

Millennium Cohort Study:

Sweep 8

Age 23 Survey

User Guide (Version 1)

February 2026

CENTRE FOR
LONGITUDINAL
STUDIES



Economic
and Social
Research Council

Contact

Data queries: help@ukdataservice.ac.uk

Questions and feedback about this user guide: clsdata@ucl.ac.uk.

Authors

Matt Brown, Emla Fitzsimons, Lucy Haselden, Morag Henderson, Sarah Kerry-Barnard, Andrew Peters, Rachel Rosenberg, Aida Sanchez-Galvez, Liam Wright.

How to cite this guide

Brown M., Fitzsimons E., Haselden L., Henderson M., Kerry-Barnard S., Peters A., Rosenberg R., Sanchez-Galvez A., Wright L. (2026) *Millennium Cohort Study Sweep 8: Age 23 Survey User Guide (Version 1)*. UCL Centre for Longitudinal Studies.

Data citation and acknowledgement

You should cite the data and acknowledge CLS following the [guidance from Citing our data webpage on the CLS website](#).

This version of the guide was published in February 2026 by the UCL Centre for Longitudinal Studies (CLS).

Centre for Longitudinal Studies

Centre for Longitudinal Studies (CLS)

UCL Social Research Institute

University College London

20 Bedford Way, London WC1H 0AL

www.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 [UCL Institute of Education](#).

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

Contents

About the Millennium Cohort Study	6
1. Introduction	7
2. The Age 23 Survey	10
3. Fieldwork	11
3.1 Issued sample.....	11
3.2 Fieldwork period.....	11
3.3 Fieldwork stages	11
3.4 Contact strategy.....	12
3.5 Pilot.....	13
3.6 Incentives.....	13
4. Response at Age 23	14
4.1 Overall response and outcome codes.....	14
4.2. Response by stage	15
4.3. Mode of completion.....	16
5. Questionnaire	18
5.1 Overview	18
5.2 Changes following the soft launch	30
5.3 Data linkage	31
5.4 Other special features	32
5.5 Scales	40
6. Research Survey Data	52
6.1 Licensing and data access.....	52
6.2 Datasets and data structure	53

6.3 Future datasets	54
6.4 Data documentation	55
6.5 Identifiers	56
6.6 Variable description	59
6.7 Derived variables	61
6.8 Income and payment unfolding brackets.....	63
6.9 Household grid.....	64
6.10 Longitudinal family and cohort member files	64
6.11 Missing values	65
6.12 Data errors and inconsistencies	66
6.13 Data de-identification	67
6.14 Output disclosure control	68
7. Response patterns and weights	69
7.1 Response patterns.....	69
7.2 Predicting response in Age 23 and deriving non-response weights.....	72
8. Mode Effects.....	81
9. References.....	83
Appendix 1: Derived Variables Guide.....	85
Appendix 2: Response Models	108
Appendix 3: Performance of Non-Response Weights	111

A tribute to Kate Smith (CLS Survey Manager)

CLS would like to pay tribute to Kate Smith who sadly passed away on 2 September 2023 while working on the eighth sweep of the Millennium Cohort Study. Kate was the Centre's longest serving member of staff and devoted her highly successful career to the development of longitudinal cohort studies, and in particular to MCS. Kate was pivotal in ensuring the success of the study from the very first sweep. Renowned for her intellect, kindness, and collaborative spirit, Kate is deeply missed by all who had the pleasure of working with her.

About the Millennium Cohort Study

The Millennium Cohort Study or MCS (also known as Child of the New Century) to participants is a longitudinal cohort study, following a nationally representative group of over 19,000 people born in the UK between September 2000 and January 2002.

Over the years, the study has built a uniquely rich record of cohort members' lives - from their early health and development, through childhood, into adolescence and now into adulthood. This long-term evidence allows researchers to understand how experiences at different stages of life shape later outcomes.

As a multidisciplinary study, MCS is used by researchers working in a wide range of fields. Findings from MCS have influenced policy at the highest level, and today the study remains a vital source of evidence on the major issues affecting young people's lives.

1. Introduction

The Millennium Cohort Study Age 23 Sweep (Sweep 8) took place between October 2023 and February 2025. It is designed and managed by the UCL Centre for Longitudinal Studies (CLS) at the UCL Institute of Education (IOE), and fieldwork was carried out by Ipsos. It was funded by the Economic and Social Research Council and a consortium of government departments including the Department of Health and Social Care, Department for Education, Department for Work and Pensions, Home Office, Welsh Government, the Department for the Economy (Northern Ireland) and the Department of Education (Northern Ireland). The Age 23 Sweep is the eighth sweep of the study. Previous sweeps were conducted when the cohort members were aged 9 months then at 3, 5, 7, 11, 14 and 17, all face-to-face. In addition, three online¹ surveys were conducted during the COVID-19 pandemic, when data was also collected simultaneously from participants in the MRC National Survey of Health and Development, 1970 British Cohort Study, 1958 National Child Development Study, and Next Steps.

The majority of fieldwork was conducted using a sequential mixed mode approach in which participants were first invited to take part online. After 3 weeks, non-responders were then contacted by face-to-face interviewers who could carry out the interview face-to-face or 'nudge' cohort members to complete the survey online. The mixed mode approach was introduced after an experiment in the first wave of fieldwork (the 'soft launch') demonstrated that this approach was more successful at maximising response than a face-to-face only design. At the end of fieldwork there was also an online 'mop-up' survey in which non-respondents were asked to complete an abbreviated version of the survey.

Interviews were completed with 9,735 cohort members. Of these 8,900 were fully completed interviews and the remaining 835 were partially completed (including 455

¹ While the first two waves were fully online, there was telephone follow up in Wave 3. See the COVID-19 Survey in Five National Longitudinal Studies Waves 1, 2 and 3 User Guide (Version 4) for details.

who completed the online 'mop up' survey). The overall response rate for the study is 72%.

The deposited safeguarded dataset contains data for 9,675 participants. The 60 participants whose data are not included in the deposit are: one participant who requested their data be deleted after taking part, 42 where a 'proxy interview' was completed (a very short version of the questionnaire completed by a proxy on behalf of the cohort member) as well as 17 cohort members who are triplets. Data from triplets are excluded from the safeguarded datasets (End User Licence) to minimise disclosure risk, but they are made available as controlled data (Secure Access). Proxy data are not included in the research datasets due to their scarcity. Please refer to section 6.1 for information on how to access the safeguarded and the controlled MCS8 data.

This User Guide provides information about the data arising from the Millennium Cohort Study Age 23 Sweep and accompanies the deposit of the data at the UK Data Service (UKDS). The data from the MCS Age 23 Sweep are large and complex and will therefore be deposited in two phases. This approach allows the majority of the data to be made available as quickly as possible in the first phase, while ensuring that more complex elements of the dataset undergo additional processing and quality checks. Further details of the content of each phase are provided in Section 6.3. An updated User Guide will accompany the Phase 2 deposit.

In addition to this user guide, the Age 23 Survey data documentation includes:

- Millennium Cohort Study Sweep 8 (Age 23) Household Grid And Proxy Questionnaire
- Millennium Cohort Study Sweep 8 (Age 23) Main Questionnaire
- Millennium Cohort Study Sweep 8 (Age 23) Resident Partner Questionnaire
- Millennium Cohort Study Sweep 8 (Age 23) Technical Report
- Millennium Cohort Study Sweep 8 (Age 23) Technical Report Appendices

Data, questionnaires and user guides for all previous sweeps are also available at UKDS. All MCS8 datasets use common ID, **MCSIDCMH**, which is a concatenation of the family ID (MCSID) and the cohort member number (CNUM). Please see section 6.5 for details.

Note that there is no separate questionnaire documentation for the mop up survey. Instead, questions included in the mop-up questionnaire are denoted by an asterisk beside the question name in the Millennium Cohort Study (Sweep 8) Age 23 Questionnaire.

Important note about figures in this document

Figures that are presented in this document may differ from totals in deposited datasets. This happens for various reasons including resolution of duplicate cases and availability of specific cases on the safeguarded data (End User Licence), such as data collected from triplet families being only available as controlled under Secure Access. The DATA_LICENCE variable in the mcs_longitudinal_family_file_1-7 (MCS 1-7) and on the mcs_longitudinal_cmh_file (MCS 1-8) indicates which cases are available in the safeguarded data, and DATA_AVAILABILITY indicates which cases are available for research use. These help users to estimate the final sample size that can be used for research purposes.

2. The Age 23 Survey

The Age 23 Survey aimed to provide data for research and policy on the lives of this generation of adults in their early twenties. Since the last sweep at Age 17, many cohort members will have experienced key transitions in education and training, work and careers, relationships, parenthood and housing. Yet pathways to adulthood for this generation are varied and often non-linear. Some move between education and employment several times and many continue living with their parents after finishing education. To reflect this diversity, the sweep collected detailed information on household composition, housing, education and employment. The Age 23 Survey also collected information on wider aspects of cohort members' lives, including health and wellbeing, identity and attitudes, political and social participation, personality and resilience. It also included cognitive assessments to measure working memory and selective attention.

Participants were asked for consent to link information from their health, economic, education and police and criminal justice records to their survey data (if they had not previously given consent at age 17). Those who were parents were asked an additional set of questions about their children, their approach to parenting and for consent to link information from their child's health and education records.

Study members who were living with a partner were asked for consent to contact their partners to ask them to complete a short web survey about their own lives.

The Age 23 Survey was conducted by Ipsos and funded by the Economic and Social Research Council (ESRC) and a consortium of government departments including the Department of Health and Social Care, Department for Education, Department for Work and Pensions, Home Office, Welsh Government, the Department for the Economy (Northern Ireland) and the Department of Education (Northern Ireland).

Ethical approval was provided by North West Haydock Ethics Committee. (REC reference: 23/NW/0213).

3. Fieldwork

3.1 Issued sample

A total of 13,713 study members were issued to the Age 23 Survey. The issued sample was comprised of all study members who had previously taken part in MCS other than 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.

3.2 Fieldwork period

The Millennium Cohort Study Age 23 Survey took place between 10 October 2023 to 23 February 2025.

3.3 Fieldwork stages

To make fieldwork more manageable, the issued sample was divided into three waves, released to the field in sequence. The first wave was designated as a 'soft launch' to test survey processes and run an experiment assessing the impact of a sequential web-first mixed mode approach compared with face-face only interviewing.

