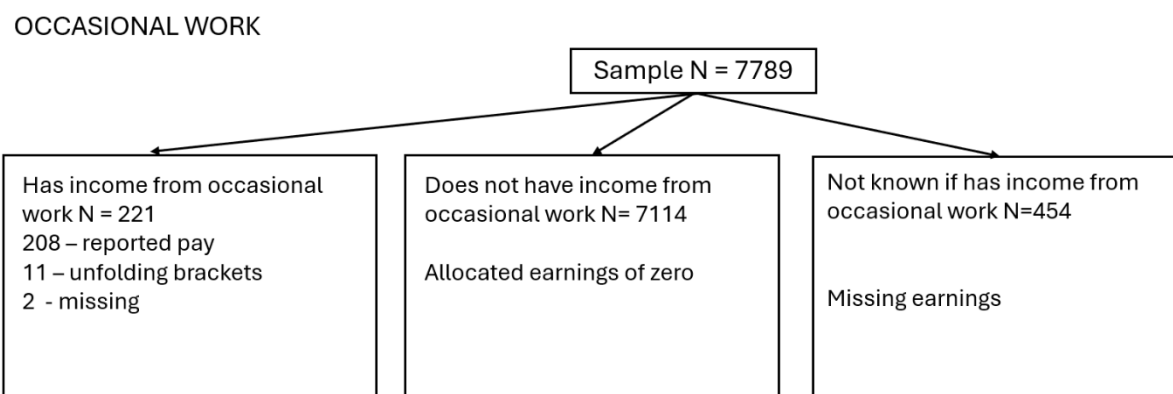
OCCASIONAL WORK

```
                        ┌─────────────────────┐
                        │  Sample N = 7789    │
                        └─────────────────────┘
```

Has income from occasional work N = 221	Does not have income from occasional work N= 7114	Not known if has income from occasional work N=454
208 – reported pay		
11 – unfolding brackets	Allocated earnings of zero	
2 - missing		Missing earnings

Figure 5: Sample sizes for earnings from other occasional work

The values of this income are converted to weekly and used where available and if missing unfolding brackets are used. Those who report not to have any income from occasional work are allocated earnings of zero and in cases where we do not know whether CMs have other regular work, the income is left as missing. These values are also top and bottom coded as described above.

A2.2.2 CM's partner's net earnings

All CMs who have a cohabiting partner are also asked about their partner's economic activity and, if they are employed or self-employed are asked to report their partner's earnings. Figure 6 shows the sample in the respective groups, broken down by how the estimates for partner's earnings were derived.
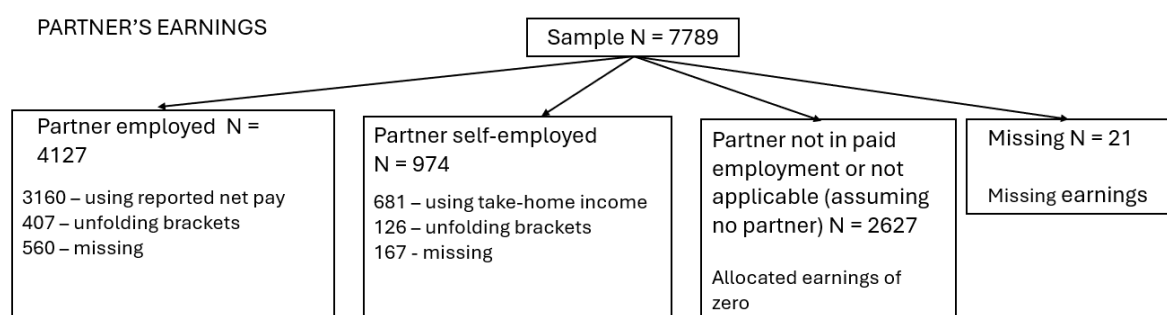
PARTNER'S EARNINGS

Sample N = 7789

Partner employed N = 4127	Partner self-employed N = 974	Partner not in paid employment or not applicable (assuming no partner) N = 2627	Missing N = 21
3160 – using reported net pay 407 – unfolding brackets 560 – missing	681 – using take-home income 126 – unfolding brackets 167 - missing	Allocated earnings of zero	Missing earnings

Figure 6: Sample sizes for partner's earnings

We use reported values in the first instance and, if these are not available, unfolding brackets. Partner's earnings for those who do not live with a partner or whose partner is not in paid work are coded to zero. Cases where we do not know if the partner works are coded as missing. Reported values of partner's earnings are also top and bottom coded as described above.

A2.2.3 Benefit income

All CMs were asked if they receive universal credit[13] or any other types of payment, such as jobseeker's allowance, income support, sickness, disability or incapacity benefits, pension benefits, child benefit, tax credits, housing benefit or council tax reduction, carer's allowance, income from a coronavirus state benefit or any other

---

[13] a new type of benefit which is being introduced in some parts of the country and replaced a number of other benefits

benefit. For each of these types of benefit there are follow up questions breaking these down to more specific types (see questionnaire for more details). For each benefit type mentioned, CMs are asked what the last payment amount was they received and what period this payment covered.

Based on the answers CMs provided to these questions, we construct 15 variables corresponding to the most commonly received payments and one that encompasses benefits not specified and received by a small number of people. These variables take a value of: a) the reported weekly amount received from the given type of benefit, either as reported or using information obtained from unfolding brackets; b) zero, in cases where the given type of benefit is not mentioned or: c) missing, under the following conditions - if the amount received is not provided or the question was not asked in a survey stage, CM refused to answer, or doesn't know if they receive a specific benefit. Each variable is top and bottom coded as described above. Table 14 provides the number of people in each of these categories.