From an overall sample of 13,713 cohort members, Ipsos selected a stratified random sub-sample of 3,000 cases for the soft launch. Stratification was based on prior participation, sex and region. These 3,000 cases were then randomly allocated, using the same stratification variables, into two equally sized groups:

- **Group 1: Face-to-face data collection only:** 75-minute face-to-face interview including physical measurements (height and weight).
- **Group 2: Sequential mixed mode:** Participants first invited to participate online with non-respondents followed up by face-to-face interviewers. No physical measurements were conducted with this group which reduced the length of face-to-face interviews to 65 minutes.

The sequential mixed-mode arm achieved a significantly higher response rate than the face-to-face only arm (66.8% vs 53.5%). As a result, the decision was taken to adopt mixed-mode data collection for the remaining two stages of fieldwork.

The soft launch also provided an opportunity to test the questionnaire and fieldwork protocols. Minor refinements were made following this phase, and any changes to questionnaire content are documented in the questionnaire materials.

Throughout fieldwork, where appropriate, 'soft' refusals and non-contacts, were later reissued to another interviewer to allow them to try again.

Finally, participants who had not taken part after the main fieldwork period, or who had only partially completed an interview (below a minimum threshold), were invited to take part in a shorter online-only "mop-up" survey. This shorter "mop-up" version, which took around 20 minutes to complete, focused on core measures of current circumstances and did not include the cognitive assessments, data linkage consents or partner components included in the main survey. Study members were offered the same incentive (£30) for completing the mop-up as for the main survey. The mop-up survey ran from 27 January to 23 February 2025

3.4 Contact strategy

The Millennium Cohort Study makes continual efforts to maintain contact with study members, including regular mailings between sweeps and encouraging participants to notify the study team of any change of address.

At the start of each wave study members were sent an invitation letter by both email and post which explained the purpose of the survey, what participation would involve, and (for those invited to take part online) provided a link and login details for completing the online questionnaire. Where appropriate, reminders to complete the web survey were sent by text and email. For further details see the Millennium Cohort Study Sweep 8 (Age 23) Survey Technical Report. Three weeks after the launch of web fieldwork, non-respondents were issued to interviewers, who attempted contact by telephone or by personal visit. Interviewers could either complete the interview in-person or encourage study members to participate online. The online survey remained open throughout the full main stage fieldwork period.

When interviewers established that a study member had moved, they undertook field tracing, primarily by contacting parents, stable contacts or speaking with current occupiers (see Section 7.4 of the Millennium Cohort Study Sweep 8 (Age 23) Survey Technical Report). Study members that interviewers were unable to trace were returned to CLS for office-based tracing.

Following the face-to-face fieldwork period, study members who had not completed the main survey were invited to complete an abbreviated online questionnaire to maximise participation (described above).

3.5 Pilot

Main stage fieldwork was preceded by a small-scale pilot study, conducted between 11th and 30th July 2023. The pilot aimed to test the questionnaire routing and survey protocols with individuals who were not study members but were a similar age. As the pilot did not involve Millennium Cohort Study participants the data collected are not included in the deposited dataset. Full details of the pilot are provided in Section 3.3 of the Millennium Cohort Study Sweep 8 (Age 23) Survey Technical Report.

3.6 Incentives

For the first time in the study, all study members (including those who took part in the shorter online mop-up survey) were offered a £30 voucher that could be used in many high street and online stores as a thank-you for completing the survey. Partners who completed the shorter partner questionnaire were given a £10 voucher.

4. Response at Age 23

4.1 Overall response and outcome codes

A total of 13,713 study members were issued to the Age 23 Survey, of whom 9,735 took part. Of those who participated, 8,900 interviews were classified as 'fully productive' and 835 as 'partially productive'.

Interviews were defined as fully productive if the participant completed all main elements of the questionnaire. Interviews were classified as partially productive if the participant completed the household relationships section of the main questionnaire, or at least the household relationships section of the mop-up survey.

Of the interviews classified as fully productive, 42 were proxy interviews conducted with someone other than the cohort member. Proxy interviews were conducted where the cohort member was unable to understand the questions or respond independently. These proxy interviews were very short and the data they generated are not included in the available research data.

Table 1 below shows the overall response rate for the cases issued to fieldwork. The overall response rate was 71.5%. The main reason for non-response was refusal to the survey (14% of the total issued sample).

Full breakdowns of response rates including by country, prior participation and original sampling stratum are provided in the Millennium Cohort Study Sweep 8 (Age 23) Survey Technical Report (Chapter 10).

Table 1. MCS Age 23 Survey – Overall Response Rate

Outcome	Frequency	Percent (of eligible)
Productive	9,735	71.5%
<i>Fully productive (main interview)</i>	<i>8,900</i>	<i>65.3%</i>
<i>Partially productive (main interview only)</i>	<i>338</i>	<i>2.5%</i>
<i>Fully productive (mop-up)</i>	<i>428</i>	<i>3.1%</i>
<i>Partially productive (mop-up)</i>	<i>27</i>	<i>0.2%</i>
Refusal	1,963	14.4%
Non-contact	988	7.2%
Untraced	885	6.5%
Other unproductive	35	0.3%
Total eligible sample	13,606	100.0%
Ineligible (deceased, in prison/probation, or no email)	107	—
Total issued sample	13,713	—

4.2. Response by stage

Soft launch

A total of 3,073 cases were issued to the soft launch. This included 3,000 cohort members randomly selected for the mode experiment, which sought to compare the impact of a sequential mixed-mode approach with face-to-face only interviewing. 73 additional participants were issued to the soft-launch who were twin or triplet siblings of those selected for the soft-launch – in order to increase fieldwork efficiency.

1,500 were randomised into each arm, and the experiment demonstrated a clear difference in response rates between the two arms: 53.5% of those issued to the

face-to-face arm completed the survey during the soft launch, compared with 66.8% in the sequential mixed-mode arm.

Once it was confirmed that sequential mixed-mode approach resulted in higher response rates, this approach was adopted for the remainder of fieldwork. Some cases originally allocated to the face-to-face arm were later reissued and given the opportunity to complete the survey online, and if still unproductive, were invited to complete the mop-up survey. The overall soft-launch response rate was 69.9%.

Main stage

During the main stage, the sequential mixed-mode approach was used for all cases. 6,455 cases were issued in Wave 2 and 4,185 cases were issued in Wave 3. Wave 2 achieved a final response rate of 69.3%, while Wave 3 achieved a final response rate of 74.3%. The lower response rate in Wave 2 resulted from the fact that it contained a larger proportion of cases who were expected to be more challenging to locate and persuade to take part based on prior participation history.

Online mop-up stage

A total of 3,203 non-respondents were issued to the online mop-up stage. The mop-up survey was issued to all non-responders except hard refusals and permanent withdrawals. It was also issued to partially productive cases from the soft launch and main stage who had not reached the start of the CASI section.

In total, 455 interviews were achieved at the mop-up stage, giving a response rate of 14.2%. The online mop-up increased the overall survey response rate by 3.3 percentage points.

4.3. Mode of completion

Web was the most common mode by far, with 78.8% completing the survey in this mode.

Table 2. Mode of response

	Number	Percentage
Online (total)	7,638	78.8%
<i>Online (main survey)</i>	7183	74.1%
<i>Online (mop-up)</i>	455	4.7%
Face-to-face	2,055	21.2%
Total productive (excluding proxies)	9,693	100.0%

5. Questionnaire

5.1 Overview

The Age 23 survey consisted of 21 modules or topic areas. A summary of the content of each module is provided below in Tables 3.1 – 3.21. Each table also shows which content was answered by self-completion during face-to-face interviews (usually because the questions were of a more sensitive nature) and which content was included in the abbreviated version of the questionnaire used during the final online mop-up.

The questionnaire took around 45 minutes on average to complete online and around 60 minutes when completed face-to-face.

Table 3.1 Questionnaire topics in module – Household grid

Method of administration in face-to-face interviews: Interviewer	
Topic	Included in mop-up?
Confirmation of cohort member name and address	Yes
Preferred pronouns	Yes
Type of residence	Yes
Nation living in	Yes
Whether still living with the same people as last time	Yes
Details of new people living with (including sex, age and relationship to study member)	Yes
Marital status and date of marriage	Yes
Resident and non-resident children	Yes

Table 3.2 Questionnaire topics in module – Housing

Method of administration in face-to-face interviews: Interviewer	
Topic	Included in mop-up?
Whether moved since previous interview	Yes
Type of accommodation	Yes
Number of rooms	Yes
Tenure	Yes
Home ownership (purchase price, year of purchase, amount borrowed, sources of finance)	No
Housing expenses (rent, mortgage payments, financial help)	No
Whether the accommodation has problematic damp	Yes
Whether accommodation has access to outdoor space	No
Periods of homelessness	No

Table 3.3 Questionnaire topics in module – Health

Method of administration in face-to-face interviews: Interviewer	
Topic	Included in mop-up?
General health	Yes
Limiting longstanding illness	Yes
Exercise	Yes
Long covid	No

Table 3.4 Questionnaire topics in module – Activity and Employment

Method of administration in face-to-face interviews: Interviewer	
Topic	Included in mop-up?
Current economic activity	Yes
Work hours (full or part time)	Yes
Study hours (full or part time)	Yes
Institution of study	Yes
Start date of current economic activity	Yes
Occupation details	Yes
Usual weekly working hours	Yes
Overtime	Yes
Shift work	No
Autonomy over working hours	No
Zero hours contract	No
Place of work	No
Type of contract (permanent or temporary)	No
Whether has a second job	No
Hours a week in second job	No
Whether currently looking for work	No

Table 3.5 Questionnaire topics in module – Education

Method of administration in face-to-face interviews: Interviewer	
Topic	Included in mop-up?
Whether has a baccalaureate type qualification (including type)	Yes
Baccalaureate details (subjects, grades/scores, institution, completion date, COVID impact)	No
School-level academic qualifications not previously reported (level, type, number)	Yes
School-level academic qualifications not previously reported (subjects, grades, dates, COVID impact)	No
Whether has post-secondary academic qualifications (and type)	Yes
Post-secondary academic qualification details (institution, subjects, grades, COVID impact)	No
Choice of institution (Bachelor's degree)	No
Sandwich courses and gap years	No
Vocational qualifications (type and number)	Yes
Vocational qualification details (subjects, grades, dates, institutions)	No
Academic qualifications currently being studied (type and number)	Yes
Academic qualifications currently being studied (subjects, institution, dates)	No
Vocational qualifications currently being studied (type, number, subjects, levels, institutions, dates)	No
Uncompleted courses (number, types, reasons)	No
Whether accrued educational debt (and amount)	Yes
Educational debt – who owed	No
Educational debt – future expectations	Yes

Table 3.6 Questionnaire topics in module – Income

Method of administration in face-to-face interviews: Interviewer	
Topic	Included in mop-up?
Pay (Gross, Net, Usual)	No
Self-employed take home income	No
Other income from work	No
Economic activity of partner	Yes
Partner's working hours	Yes
Partner's studying hours	Yes
Benefit receipt	Yes
Other sources of income	No
Total household income	Yes

Table 3.7 Questionnaire topics in module – Cognitive assessment

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Backward Digit Span	No

Table 3.8 Questionnaire topics in module – Identity and attitudes

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Gender identity	Yes
Sexual identity	Yes
Welsh language use	No
Attitude to government redistributing income	Yes
Attitude to democracy	No
Voting in general election	No
Attitude to environmental issues	Yes

Table 3.9 Questionnaire topics in module – Mental health part 1

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Kessler-6 Psychological Distress Scale	Yes
Life satisfaction	Yes
Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS)	No
Whether sought help for mental health problem past year	No
Whether received help for mental health problem past year	No
Whether currently receiving medication or counselling for mental health problem	No
ONS-4 Loneliness Scale, which includes the UCLA-3	No
Eating Disorder Examination Questionnaire Short (EDE-QS): 5 items	No
Short Grit Scale: 8 items	No

Table 3.10 Questionnaire topics in module – Health behaviours

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Sedentary activities	No
Sleep	No
Diet	No
Perception of weight	No
Smoking	Yes
Vaping	Yes
Alcohol – Alcohol Use Disorders Identification Test – Consumption (AUDIT-C), the 3-item short form of the WHO AUDIT	Yes
Drug taking	No
Puberty – age started puberty	No
Contraception	No
Sexually transmitted infections	No
Pregnancies (number, whether planned, outcome, dates)	No
Where baby is living, ethnicity and gender	No

Table 3.11 Questionnaire topics in module – Family and relationships

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Ever lived in care	No
Whether mother is still alive	No
Closeness to mother	No
Other maternal figure	No
Closeness to other maternal figure	No
Whether father is still alive	No
Closeness to father	No
Other paternal figure	No
Closeness to other paternal figure	No
Reason for living at home	No
Caring responsibilities	No
Whether in a non-cohabiting relationship	Yes
Sex of non-cohabiting partner	No
Satisfaction with relationship with non-cohabiting partner	Yes
Length of relationship with non-cohabiting partner	No
Experience of domestic violence	No
How managing financially	Yes

Table 3.12 Questionnaire topics in module – Cognitive assessment

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Stroop	No

Table 3.13 Questionnaire topics in module – Mental health part 2

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Self-harm: 6 items from the Edinburgh Study of Youth and Transitions	No
Suicide attempts	No
Social provisions: 4 items from the Social Provisions Scale	No
Generalised anxiety disorder scale (GAD-2)	No
Patient Health Questionnaire – 2 item (Depressive disorder)	No
The Positive Childhood Experiences (PCE) Questionnaire (7 items)	No
Body dysmorphia: 3 items from the Dysmorphic Concern Questionnaire (DCQ)	No

Table 3.14 Questionnaire topics in module – Financial literacy

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Financial literacy	No

Table 3.15 Questionnaire topics in module – Activities and risky behaviour

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Social media use and addiction	No
Victimhood	No
Gambling	No
Carrying a knife	No
Interaction with police	No
Interaction with courts and justice system	No
Reading for enjoyment	No
Spending time with friends	No

Table 3.16 Questionnaire topics in module – Future aspirations and living expenses

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Expected achievements by age 30	No
Expected career at age 30	No
Fertility plans	No
Received money to pay for living expenses in last month	Yes

Table 3.17 Questionnaire topics in module – Resident children

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Whether child born early, late or on time	No
Type of birth	No
Birth weight	No
Smoking, drinking and drug taking during pregnancy	No
Infant feeding	No
Milestones	No
Parental activities with child	No
Strength and Difficulties Questionnaire for children aged 2+	No
The Ages and Stages Questionnaire (ASQ): communication skills	No
Study member's relationship with child	No
Study member's feelings about caring for child	No
Study member's emotional regulation around child	No
Study member's views on parenting	No

Table 3.18 Questionnaire topics in module – Non-resident children

Method of administration in face-to-face interviews: self-completion	
Topic	Included in mop-up?
How often study member sees non-resident children	No
Nature of visits with non-resident children	No
Financial support for non-resident children	No

Table 3.19 Questionnaire topics in module – Child data linkage

Method of administration in face-to-face interviews: Self-completion	
Topic	Included in mop-up?
Request for consent for data linkage for health and education records of resident children	No

Table 3.20 Questionnaire topics in module – Physical measurements

Method of administration in face-to-face interviews: Interviewer administered in face-to-face arm of soft launch; self-reported by self-completion after soft launch	
Topic	Included in mop-up?
Height	Yes
Weight	Yes
Body fat (in face-to-face arm of soft launch experiment only)	No

Table 3.21 Questionnaire topics in module – Data linkage

Method of administration in face-to-face interviews: Interviewer	
Topic	Included in mop-up?
Those who had not previously consented asked for consent to link to: <ul style="list-style-type: none"> • health records • education records • economic records • police and criminal justice records. 	No

5.2 Changes following the soft launch

Following the soft launch, a small number of amendments were made to the instrument. Most of these related to routing and validation checks, but several substantive changes were also introduced:

- **Removal of physical measurements** previously carried out by interviewers, replaced with self-reported measures.
- **Addition of new questions, including:**
 - **MAN** – whether the respondent has managerial duties at work
 - **EDUAFQ** – full details of all non-UK high-school qualifications (name, country obtained, subjects studied, and grades)
 - **VCQFQ** – full details of all non-UK vocational qualifications (name, country obtained, subjects studied, and grades)
 - **VCQAFQ** – full details of any non-UK vocational qualifications currently being studied (name, subjects, and expected awarding country)
 - **DEM1** – attitude to democracy: *“Democracy may have its problems but it is the best system of government.”*
 - **DEM2** – perception of how democratic the UK is today
 - **VOT2** – whether the respondent voted in the general election (asked only after 4 July 2024)
 - **GAD1** – felt nervous, anxious or on edge in the past two weeks
 - **GAD2** – not being able to stop or control worrying in the past two weeks
 - **PHQ1** – little interest or pleasure in doing things in the past two weeks
 - **PHQ2** – feeling down, depressed or hopeless in the past two weeks.

5.3 Data linkage

During the Age 17 Sweep participants were asked to provide consent to link their survey responses with data held by various government departments and agencies. In the Age 23 Sweep those who had not previously consented, either because they did not take part at Age 17 or chose not to give consent at that time, were asked again. Table 4 summarises the records for which consent was requested.

Table 4. Data linkage consent requests

<p>Education records</p>	<p>Department for Education (England)</p> <p>Knowledge and Analytical Services within the Welsh Government (Wales)</p> <p>Education Analytical Services Division of the Scottish Government, the Scottish Funding Council and the Student Awards Agency Scotland (Scotland)</p> <p>Department of Education and the Department for the Economy (Northern Ireland)</p> <p>Higher Education Statistics Agency, the Universities and Colleges Admissions Service (UCAS) and the Student Loans Company (All countries).</p>
<p>Health records</p>	<p>NHS England</p> <p>Digital Health and Care Wales</p> <p>Public Health Scotland and NHS National Services Scotland</p> <p>Health and Social Care (HSC) in Northern Ireland</p>
<p>Economic records</p>	<p>Department for Work and Pensions (England, Scotland, Wales)</p> <p>Department for Communities (Northern Ireland)</p> <p>His Majesty's Revenue and Customs (HMRC) (All countries)</p>

Police & criminal justice records	Ministry of Justice (Ministry of Justice) Police Scotland, the Crown Office and Procurator Fiscal Service (Scotland) Police Service and the Department of Justice (Northern Ireland)
--	--

In addition, if the cohort member was a parent, they were also asked for permission to link their survey responses to information held in their child’s health records (up to age 14) and education records (up to age 16).

A full description of the consent process and consent rates obtained is provided in the Millennium Cohort Study Sweep 8 (Age 23) Survey Technical Report. Section 4.3 describes the process for collecting consent and Section 10.12.3 describes the consent rates.

For up-to-date information about the availability of linked data for research visit the [Linked data webpage on the CLS website](#).

5.4 Other special features

5.4.1 Partner Survey

The Age 23 Survey included a Partner Survey designed to collect information on the circumstances and experiences of resident partners. This was the first time a Partner Survey had been conducted as part of MCS. Study members with a resident partner² were asked whether they were willing for their partner to be invited to complete a short online questionnaire. Where permission was given, the partner was provided

² Resident partners were identified as sharing the same permanent address as the cohort member in the Household relationships module of the interview. There were no restrictions on how long a partner had been living with the cohort member.

with login details for the web survey and an information booklet explaining what participation would involve.

During face-to-face interviews, interviewers provided these materials directly to the partner if they were present at the time of the interview or left them with the cohort member if they were not. Where the cohort member participated online, this information was sent to the partner by post or email.

The Partner Survey primarily comprised questions that were also asked of cohort members and took an average of 15 minutes to complete. Partners were offered a £10 voucher as a thank you for taking part. Topics covered included ethnicity; education and qualifications; activity and employment; income; physical health; risky behaviours (including alcohol consumption, drug use and gambling); mental health and wellbeing; and family background.

In total, 1,403 study members (15% of those completing a fully or partially productive interview) were identified as having a resident partner. Permission to invite the partner to take part was given by 904 study members (64.4% of identified partners). Of those invited, 570 partners completed the survey, representing 40.6% of all identified partners and 63.1% of those invited to participate.