The benefits which were only been received by a small number of people are included in the "other state benefits" category. Those who do not mention any of these benefits are allocated a value of zero. Those who mention the value of at least one, and are not missing on any others, are allocated a total obtained by summing the reported values. For those who mention more than one of these benefits, but where at least one of these values is missing, the total would be underestimated. Therefore, in these cases, we rely on the rich imputation to give a better estimate (see section A2.3 for more details).

Table 14: Sample sizes for each type of benefit received

Benefit type	N mention benefit type	N amount received provided	N benefit type not mentioned (allocated value of zero)	N missing
Universal credit	363	328	6916	545
Jobseekers allowance	30	27	7293	469
Income support	55	48	7268	473
Incapacity benefit	23	17	7300	472
Employment and support allowance	188/189	169	7134	486
Severe disablement allowance	21	15	7302	472

Personal Independence payment	354	322	6969	498
Carer's Allowance	200	183	7131	484
Disability Living Allowance	101	84	7230	484
NI pension or State retirement pension	37	29	7286	474
Child benefit	1743	1655	5580	554
Working tax credit	137	126	7183	480
Child tax credit	178	137	7142	510
Housing benefit	188	147	7135	507
Council Tax Reduction	295	215	7027	547
Other State benefit[14]	NA	35	7278	511

A2.2.4 Other regular income

All CMs were asked if they currently receive a regular payment from any of the sources, such as those listed in Table 15. If CM reported to receive income from any of these sources, they were subsequently asked about the amount received in the last month, as in the case of benefit income.

Based on the answers CMs provided to these questions, we construct twelve variables which take a value of: a) the reported weekly amount received from the given source; b) zero, in cases where the given source is not mentioned by the CM or c) missing, if the amount received is not provided, the questions were not been asked in a survey stage, CM refused to answer or doesn't know if they received income from the given source. When missing unfolding brackets are used where available. Each variable is top and bottom coded as described above. Table 15 provides the sample sizes for each of these categories and each type of income source mentioned.

Table 15: Sample sizes for each type of regular income received

Other regular income type	N mention income type	N amount received provided	N income type not mentioned	N missing

---

[14] Includes return to work allowance, attendance allowance, industry injury disablement benefit, war disablement benefit, sickness and accident insurance, pension credit (including guarantee credit & savings credit), widow's and war widow's pension, widowed parent allowance, foster allowance, lone parents in work credit as well as other family related and not specified benefits.

			(allocated value of zero)	
Income from investment	1151	1059	6185	545
Education grants/studentships including training or government training scheme	27	26	7310	454
Rent from boarders, lodgers or sub-tenants/rent from other property	920	901	6416	472
Allowance for a foster child	36	33	7300	556
Regular cash help from parents	102	101	7234	454
Regular cash help from children	137	133	7199	457
Regular cash help from other relatives or friends outside the household	44	41	7292	456
Income received from a pension scheme transferred to you from another individual such as a previous spouse, partner or parent	41	40	7295	454
Other income from organisations or other persons outside the household	73	72	7263	454
Child maintenance	36	35	7300	454
Ex-partner maintenance	11	11	7325	453
Other specific answer not in code frame or vague or irrelevant answers	13	13	7323	453

A2.3 Missing data strategy

A dataset with variables described above was created, which also included additional auxiliary variables including part-time and full-time indicator for current CMs employment and the current employment of their partner, if applicable; four indicators for full-time employment, part-time employment, full-time self-employment and part-time self-employment from the two previous sweeps of data collection; and a set of indicators known to predict non-response (see table 2 of UCL-Cohorts-COVID-19-Survey-user-guide.pdf). The sample was restricted to the N = 7789 productive interviews from GB residents and missing data in in these variables were imputed using multiple imputations with chained equations. All components of income were imputed using predictive mean matching with five nearest neighbours, while auxiliary variables were imputed using linear regression. We imputed five

datasets, summed the components of income imputed for each dataset and averaged the income across the five imputations.

81.26% of observation in our sample have complete information on all components of their income, but 10.01% have one of the components missing and 8.73% have more than one component missing, with large proportion of those observations (66.18%) being the participants of mop-up survey. Thus, 18.74% of observations rely of the richness of the BCS70 data included in the imputation to estimate at least one of these components, which are then averaged across imputations. While this approach is consistent with Rubin's standard rules in terms of estimating the average income of a given cohort member and therefore their rank on the income distribution, the uncertainly about these estimates is not taken into consideration in the derived variable. Therefore, we recommend that researchers intending to use this variable omit the cases where income is imputed when using the variable as continuous and perform their own analysis-specific imputation which can also be congenial with their substantive model of interest, or categorise the continuous income variable according to a metric most suitable for their research (e.g. quintiles, households below or above the poverty line etc.). A flag variable indicating whether none, one, or more than one component(s) of income were missing in the reported data is provided in the dataset.

Appendix 3: Response Models

Table A3: Response Models by Interview Type. Logistic regression models based upon twenty multiply imputed datasets pooled using Rubin's (1987) rules.