Table 5. Partner response rates

Outcome	Total
Fully or partially productive study members	9,323
Cohort members living with a partner	1,403
Partner invited	904
Productive partner interview	570
No partner interview	334
Eligibility, placement and response rates (%)	
% Eligible study members	15.0%
% Placement of partner interviews	64.4%
% Productive partner interview within eligible	40.6%

Outcome	Total
% Productive partner interview within placed	63.1%

5.4.2 Occupation coding

Study members were asked to provide details about their current job and about the job they expected to have at age 30. Current jobs were coded to the Standard Occupational Classification (SOC 2010 and SOC 2020) and the Standard Industrial Classification (SIC 2007). Partners were also asked to provide information about their own jobs and about the jobs of their parents, which were also coded using the same classification systems. Study member’s current occupation was also coded to the National Statistics Socio-economic Classification (NS-SEC).

During the soft launch, a script error meant that the question **EMPS** (“Number of people that work for the employer”) was asked using incorrect response options. These soft launch response options did not map neatly onto the categories required for deriving NS-SEC. The error was corrected for the main stage, where the intended response options were used. As a result, data collected at the two stages cannot be combined and have therefore been provided separately as *hsempsversion* (soft launch) and *hsempsversion2* (main stage).

The same script error affected the question **SENEMP** during the soft launch and was likewise corrected for the main stage. The soft launch and main stage versions are provided as *hssenempversion1* and *hssenempversion2* respectively. Downstream, this affected the derivation of NS-SEC codes: respondents interviewed at the soft launch who selected the response option “10–49 employees” were treated as having selected the option “10–25 employees” for the purposes of NS-SEC derivation.

Resident partners (of the cohort member) were not asked about special training required for their (and their parents’) jobs and therefore this information was not available for the SOC coding process.

Table 6 summarises the occupation coding variables available via the UK Data Service (see section 6.1). To minimise disclosure risk, truncated SOC codes are

included in the safeguarded dataset. The 4-digit SOC codes are available under Secure Access (see Section 6.2).

Partner occupation and parental occupation (for the partner's parents) will be included in the Partner Interview data as part of the Phase 2 deposit. Full details of the partner and parent-of-partner occupation variables will be provided in an updated version of this User Guide when these data are released.

Table 6. Variable names relating to occupation

	Current Job	Partner's job	Partner's father's job	Partner's mother's job
3 digit SOC code (per SOC2020)	Hssoc2020_r	To be deposited in Phase 2		
3 digit SOC code (per SOC2010)	Hssoc2010_r	To be deposited in Phase 2		
NS-SEC (SOC 2020)	hsnssec	n/a	n/a	n/a
NS-SEC 8 (SOC 2020)	hsnssec8	n/a	n/a	n/a
NS-SEC 5 (SOC 2020)	hsnssec5	n/a	n/a	n/a

5.4.3 Cognitive assessments

The Age 23 Survey involved two cognitive assessments – the Backward Digit Span and the Stroop task.

Backward Digit Span

Participants were asked to complete the Harvard's Test MyBrain Backward Digit Span task. The task is a measure of working memory (Singh et al., 2021) which was also included in the Next Steps Age 32 Survey. The task involved the participant memorising a sequence of digits and then recalling them in the reverse order (for example, '1', '2' would be recalled as '2', '1'). The sequence started with two numbers and increased by one digit at a time as it progresses, reaching a maximum of 11 digits. There were two sequences at each length. Study members started with 'practice' trials and then moved on to 'test' trials for the actual exercise. The task ended when a study member failed to correctly recall either of the two sequences of a particular digit length, or correctly entered all displayed numbers up to the maximum number of digits. They also had an option to close the exercise before it was complete if they wished to do so.

The key variables related to the Backwards Digit Span task can be found in the deposited mcs8_23y_cm_survey dataset and are shown in Table 7.

Table 7. Backward Digit Span variables

Variable name	Variable label
HSCOGSCORE	'Backward Digit Span: Score' This is the aggregate score (highest number of digits recalled correctly) for cohort members who completed the 'test' trials and ranges between 0 and 11.
HSNUMCORRECT	Backward Digit Span: Total number correctly recalled sequences This is the total number of correctly recalled sequences (excluding practice trials) and ranges between 0 and 20.

Variable name	Variable label
HSCOGDEVICE	<p>Backward Digit Span: Entry method (1st test trial)</p> <p>This shows the type of input action used by the study member on their response to the first 'test' trial, required by their device (keyboard, pen/stylus, touch).</p>

A summary of the scoring (HSCOGSCORE) is given below:

Table 8. Summary of Backward Digit Span Scores

Score	Explanation
1	Failed both attempts when sequence length of response is 2
2	Failed both attempts when sequence length of response is 3
3	Failed both attempts when sequence length of response is 4
4	Failed both attempts when sequence length of response is 5
5	Failed both attempts when sequence length of response is 6
6	Failed both attempts when sequence length of response is 7
7	Failed both attempts when sequence length of response is 8
8	Failed both attempts when sequence length of response is 9
9	Failed both attempts when sequence length of response is 10
10	Failed both attempts when sequence length of response is 11
11	Successfully completed either attempt when sequence length of response is 11

Trial-by-trial data are available in a separate dataset, `mcs8_23y_cm_bds_long`, which contains responses for each practice and test trial of the Backward Digit Span task. The dataset includes reaction time and the duration taken to complete each trial and covers both practice and test trials. An additional variable, `HSSTATE`,

records the input action used by study members for each trial (for both practice and test trials) and uses the same response options as HSCOGDEVICE. **HSBDSIDX** uniquely identifies each trial, in combination with the cohort member identifier, **MCSIDCMH**.

Stroop Task

Study members were asked to complete a second cognitive task, the Stroop task, (Stroop, 1935) described to participants as the '*colour naming task*'. The purpose of the Stroop task is to measure selective attention and inhibitory control – that is, how well people can focus on relevant information while suppressing distracting or conflicting information. In this task, participants were asked to identify the colour in which a word was displayed on the screen, rather than the meaning of the word itself. For example, if the word 'red' was shown on the screen but the word was coloured blue then the answer is blue. There was no demonstration or practice run; the instructions were presented on the first screen. Study members with substantial vision problems were not eligible to complete the task.

If a cohort member requested help, interviewers were allowed to share the screen and ask the participant to say the answer aloud for each item; the interviewer then selected the response on their behalf. Interviewers were explicitly instructed not to read out any of the words shown on the screen.

The variable **HSSTROOP_BLOCKEDIDX** represents an individual Stroop trial, defined by a specific word–colour combination. Although respondents were presented with these trials in a random order, the order itself was not recorded in the dataset. There are 32 possible word/colour combinations, and each **HSSTROOP_BLOCKEDIDX** uniquely corresponds to one of these combinations. Not all cohort members completed all trials, so some participants have only partial data. These partial responses remain in the dataset.

5.4.4 Geographical Indicators

Region and country of residence are available as part of the safeguarded data (End User Licence).

In addition, the detailed geographical identifiers (e.g. LSOAs, Local Authorities, etc) will be available from the UK Data Service as controlled data (Secure Access) at a later stage, and will be documented in a user guide once the data are available.

5.5 Scales

The Age 23 Survey included several established scales which are listed below. Overall scores for each scale have been derived and included within the data deposit (mcs8_23y_cm_derived) and described below. Further details regarding the derivation of the scores can be found in Appendix 1, and original wording used in the scales can be found in the Millennium Cohort Study Sweep 8 (Age 23) Questionnaire.

5.5.1 ONS long lasting health conditions and illnesses: Impairments and Disability

The Age 23 Sweep included a sub-set of the ONS harmonised (ONS 2015) set of questions on Long-lasting Health Conditions and Illnesses including Impairments and Disability. These are described in detail in the [long lasting health conditions and illness harmonised standard on the Government Analysis Function website](#). The three items listed below are used to derive variables indicating whether cohort members are disabled using the Equality Act 2010 definition (**HDDISEQ**) and whether they have a long-standing illness or condition using the European Union's Statistics on Income and Living Conditions (EU-SILC) definition (**HDDISLS**). **HDDISEQ** identifies individuals as disabled or not, **HDDISLS** identifies individuals as having no long-standing health condition, having a condition which hampers daily activities to an extent and having a condition which severely hampers daily activities.

Table 9. Long lasting health conditions and illnesses variables

Variable name	Variable label
HSCLSI00	Has longstanding illness “Physical or mental health conditions or illnesses lasting 12 months or more”
HSCLSL00	Reduced day-to-day activities as result of longstanding illness
HSCLSP00	Length of time day-to-day activities affected by longstanding illness
HDDISEQ	(Derived) Disability classification Equality act (2010)
HDDISLS	(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 (**HSCLSI00**) and have a reduced ability to carry out day-to-day activities as a result of their illness (**HSCLSL00**).

According to the EU-SILC definition, a cohort member is considered to be disabled if they report a longstanding illness (**HSCLSI00**), have a reduced ability to carry out day-to-day activities as a result of their illness (**HSCLSL00**) and this reduced ability has lasted for more than 6 months (**HSCLSP00**). This variable also distinguishes between those that are disabled to some extent, and those that are severely hampered (from **HSCLSL00**).

Note that this scale was also used in the Partner Survey.

5.5.2 Kessler 6 – Psychological distress

Information on scoring and interpretation of this scale can be found on the [K6 Scales webpage on the Harvard website](#).

The Kessler 6 (K6) scale is a quantifier of non-specific psychological distress (Kessler et al., 2003). It consists of six questions about depressive and anxiety symptoms that a person has experienced in the last 30 days. The scale was used to measure psychological distress amongst parents between MCS2 (age 3) and MCS7 (age 17) and was answered by study members themselves for the first time at MCS7 (age 17).

The Kessler 6 (K6) scale is a quantifier of non-specific psychological distress. It consists of six questions about depressive and anxiety symptoms that a person has experienced in the last 30 days. For each question, respondents were offered a self-report scale of five possible answers plus don't know/don't wish to answer:

1. All of the time
2. Most of the time
3. Some of the time
4. A little of the time
5. None of the time
6. Don't know/Don't wish to answer

The questions are preceded by the statement: 'The next few questions are about how you have felt over the last 30 days'. The six questions are:

Table 10. Kessler 6 – Psychological Distress variables

Variable name	Variable Label
HSPHDE00	Kessler – how often felt depressed past 30 days
HSPHHO00	Kessler – how often felt hopeless past 30 days
HSPHRF00	Kessler – how often felt restless or fidgety past 30 days
HSPHEE00	Kessler – how often felt that everything was an effort past 30 days
HSPHWO00	Kessler – how often felt worthless past 30 days
HSPHNE00	Kessler – how often felt nervous past 30 days
HDKESSLER6	(Derived) Kessler K6 scale

Note that this scale was also used in the Partner Survey.

5.5.3 Warwick-Edinburgh Mental Wellbeing Scale (Short WEMWBS)

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 (Hanzlová et al., 2023). 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.

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

The scale was also used to measure well-being at MCS7 (Age 17).

Table 11. Short WEMWBS variables

Variable Name	Variable label
HSWWOP00	Whether felt optimistic past 2 week
HSWWUS00	Whether felt useful past 2 week
HSWWRE00	Whether felt relaxed past 2 weeks
HSWWDE00	Whether been dealing with problems well past 2 weeks
HSWWTH00	Whether been thinking clearly past 2 weeks
HSWWCL00	Whether been feeling close to other people past 2 weeks
HSWWMN00	Whether been able to make up own mind about things past 2 weeks
HDWEMWB	(Derived) Warwick Edinburgh Mental Well-Being scale

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

5.5.4 Self-completion module: ONS Loneliness 4 (UCLA loneliness 3 item + overall lonely item)

Three items from the 20-item UCLA loneliness scale (Russell, 1996 and Hughes et al., 2004) were asked of all cohort members. Participants were asked to report how

often they experienced different aspects of loneliness using 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 which is not part of the UCLA scale, but has been used in the Millennium Cohort Study in accordance with ONS recommendations to use a direct measure as well as the 3 item scale (ONS, 2018).

Table 12. Loneliness variables

Variable Name	Question wording
HSLNLY100	How often do you feel that you lack companionship?
HSLNLY200	How often do you feel left out?
HSLNLY300	How often do you feel isolated from others?
HDLONELINESS	(Derived) UCLA loneliness 3 item
HSLNLY400	How often do you feel lonely?

Above items have also been asked to Millenium Cohort Study cohort members in the COVID-19 surveys.

5.5.5 Eating Disorder Examination – Questionnaire Shortened (EDE-QS)

Five items from the EDE-QS were used. The Eating Disorder Examination Questionnaire (EDE-Q) (Gideon et al., 2016) is a widely used measure of eating disorder symptom severity. To minimise respondent burden and potential distress

when asking about sensitive experiences, shorter forms of the scale have been developed (Lantos et al., 2024). Building on prior work that reduced the original 28-item EDE-Q to a validated 12-item short form (EDE-QS), further methodological work by the Centre for Longitudinal Studies in collaboration with the Department of Experimental Psychology, UCL used item response theory to identify a 5-item subset. This reduced version was designed to retain the key information captured by the longer scale while being suitable for inclusion in time-constrained survey questionnaires.

Table 13. Eating Disorder Examination variables

Variable Name	Variable Label
HSEDEQA00	Thinking about weight has made it difficult to concentrate
HSEDEQB00	Fear of gaining weight
HSEDEQC00	Strong desire to lose weight
HSEDEQD00	Has weight or shape influenced view of self
HSEDEQE00	How dissatisfied with weight or shape

Respondents were asked on how many days over the past week they had had certain feelings about their weight or shape e.g. on how many of the past 7 day has thinking about your weight or shape made it very difficult to concentrate on things you are interested in (such as working, following a conversation or reading)?

Respondents were able to answer:

1. 0 days
2. 1-2 days
3. 3-5 days
4. 6-7 days

5.5.6 Self-completion module: GRIT-8

The grit scale was developed to measure perseverance and passion for long-term goals, originally as a 12 item scale (Duckworth et al., 2007). An 8-item version was developed later (Duckworth et al., 2009), which was used in the Age 23 Sweep.

The scale consists of 4 positively and 4 negatively worded phrases, and responses are provided using a 5-point scale:

1. Very much like me
2. Mostly like me
3. Somewhat like me
4. Not much like me
5. Not like me

Responses were scored on a 1–5 scale, with higher scores indicating greater grit. Positively worded items were scored from 5 (“Very much like me”) to 1 (“Not like me”), while negatively worded items were reverse-scored. Scores were summed across all eight items and divided by eight to produce a mean scale score ranging from 1 (not at all gritty) to 5 (extremely gritty).

Table 14. Grit variables

Variable Name	Variable label
HSSGSA00	Distracted by new ideas and projects
HSSGSB00	Setbacks don’t discourage me
HSSGSC00	Obsessed with idea then lost interest
HSSGSD00	I am a hard worker
HSSGSE00	Often set a goal then pursue a different one
HSSGSF00	Difficulty maintaining focus on projects
HSSGSG00	I finish whatever I begin
HSSGSH00	I am diligent
HDGRIT18	DV: Short GRIT scale (to be delivered in Phase 2)

5.5.7 AUDIT-C

The AUDIT-C was used to capture alcohol consumption, problems, and dependency (Bush et al., 1998). Responses to the 3 questions below are scored from 0 to 4

giving a maximum score of 12 (HDAUDIT3). Scores of 5 or more are considered AUDIT-C positive and associated with increasing or higher risk drinking.

Table 15. AUDIT C variables

Variable Name	Variable label
HSALDR00	How often drinks alcohol
HSAUDIT200	Number of drinks on typical day
HSAUDIT300	How often had six or more drinks on one occasion in the past year
HDAUDIT3	DV: Alcohol Use Disorders Identification Test Consumption (AUDIT-C) score
HDAUDG3	DV: AUDIT-C Group

5.5.8 Social Provisions Scale (3-items)

Three items were included from the 10-item Social Provisions Scale (Cutrona, 1987). 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 three statements described their current relationship with other people from the following responses: 1. Very true 2. Partly true 3. Not true at all.

These questions were also asked at age 14 (MCS6) and age 17 (MCS7).

In addition, participants were also asked whether there was someone who would provide financial support if they needed it (variable: **MONY**).

Table 16. Social Provisions variables

Variable name	Question
HSSAFF00	I have family and friends who help me feel safe, secure and happy
HSTRSS00	There is someone I trust whom I would turn to for advice if I were having problems
HSNCLS00	There is no one I feel close to

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

The Generalized Anxiety Disorder 2-item (GAD-2) is a brief initial screening tool for generalized anxiety disorder. The GAD-2 (Kroenke et al., 2007) was based on the GAD-7 (Spitzer et al., 2006).

Respondents are asked how often they have been bothered by a) “Feeling nervous, anxious or on edge” and b) “Not being able to stop or control worrying” in the past 2 weeks, with the following response options:

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

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

The score for each question is:

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

Table 17. General Anxiety Disorder variables

Variable name	Variable label
HSGAD100	Felt nervous, anxious or on edge last 2 weeks
HSGAD200	Not being able to stop or control worrying last 2 weeks
HGDAD2	(Derived) Generalised Anxiety Disorder 2-item

These questions were also asked of Millennium Cohort Study cohort members in the COVID-19 surveys.

Please note that these questions were not asked in the soft launch.

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

The PHQ-2 (Kroenke et al., 2007) 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 (Spitzer et al., 2006). Respondents are asked how often they have been bothered by problems over the last 2 weeks:

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

Please note that these questions were not asked in the soft launch.

Table 18. Patient Health Questionnaire variables

Variable Name	Variable Label
HSPHQ100	Little interest or pleasure in doing things last 2 weeks
HSPHQ200	Felt down, depressed or hopeless last 2 weeks
HDPHQ2	(Derived) Patient Health Questionnaire 2-item

5.5.11 Positive Childhood Experiences

The Positive Childhood Experiences (PCEs) scale is a 7-item self-report measure designed to retrospectively assess positive childhood experiences from birth to age 18 years. Developed within a developmental psychopathology framework, the PCEs scale evaluates interpersonal experiences that serve as assets and resources for developing children (Bethell et al., 2019). For each statement study members were asked to think back to before the age of 18 and to reply yes or no.

Table 19. Positive Childhood Experiences variables

Variable Name	Variable labels
HSPCE100	Was able to talk to family about feelings
HSPCE200	CM felt family would stand by them
HSPCE300	Enjoyed participating in community traditions
HSPCE400	Felt a sense of belonging in secondary school
HSPCE500	Felt supported by friends
HSPCE600	Had at least two interested non-parent adults
HSPCE700	Felt safe and protected by an adult in home

5.5.12 Dysmorphic Concern Questionnaire

The Dysmorphic Concern Questionnaire (DCQ) is a widely used measure of concerns related to body image and appearance (Mancuso et al., 2010). To minimise respondent burden and allow inclusion within a time-constrained questionnaire, a shortened version was used. The original DCQ comprises seven items; methodological work by Centre for Longitudinal Studies in collaboration with the Department of Experimental Psychology, UCL (Lantos et al., 2024) used item

response theory to select three items that retain the core information captured by the full scale.

Table 20. Dysmorphic Concern Questionnaire variables

Variable name	Variable label
HSDCQA00	Concerned about physical appearance
HSDCQB00	Spent time worrying about appearance/bodily defect
HSDCQC00	Spent time covering up defects in your appearance / bodily functioning.

For each of these respondents could answer:

1. Not at all
2. Same as other people
3. More than most people
4. Much more than most people

6. Research Survey Data

6.1 Licensing and data access

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

All users of the data need to be registered with the UKDS. Details of how to do this are available on the [registration page of the UK Data Service website](#).

6.1.1 Safeguarded data (EUL)

The MCS8 datasets listed in Table 21 are available from the UKDS as safeguarded data, which can be downloaded from the UKDS once the End User Licence (EUL) access conditions have been accepted by the user.

Please refer to section 6.14 for information on how the safeguarded data have been de-identified for safe sharing. These safeguarded data exclude information that present a high risk of disclosure and/or are highly sensitive, such as open-text responses, personal identifiers, detailed health data, variables with very low counts, etc.

6.1.2 Controlled data (Secure Access)

Disclosive and/or highly sensitive MCS8 data can be accessed as controlled data from the UKDS SecureLab. Applicants wishing to access this data need to abide by the terms and conditions of the UKDS 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 [to access non-ONS controlled data can be found on the UK Data Service website](#).

6.2 Datasets and data structure

The list of the MCS8 datasets available from the UKDS is shown below. Most of the questionnaire data is contained in the datasets `mcs8_23y_cm_survey` and `mcs8_23y_cm_education`.