	Variable	Main Interview	Self-Completion	Diet Questionnaire
Sex (Ref: Male)	Female	0.8 (0.72, 0.88)	0.99 (0.9, 1.09)	0.95 (0.85, 1.05)
Highest Education (Ref: None)	Level 1	0.86 (0.71, 1.04)	0.95 (0.78, 1.15)	1.01 (0.82, 1.25)
	Level 2	0.68 (0.58, 0.8)	0.76 (0.65, 0.89)	0.91 (0.77, 1.08)
	Level 3	0.64 (0.53, 0.77)	0.76 (0.64, 0.91)	0.87 (0.72, 1.06)
	Level 4	0.85 (0.72, 1)	0.82 (0.7, 0.96)	0.91 (0.77, 1.08)
	Level 5	1.02 (0.79, 1.33)	0.93 (0.74, 1.17)	1.08 (0.86, 1.37)
	Number of Rooms @ Age 0y	1.03 (0.98, 1.09)	1 (0.96, 1.05)	0.98 (0.93, 1.02)
Social Class @ Age 10y (Ref: Professional or Managerial)	Intermediate	0.95 (0.84, 1.09)	0.96 (0.86, 1.09)	0.98 (0.87, 1.1)
	Partly-Skilled or Unskilled	1.02 (0.87, 1.19)	0.97 (0.84, 1.13)	0.98 (0.84, 1.16)
	Cognitive Ability @ Age 10y	1.1 (1.02, 1.19)	1.13 (1.06, 1.2)	1.18 (1.1, 1.26)
	Malaise @ Age 16y	1.02 (1, 1.03)	1.01 (1, 1.03)	1 (0.98, 1.02)
Vote @ Age 42y (Ref: Didn't vote)	Voted	1.28 (1.14, 1.43)	1.27 (1.11, 1.45)	1.33 (1.17, 1.5)
	Number of Organisations @ Age 42y	1.06 (1.01, 1.12)	1.04 (0.99, 1.09)	1.05 (1.01, 1.1)
Activity @ Age 46y (Ref: Employed)	Not Employed	0.82 (0.65, 1.04)	0.94 (0.79, 1.13)	0.92 (0.77, 1.11)
Household Income @ Age 46y (Ref: Quintile 1)	Quintile 2	1.09 (0.91, 1.31)	1.07 (0.91, 1.27)	1.06 (0.89, 1.25)
	Quintile 3	1.07 (0.86, 1.32)	1.09 (0.92, 1.29)	1.14 (0.94, 1.37)
	Quintile 4	1.07 (0.88, 1.3)	1.16 (0.97, 1.4)	1.17 (0.97, 1.43)

Variable	Main Interview	Self-Completion	Diet Questionnaire
Quintile 5	1.33 (1.06, 1.69)	1.27 (1.06, 1.53)	1.17 (0.97, 1.41)
Internet Use @ Age 46y (Ref: None/Little) — Medium	0.93 (0.8, 1.09)	0.9 (0.79, 1.02)	1.03 (0.89, 1.18)
Internet Use @ Age 46y (Ref: None/Little) — Lots	0.93 (0.8, 1.08)	0.89 (0.79, 1.02)	1.11 (0.96, 1.27)
Marital Status @ Age 46y (Ref: Never Married) — Married	1.22 (1.06, 1.4)	1.22 (1.07, 1.39)	1.05 (0.93, 1.2)
Marital Status @ Age 46y (Ref: Never Married) — Separated / Widowed	1.41 (1.21, 1.65)	1.16 (0.99, 1.37)	1.15 (0.98, 1.35)
Social Visits Freq. @ Age 46y (Ref: Very Frequently) — Fairly Frequently	1.07 (0.94, 1.2)	1.01 (0.91, 1.13)	1.01 (0.9, 1.13)
Social Visits Freq. @ Age 46y (Ref: Very Frequently) — Never / Rarerly	0.93 (0.8, 1.08)	1.04 (0.92, 1.18)	0.97 (0.84, 1.11)
Social Support @ Age 46y (Ref: A Little/ Not at All) — Somewhat	1.45 (1.14, 1.84)	1.48 (1.18, 1.85)	1.27 (0.99, 1.63)
Social Support @ Age 46y (Ref: A Little/ Not at All) — A Great Deal	1.36 (1.09, 1.69)	1.35 (1.1, 1.66)	1.34 (1.08, 1.67)
Malaise @ Age 46y	0.99 (0.96, 1.03)	0.98 (0.95, 1.01)	0.98 (0.95, 1.01)
BMI @ Age 46y	1.01 (1, 1.02)	1 (0.99, 1.01)	0.98 (0.97, 0.99)
Smoking Status @ Age 46y (Ref: Never Smoked) — Ex-Smoker	0.91 (0.8, 1.03)	0.91 (0.81, 1.01)	0.88 (0.78, 0.99)
Smoking Status @ Age 46y (Ref: Never Smoked) — Current Smoker	0.95 (0.82, 1.08)	0.84 (0.74, 0.96)	0.84 (0.73, 0.97)
Self-rated Health @ Age 46y (Ref: Excellent / Very Good) — Good	0.94 (0.83, 1.07)	0.91 (0.81, 1.02)	0.96 (0.85, 1.08)
Self-rated Health @ Age 46y (Ref: Excellent / Very Good) — Fair / Poor	0.9 (0.77, 1.06)	0.86 (0.74, 1)	0.96 (0.82, 1.13)
Consent to Biological Samples @ Age 46y (Ref: Consented) — Did Not Consent	0.67 (0.56, 0.82)	0.74 (0.62, 0.89)	0.85 (0.69, 1.04)
Consent to Diet Diary @ Age 46y (Ref: Completed) — Did Not Complete			0.25 (0.23, 0.28)
# Main Sweeps Non-Response	0.63 (0.62, 0.65)	0.66 (0.64, 0.67)	0.78 (0.76, 0.8)
# COVID-19 Sweeps Non-Response	0.31 (0.3, 0.33)	0.46 (0.44, 0.48)	0.52 (0.5, 0.54)