In addition, there are several hierarchical (long format) datasets, which contain multiple records for each cohort member. For example, the results from cognitive tests, where each record relates to a separate trial.

The identifiers are described in section 6.5.

Table 21. List of safeguarded datasets (End User Licence)

Dataset name	Contents	Identifier(s)
<code>mcs8_23y_cm_survey</code>	Main cohort member survey	MCSIDCMH, MCSID, HCNUM00
<code>mcs8_23y_cm_derived</code>	Variables derived from cohort member interview	MCSIDCMH, MCSID, HCNUM00
<code>mcs8_23y_cm_education</code>	Education information from cohort member survey	MCSIDCMH, MCSID, HCNUM00
<code>mcs8_23y_cm_bds_long</code>	Cognitive test: backwards digit span test results (long format)	MCSIDCMH, MCSID, HCNUM00, HSBDSIDX
<code>mcs8_23y_cm_stroop_long</code>	Cognitive test: Stroop test results (long format)	MCSIDCMH, MCSID, HCNUM00, HSSTROOP_ BLOCKEDIDX

Table 22. List of controlled datasets (Secure Access)

Name	Contents	Identifier(s)
mcs8_23y_cm_restricted	Disclosive or sensitive variables for the full MCS8 sample	MCSIDCMH, MCSID, HCNUM00
mcs8_23y_cm_survey_triplets_restricted	Main cohort member survey for triplets	MCSIDCMH, MCSID, HCNUM00
mcs8_23y_cm_derived_triplets_restricted	Variables derived from cohort member interview for triplets	MCSIDCMH, MCSID, HCNUM00
mcs8_23y_cm_education_triplets_restricted	Education information from cohort member survey for triplets	MCSIDCMH
mcs8_23y_cm_bds_long_triplets_restricted	Cognitive test: backwards digit span test results (long format) for triplets	MCSIDCMH HSBDSIDX
mcs8_23y_cm_stroop_long_triplets_restricted	Cognitive test: Stroop test results (long format) for triplets	MCSIDCMH HSSTROOPBLOCKEDIDX

6.3 Future datasets

Given the size and complexity of the MCS Age 23 Sweep, the data are being deposited with the UK Data Service (UKDS) in two phases. This initial deposit includes the majority of the data collected at Age 23.

A second data release to the UKDS (Phase 2) is planned for the remaining MCS8 data that require more extensive processing and quality checks and could not be included in the first phase.

The following data will be added in the Phase 2 deposit which is likely to be in April 2026:

- Outcome variables
- Household grid
- Partner interview
- Income (unfolding brackets)
- Education subjects and qualifications
- Combined school level qualifications
- Pregnancy and child data (controlled)
- Non-resident children
- Covid education impact assessment
- Detailed geographical indicators.

6.4 Data documentation

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

Table 23. Data documents

Name of the document	Content summary
Millennium Cohort Study Sweep 8 (Age 23) Household Grid And Proxy Questionnaire	Questions asked in the Age 23 sweep for the household grid and proxy questionnaires, including details on any routing, and mode specific adjustments.
Millennium Cohort Study Sweep 8 (Age 23) Main Questionnaire	Questions asked in the Age 23 sweep full/long version of the questionnaire, including detailed routing and mode specific adjustments. It also reflects any changes made to the questionnaire after the Soft Launch. There is no separate questionnaire for the mop up survey but questions included in the mop up survey are denoted by a *.
Millennium Cohort Study Sweep 8 (Age 23) Resident Partner Questionnaire	Questions asked of resident partners.

Name of the document	Content summary
Millennium Cohort Study Sweep 8 (Age 23) Technical Report Millennium Cohort Study Sweep 8 (Age 23) Technical Report Appendices	Documents authored by Ipsos which provide all technical details regarding the design and implementation of the survey.

6.5 Identifiers

6.5.1 MCSID, CNUM and PNUM

MCSID is the anonymised identifier for each family/household. A family holds the same MCSID across sweeps.

Individuals within a family can be identified by using the cohort member number CNUM (some families have more than one, e.g. in case of twins), and the person number PNUM, which is used to identify for any individual other than the actual cohort members. Individuals have the same PNUM and CNUM across sweeps.

6.5.2 Twins and triplets

MCS contains singletons, twins and triplets. For data disclosivity reasons, safeguarded data available on the UK Data Service only contain cases from singletons and twins. Data from triplets are made available as controlled data (Secure Access).

The MCS cohort members are now young adults. From MCS8 onwards, cohort members from original twin and triplet sets are now considered to have their own separate family units, even if they live in the same physical household as their twin/triplet siblings. For all sets of twins/triplets interviewed at MCS8, household composition may differ for each as they will have been interviewed on different dates from each other and possibly by different modes (interviewer face-to-face or online).

The number of leavers and new residents will depend on status at the date each cohort member was interviewed, as well as the order in which new household members were listed. It is therefore now possible that cohort member twins/triplets, even if living in the same physical household, may have different people in the same

household grid PNUM slot. For example, twin 1 adds new step-sibling with pnum=5, whereas twin 2 adds new step-dad with pnum=5 in their respective household grids.

6.5.3. The new MCSIDCMH identifier

Given the new setup for twins and triplets, all MCS8 datasets now contain a new common ID **MCSIDCMH**, which combines original MCSID with cohort member number CNUM to make a unique "cohort member household" ID going forwards.

Combining **MCSIDCMH** and person number PNUM will make a unique identifier for every "household member", i.e. this is equivalent to combining MCSID+CNUM+PNUM.

A hypothetical example is described below to illustrate the new IDs.

In this example, twins Emma (Twin 1) and Jack (Twin 2) grew up in the same household. At MCS8 (age 23) both twins had moved out of the family home and formed new households. Emma (Twin 1) now lives with her partner, Harry, and two other residents. Jack (Twin 2) lives with his partner, Maria.

Emma's household record (Twin 1)

Table 24.1 below describes Emma's household record. She no longer lives with persons A, B, C, or Jack (Twin 2). In Emma's new household at MCS8, she lives with partner, Harry, and two new residents D and E. Her partner Harry is new at MCS8 and has PNUM=4 in Emma's household.

Table 24.1 Household record for Emma (Twin 1)

MCSID (Origin household)	MCSIDCMH (CM household)	CNUM (CM number)	PNUM (Person number)	First name	Birth year	Present at MCS7?	Present at MCS8?	Relationship to cohort member
M44444Z	M44444Z1	1	100	EMMA	2001	1	1	self
M44444Z	M44444Z1	1	1	A	1968	1	0	parent
M44444Z	M44444Z1	1	2	B	1962	1	0	parent
M44444Z	M44444Z1	1	3	C	2005	1	0	sibling
M44444Z	M44444Z1	1	200	JACK	2001	1	0	sibling
M44444Z	M44444Z1	1	4	HARRY	2000	.	1	partner
M44444Z	M44444Z1	1	5	D	2002	.	1	other
M44444Z	M44444Z1	1	6	E	2003	.	1	other

Jack's household record (Twin 2)

Table 24.2 below describes Jack's household record. He also no longer lives with persons A, B, C, or his twin sister Emma (Twin 1). In Jack's new household at MCS8, he lives with his partner Maria only. Maria is new at MCS8 and has PNUM=4 in Jack's household.

Table 24.2 Household record for Jack (Twin 2)

MCSID (Origin household)	MCSIDCMH (CM household)	CNUM (CM number)	PNUM (Person number)	First name	Birth year	Present at MCS7?	Present at MCS8?	Relationship to cohort member
M44444Z	M44444Z2	2	200	JACK	2001	1	1	self
M44444Z	M44444Z2	2	1	A	1968	1	0	parent
M44444Z	M44444Z2	2	2	B	1962	1	0	parent
M44444Z	M44444Z2	2	3	C	2005	1	0	sibling
M44444Z	M44444Z2	2	100	EMMA	2001	1	0	sibling
M44444Z	M44444Z2	2	4	MARIA	1996	.	1	partner

In this example, the twins' partners, Harry and Maria, have the same MCSID+PNUM combination at MCS8. Therefore, we must use MCSIDCMH+PNUM for merges involving PNUM (HHGRID, CHILD, PARTNER, PARENT datasets) to avoid merging people onto data from the wrong twin.

6.5.4 Merging using the original identifiers MCSID and CNUM

The original identifiers MCSID and CNUM (**HCNUM00** at MCS8) are retained in all datasets to facilitate merging with datasets from sweeps 1 to 7. As there is no overlap between new household residents and MCS1-7 data, users can continue to merge origin household level data from earlier sweeps onto MCS8 using MCSID, and cohort member data from earlier sweeps using the keys MCSID+CNUM etc.

Some researchers choose to retain only one cohort member per MCSID, by dropping all twin 2 cases where (CNUM=2) from all datasets. If you choose to do this, you will not have any duplicates on MCSID+PNUM.

Note that in the household grid $PNUM = CNUM * 100$ for Cohort Members (including twin/triplets of cohort members within each other's households) to distinguish them from non-cohort-members.

6.5.5 Cognitive test trial identifiers

The cognitive test datasets, Backwards Digit Span test (BDS) (mcs8_23y_cm_bds_long) and Stroop test (mcs8_23y_cm_stroop_long) contain multiple records per person. Each trial can be uniquely identified by combining the cohort member ID **MCSIDCMH** and the secondary ID, **HSBDSIDX** and **HSSTROOP_BLOCKEDIDX**, in the Backwards Digit Span (BDS) and Stroop datasets respectively.

6.6 Variable description

6.6.1 Variable order

The order in which variables appear in the datasets broadly follows the order of modules and sections within modules, of the CAI program as documented in the questionnaire.

6.6.2 Variable names

The variable names in the dataset are based on those used in the CAI program and are documented in the questionnaire. These variable names are prefixed with 'H' denoting the current Age 23 sweep. The remaining characters have kept as close to the questionnaire documentation as possible and therefore have not been truncated to a maximum limit.

Variables derived in the CAI program, and those derived separately and included in the derived variables dataset ('mcs8_23y_derived') have been given the prefix "HD".

6.6.3 Multi-coded variables

Multi-coded variables have been given a common prefix based on the question content. Where a single question produces more than one response, a numeric suffix has been used to identify the iteration. 01, 02, 03, ..., 19, 20 etc has been used to denote the 1st, 2nd, 3rd, ..., 19th, 20th iteration respectively. In most cases these match the categorical values shown in the questionnaire for the respective question,

with additional numbers representing missing value categories and coded “other” responses.