Appendix 4: Weights Performance

Figure A4.1: Sample proportions for categorical predictors included in response model, by variable category, population (target or respondents) and whether sample is weighted or unweighted.
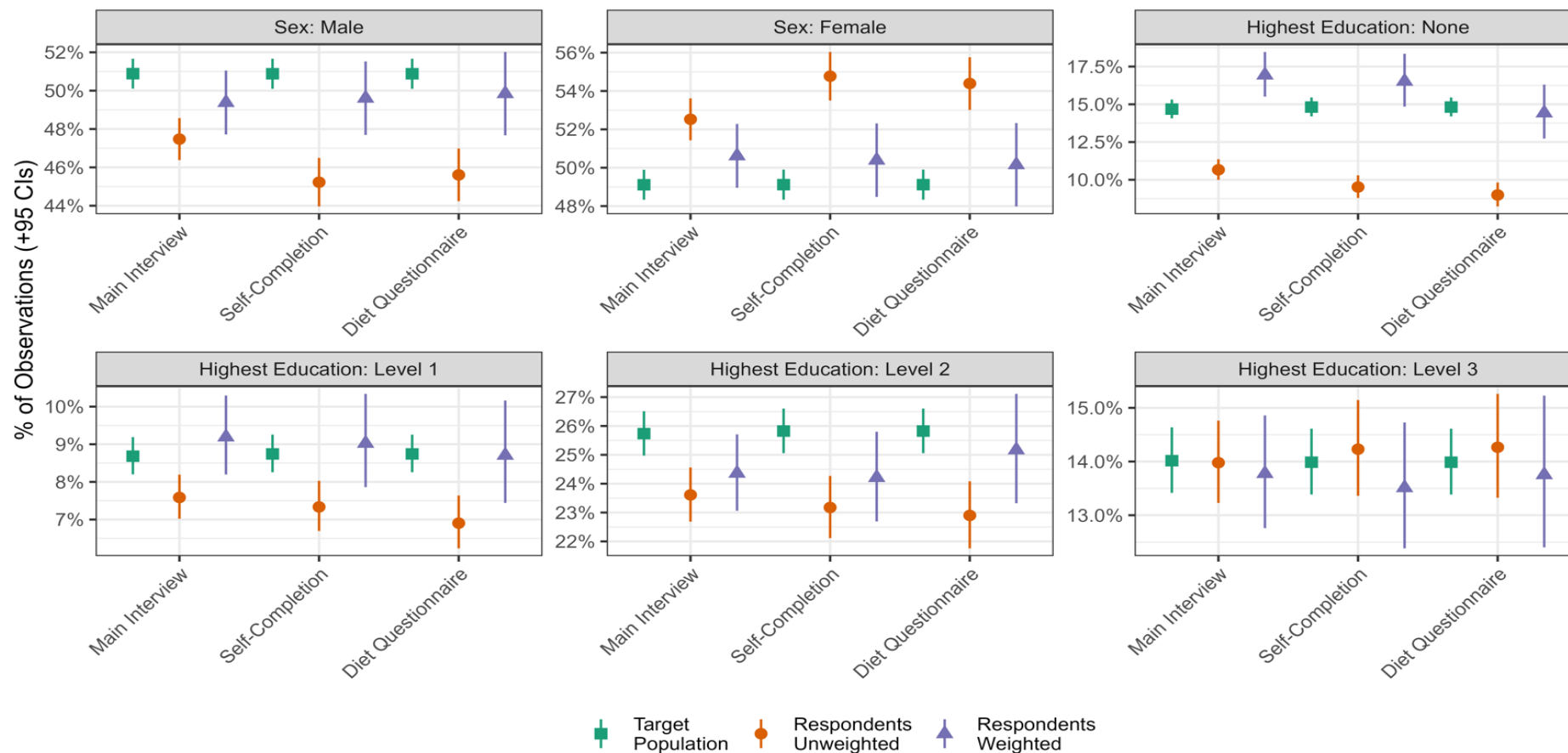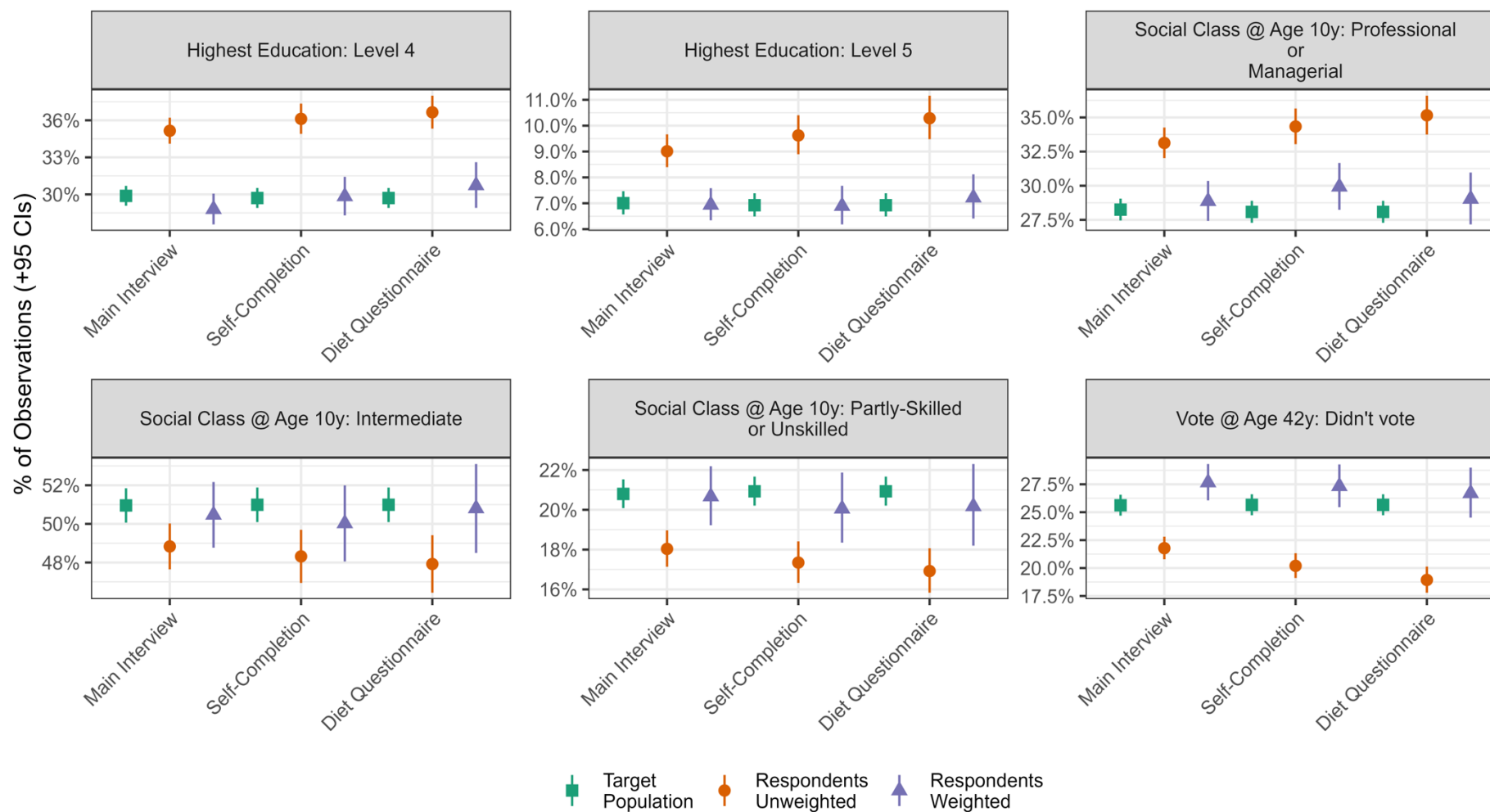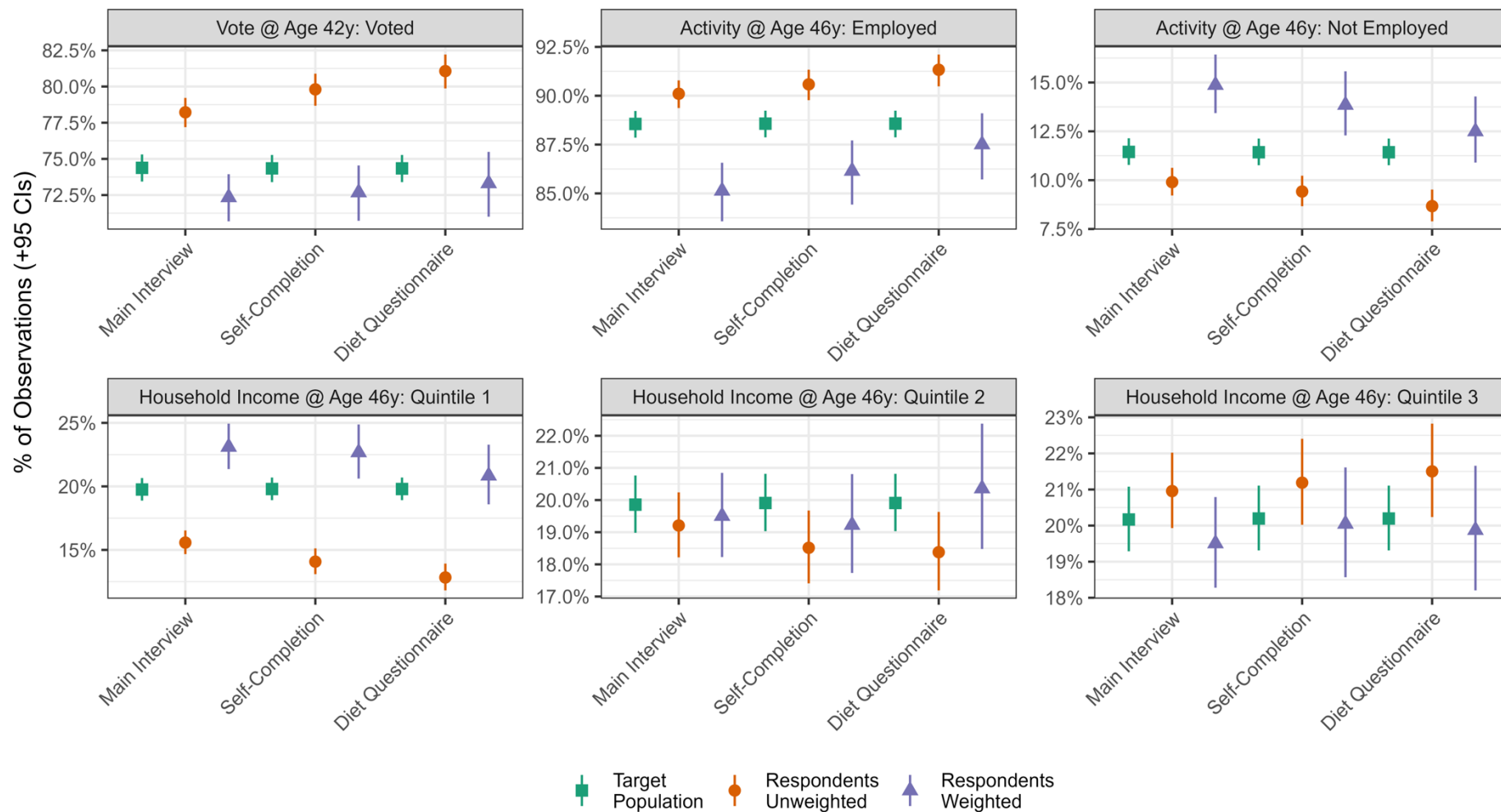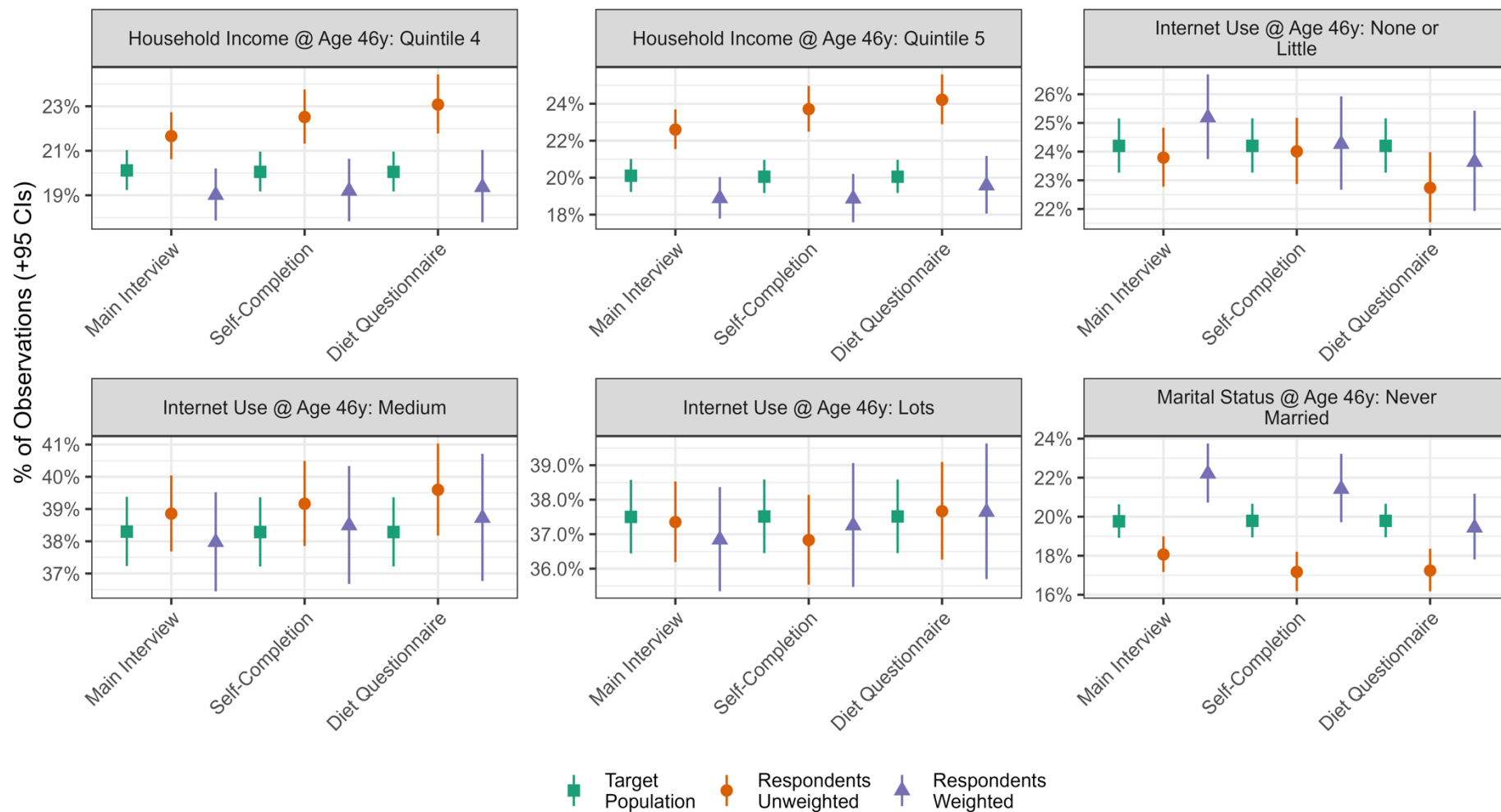
Figure A4.2: Sample proportions for categorical predictors included in response model, by variable category, population (target or respondents) and whether sample is weighted or unweighted.