Where numbers appear non sequential, this is usually due to those categories having low frequencies and therefore being available only under secure access. Variable names with suffix 00 or no numeric suffix relate to single response questions.

Examples of multi-coded variables in the questionnaire include:

Multi-coded variables	Overarching label
HSWHOTEN01 – HSWHOTEN07	Whose name accommodation is held in
HSASPR01– HSASPR12	Achievements by age 30

6.6.4 Variable labels

Variable labels are based on the wording that can be found in the questionnaire documentation. Where necessary, labels have been modified to ensure they are comprehensible and accurate.

CLS original datasets are produced in SPSS format, which allows very wide variable labels (120 characters). The UKDS translates the SPSS datasets into Stata format, which has more limited width allowance for variable labels (80 characters). As a result, this translation truncates the Stata variable labels, thus losing useful information. If using Stata files, it is therefore important to refer to the *.rtf* data dictionaries created by the UKDS that accompany the datasets, as these contain the full variable names and labels.

6.6.5 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.

6.7 Derived variables

Several derived variables have been produced based on the questionnaire data and are listed in the table below. The majority of these can be found in a separate derived variables dataset and detailed documentation on their derivation can be found in **Appendix 1**.

Table 25. List of derived variables

Variable Name	Variable Label
Paradata	
HDAGEINT	(Derived) Age in months at interview
HDAGEINTY	(Derived) Age at interview (year part)
HDAGEINTM	(Derived) Age at interview (month part)
HDSURVEY_MODE	(Derived) Survey mode of data collection
HDSC_MODE	(Derived) Mode for self-completion
Geography	
HDCNTRY	(Derived) Country of residence - CM
HDRGN	(Derived) Region of residence - CM
HDIMD	(Derived) Index of Multiple Deprivation rank Decile
Household Relationships	
HDTOTP	(Derived) Household size
HDNUMCH	(Derived) Number of children in household
HDNOCHH	(Derived) Number of biological children in household
HDNOC2A	(Derived) Number of own children aged 0 to 2 in household or absent
HDNOC4A	(Derived) Number of own children aged 3 to 4 in household or absent
HDNOC11A	(Derived) Number of own children aged 5 to 11 in household or absent
HDSIBHH	(Derived) Number of CM's siblings in household
HDFINH	(Derived) CM's natural father in household
HDMINH	(Derived) CM's natural mother in household

Variable Name	Variable Label
HDAPARNO	(Derived) Number of CM's adoptive/foster/step/parental figures in household
HDMS	(Derived) CM's Legal marital status
HDCOHAB	(Derived) Whether cohabiting as a couple
Housing	
HDTIMAD	(Derived) Time at current address (months)
Employment	
HDECONACT	(Derived) CM Current economic activity status
HDPECONACT	(Derived) Partner Current economic activity status
HDDWRKCP	(Derived) Combined labour market status
Finance	
HDINCB	(Derived) Banded weekly income of cohort member and partner
HDBENE	(Derived) Whether cohort member or partner receives any benefits
HDFINLIT3	(Derived) Financial Literacy – All three questions correct
HDFINLITC	(Derived) Financial Literacy – Number of correct answers
HDFINLITA	(Derived) Financial Literacy – Number of questions attempted
HDFINLITR	(Derived) Financial Literacy- At least one question refused
Education	
HDACDQC	(Derived) Highest academic qualification including currently studying in MCS8
HDACDQ	(Derived) Highest academic qualification obtained in MCS8
HDHACDQC	(Derived) Highest academic qualification including currently studying to age 23
HDHACDQ	(Derived) Highest academic qualification obtained to age 23
HDDDEG	(Derived) Degree or above obtained
HDPDEG	(Derived) Higher degree obtained
Health	
HDDISEQ	(Derived) Disability classification Equality act (2010)
HDDISLS	(Derived) Disability classification EU-SILC
HDHGHTM	(Derived) Height in metres – Self-reported
HDMHGHTM	(Derived) Height in metres – Measured and Self-reported combined

Variable Name	Variable Label
HDWGHTK	(Derived) Weight in kilograms – Self-reported
HDMWGHTK	(Derived) Weight in kilograms – Measured and Self-reported combined
HDBMI	(Derived) Body Mass Index – Based on self-report
HDMBMI	(Derived) Body Mass Index – Based on measure and self-report combined
HDBMIC	(Derived) (Derived) Body Mass Index classification – Based on self-report
HDMBMIC	(Derived) (Derived) Body Mass Index classification – Based on measure and self-report combined
HDSMOKE	(Derived) Smoking habits
HDAUDIT3	(Derived) Total AUDIT-C Score
HDAUDG3	(Derived) AUDIT-C Group
Mental Health	
HDKESSLER	(Derived) Kessler K6 scale
HDWEMWB	(Derived) Warwick Edinburgh Mental Well-Being scale
HDLONELINESS	(Derived) UCLA Loneliness 3-item
HDGAD2	(Derived) Generalised Anxiety Disorder 2-item
HDPHQ2	(Derived) Patient Health Questionnaire 2-item

6.8 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.

These data will be available from the UK Data Service in the Phase 2 deposit (see section 6.3).

6.9 Household grid

The MCS8 Household Grid contains details of all people who have ever been present in the cohort member's household (MCS1-7) and whether they are present at MCS8. It also contains people who are new to the household at MCS8.

At MCS7, an online follow up was conducted for those who had not responded to the main survey. This was equivalent to the MCS8 'mop up'. The Household Grid questions were not asked at the MCS7 online follow up.

The detailed MCS8 household grid will be available from the UK Data Service as controlled data (Secure Access) in the Phase 2 deposit. It will contain the original (MCS1-7) family ID MCSID and cohort member number of person being interviewed, **HCNUM00**, along with the new MCS8 Cohort Member Household ID **MCSIDCMH** described in section 6.4.

For twin and triplet cohort member households the composition of new household members may diverge from those of their cohort member siblings (see section 6.5). Therefore, it is important to use **MCSIDCMH** when merging person level data between the Household grid and any other datasets within MCS8. This is important for partner and child data, as there is a possibility of two different people sharing a single PNUM within the origin MCSID (as explained in section 6.5)

6.10 Longitudinal family and cohort member files

The MCS sweeps 1-7 survey data are accompanied by the MCS longitudinal family file (`mcs_longitudinal_family_file`), which contains the outcomes of MCS families at

each sweep and can be used to follow the families longitudinally. This dataset, now renamed as `mcs_longitudinal_family_file_1-7` is shared separately to the survey data and is available under UKDS SN8172.

From MCS8, rather than following the original MCS family units, cohort members from twin and triplet families are treated as having their own unique households, given that many will no longer live together, as mentioned in previous sections. To account for this, a new MCS1-8 longitudinal dataset has been constructed at cohort member level instead of at family level. This new dataset is called the MCS longitudinal cohort member file (`mcs_longitudinal_cm_file`).

This new longitudinal file contains information about the outcomes of individual cohort members at MCS8 and those of their family at each previous sweep (1-7), and can therefore be used for research focused on one or multiple MCS sweeps, and to follow cohort members longitudinally.

The file `mcs_longitudinal_cm_file` includes the weights that have been calculated for sweeps 1 to 8 (family level from MCS1-7 and cohort member level at MCS8) and the user can use them either for research using data from one sweep or when using data from more than one sweep. Specific information on how the weight variables have been constructed for each sweep is provided with the User Guide of the respective sweep.

The previous MCS1-7 longitudinal family file, has been renamed as `mcs_longitudinal_family_file_1-7`, and has been retained for those who need to continue using it.

The two longitudinal files and associated user guide are available under SN8172.

6.11 Missing values

Missing values are labelled as follows:

-9 = Prefer not to say

-8 = Don't know

-1 = Not applicable

-3 = Not asked at case fieldwork stage

In MCS8 cohort members who did not want to answer were coded as 'prefer not to say'. In other sweeps this was coded as 'refusal'. Users should be aware of the different variable labels when merging data.

In MCS8 the missing value -3 has been reserved for questions that were not asked in the short mop up questionnaire.

Value -1 is also used for missing responses to questions which study members would not have been asked if they only partially completed the survey.

For derived variables, value -8 is typically reserved for "Not enough information", where there is insufficient data for the variable to be derived.

6.12 Data errors and inconsistencies

Users should be aware of several data inconsistencies.

6.12.1. Missing information on twin/triplets who left the household

Due to a script error during data collection, some cohort members were not asked about twins or triplets in the household grid. This is being fully investigated and will be reported on in the next version of the user guide.

6.12.2. Snapbacks

Where a respondent was routed into a question, responded to that question but then changed their mind and hit the 'back' button, their original response may still appear in the data. This has led to some inconsistencies that may look like routing errors but are not. This problem has been named a "snapback", where a respondent has snapped back to an earlier question and changed an answer.

For example, if a cohort member responded that they smoke cigarettes every day they were routed into the question that asks for number of cigarettes smoked, to which they could have typed "12". Then they could have changed their mind and hit the 'back' button rather than delete their response. If they then responded that they did not smoke cigarettes every day, the collected data might still contain the response "12", which looks like a routing error. This occurrence did not happen

often, but it should be taken into account when users find what look like routing inconsistencies.

Every attempt has been made to resolve such inconsistencies but please contact CLS to report any additional instances of this that are found.

6.12.3. Inconsistency in country values

At MCS8, the response categories used for country questions were asked in a different order to that used in previous MCS sweeps (1-7).

Due to this questionnaire inconsistency, we are only providing the country derived variable HDCNTY, which is included in the Derived Variables dataset and has been computed from the residence postcodes. This will ensure that the values and labels are consistent with MCS1-7:

0=Unknown, 1=England, 2=Wales, 3=Scotland, 4=NI

6.13 Data de-identification

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

- The original SOC have been merged and truncated to 3 digits in line with previous sweeps and further truncated to 2 digits where counts are still < 10 after truncation to 3 digits. These variables are **HSSOC2020_R** and **HSSOC2010_R**. The SIC industry codes are given to 2 digits **HSSIC_R**.
- The long-standing conditions and illnesses questions are represented at the highest category.

The original version of the variables are either available under Secure Access or can be requested separately.