Figure A4.3: Sample proportions for categorical predictors included in response model, by variable category, population (target or respondents) and whether sample is weighted or unweighted

Figure A4.4: Sample proportions for categorical predictors included in response model, by variable category, population (target or respondents) and whether sample is weighted or unweighted.
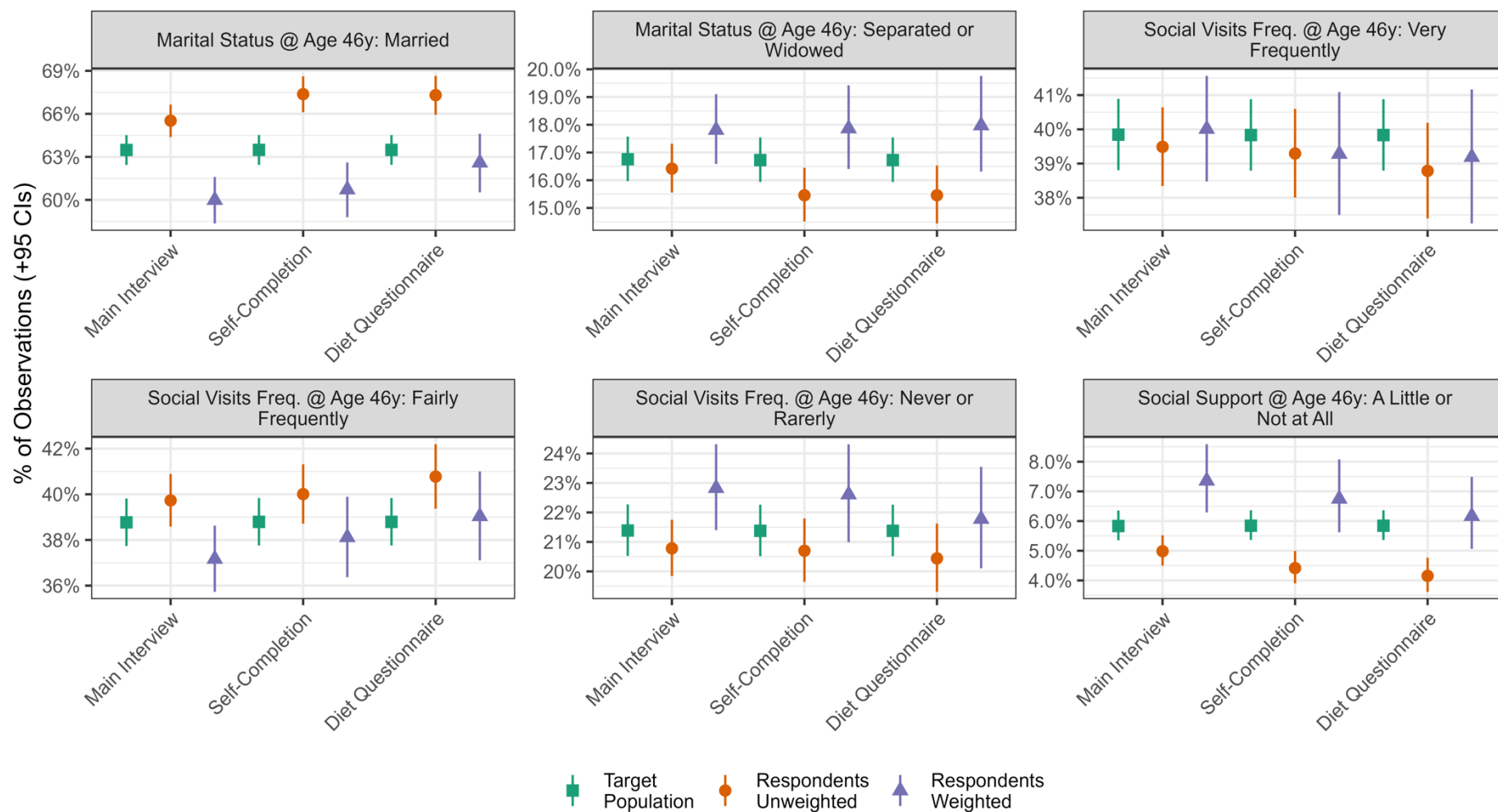
Figure A4.5: Sample proportions for categorical predictors included in response model, by variable category, population (target or respondents) and whether sample is weighted or unweighted.
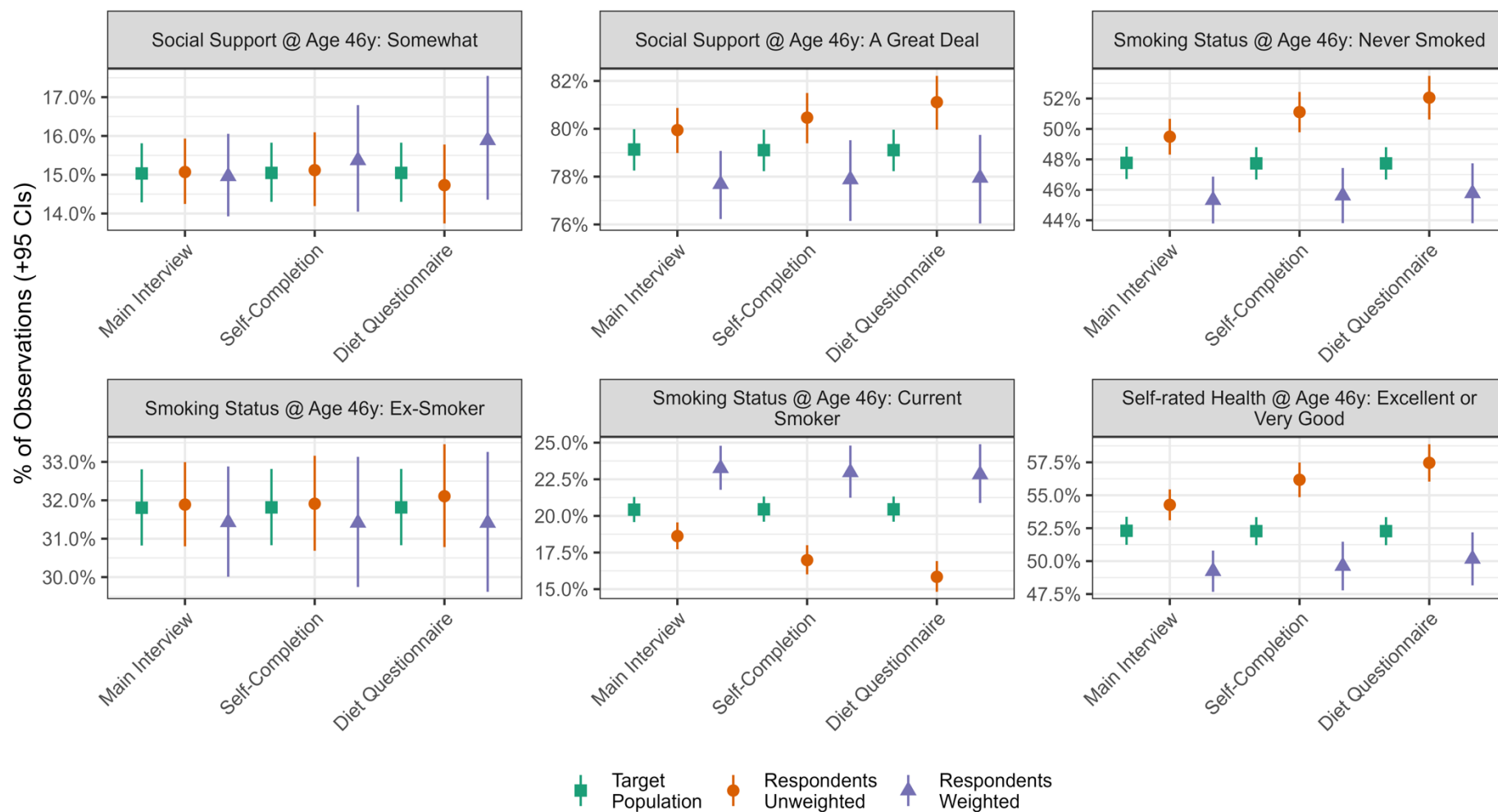
Figure A4.6: Sample proportions for categorical predictors included in response model, by variable category, population (target or respondents) and whether sample is weighted or unweighted.

Figure A4.7: Sample proportions for categorical predictors included in response model, by variable category, population (target or respondents) and whether sample is weighted or unweighted.
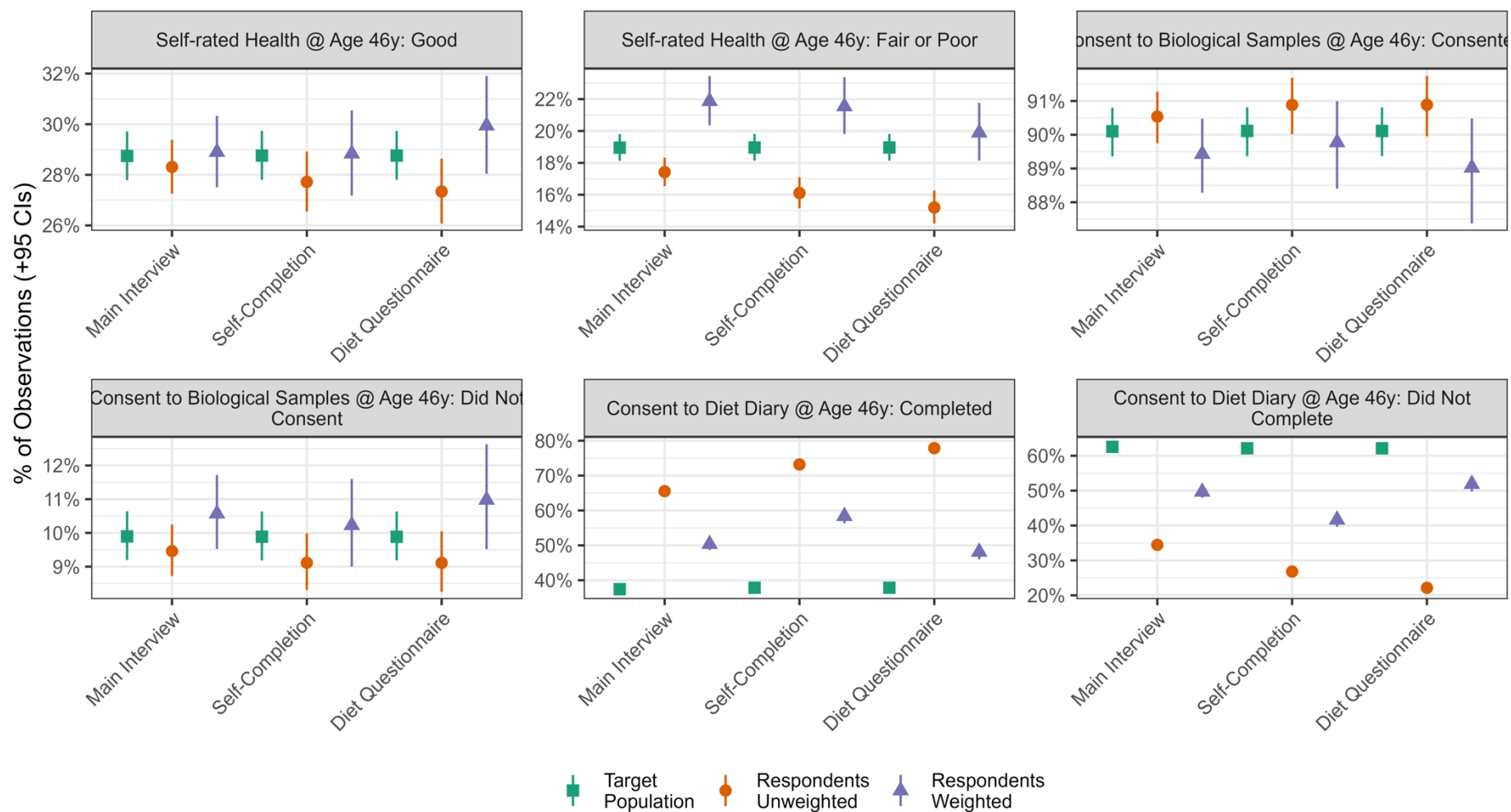
Figure A4.7: Sample means for continuous predictors included in response model, by variable, population (target or respondents) and whether sample is weighted or unweighted.