Table 26. Variables edited for de-identification purposes

Variable name	Variable label
HSNUMROOMS_R	Number of rooms in current home (top-coded)
HSNUMOTHCHILD_R	How many other children CM has (top-coded)
HSMOVINY_R	Year moved to current address (top-coded)
HSNUMHOML_R	Number of times CM has been homeless after age 16 (top-coded)
HSPRMN_R	Number of times pregnant/ fathered a pregnancy (top-coded)
HSSUNO_R	Number of times has ever hurt self on purpose in attempt to end life (top-coded)
HSBUYR_R	Year (share of) current property (top-coded)
HSPURCHPR_R	Purchase price (share of) current property (grouped)
HSBORROW_R	Amount of purchase price borrowed (grouped)
HSRENT_R	Rent amount (grouped)
HSRENT_DV_R	Computed annual rent amount (grouped)
HSMMOP_R	Rent amount (grouped)
HSSOC2020_R	DV: Current Job SOC2020 (3 digits)
HSSOC2010_R	DV: Current Job SOC2010 (3 digits)
HSSIC_R	DV: Current Job SIC code (2 digits)

6.14 Output disclosure control

Data included in Table 22 is only available via the UKDS Secure Lab. Access to this data is controlled and 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 their website](#).

The two UK Data Service Secure Lab rules of thumb that will be applied to all outputs are:

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

7. Response patterns and weights

7.1 Response patterns

The issued sample for the first sweep (9 months) of MCS was 20,646 families. A total of 18,552 families were interviewed (comprising 18,818 cohort members) in the initial sweep, representing 90% of the target sample. At Sweep 2 (Age 3y), 691 further families were added, taking the total number of families who have taken part in the study up to 19,243³.

MCS contains singletons, twins and triplets. For data disclosivity reasons, data are only available on the UK Data Service for singletons and twins (see section 6.4). In what follows, we discuss response patterns and non-response weights derived for the group of singletons and twins (19,482 cohort members from 19,231 families).

9,717 cohort members participated at Sweep 8 (Age 23y, 49.9% of the total sample).⁴ Sample sizes and response rates (as a proportion of all cohort members ever participating) for each sweep are shown in Figure 1.

³ One original family was mistakenly interviewed as a new family in Sweep 2 and has subsequently been kept under their Sweep 2 serial number. This was only discovered in latter waves, which is why earlier user guides mention 692 families joining in Sweep 2 and 19,244 families being included in the MCS sample overall.

⁴ Section 1 notes that 9,735 cases completed an interview, which include 17 individuals from triple families and 1 case subsequently requesting their data to be deleted ($9,735 = 9,717 + 17 + 1$). As also discussed in Section 1, deposited safeguarded dataset contains data for 9,675 participants. This does not include 42 proxy interview cases, which are included in the 9,717 headline figure here ($9,675 = 9,717 - 42$).

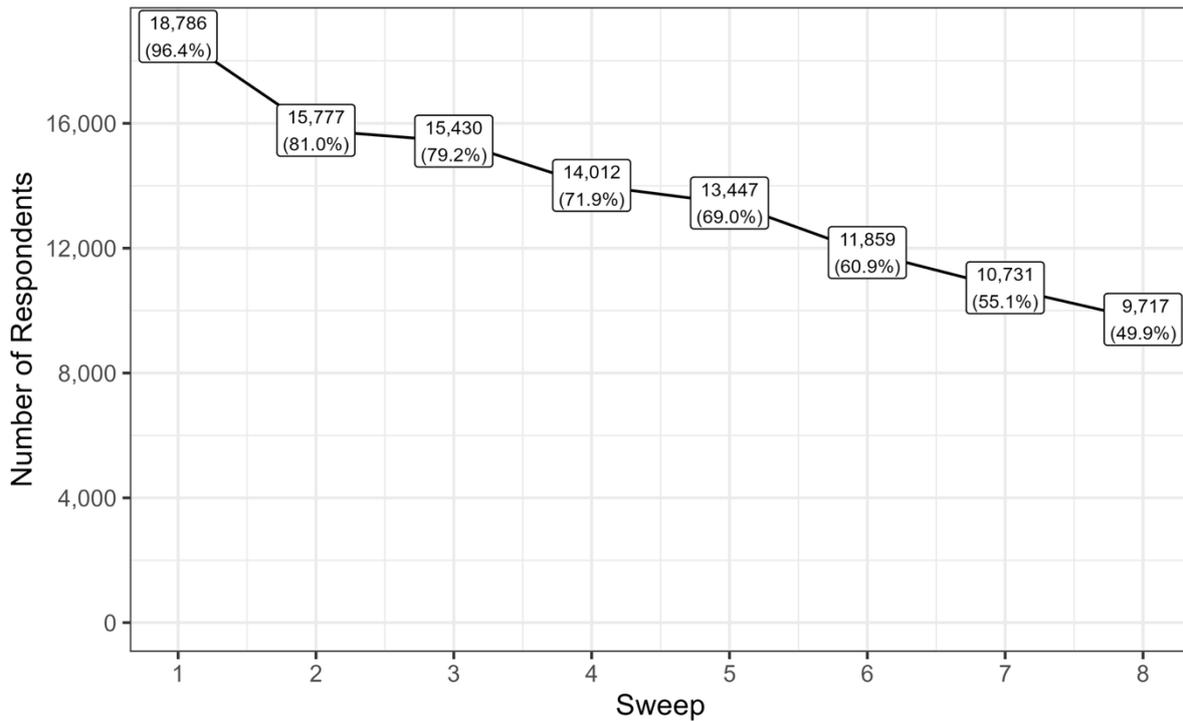


Figure 1. Sample size and response rates by sweep (productive participants)

Note: Response rates calculated as proportion of total ($n = 19,482$) sample of cohort members from singleton or twin families. Cohort members from 'New Families' first participating in Sweep 2 (3y) are therefore included in Sweep 1 when these participants were not issued for inclusion in the study.

In each sweep, cohort members have been issued to participate regardless of prior non-response. As such, not all participants in a given sweep have participated in the sweep immediately prior. Figure 2 below includes three plots that are useful for understanding non-response in MCS. Panel A shows the proportion of participants at a given sweep who had participated at all prior sweeps. Panel B shows the proportion of Age 23 participants who participated at a given prior sweep. Both are useful for understanding missingness in data from prior sweeps. Panel C shows the proportion of participants at a given sweep who participated at Age 23. These figures are useful for understanding attrition from the survey.

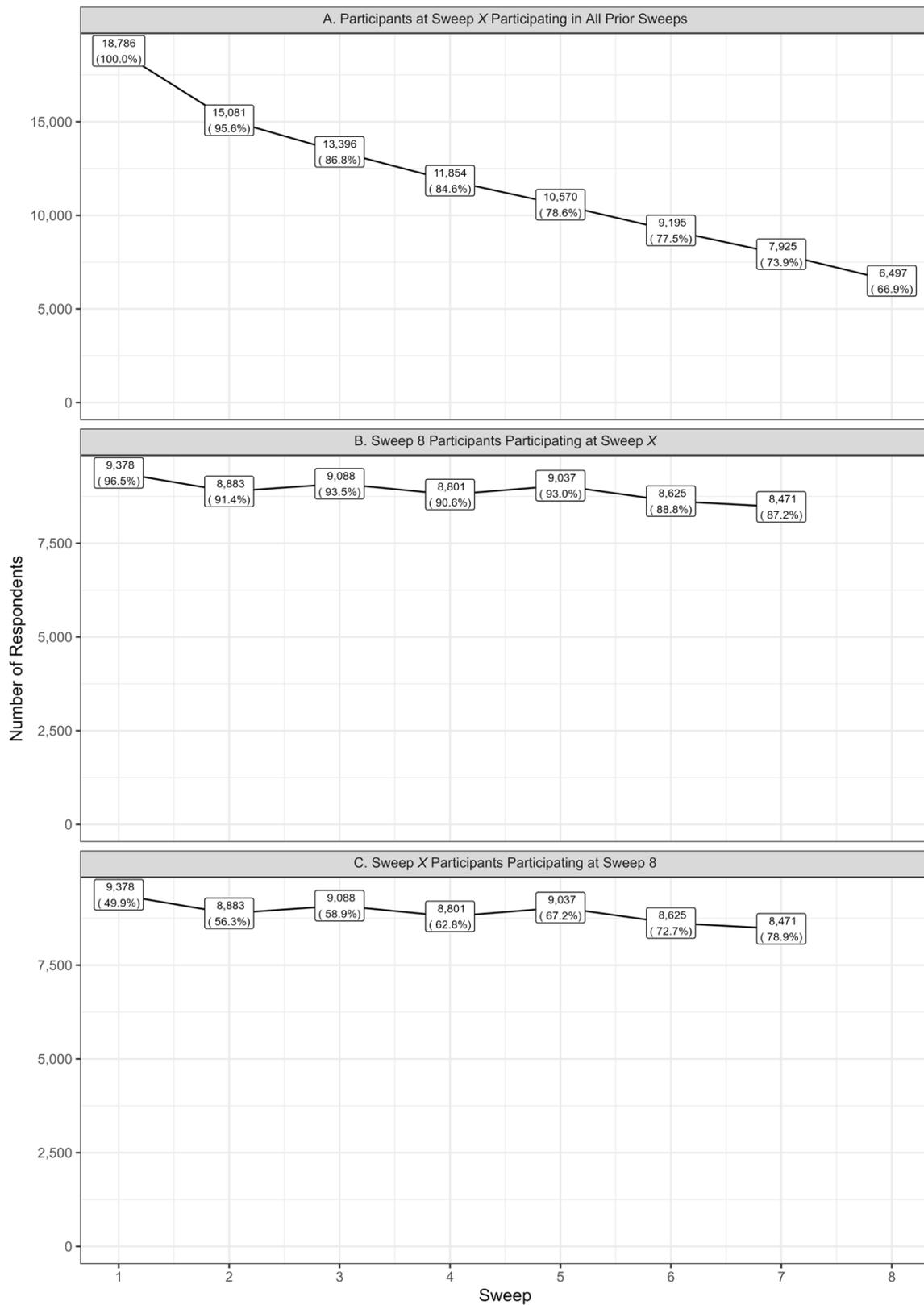


Figure 2. Response rates by sweep

Note: Although cohort members from ‘New Families’ first participated in Sweep 2 (3y), all Sweep 9 (3y) participants are included in the denominator in Panel B which is why the proportion of Sweep 9 respondents who took part in Sweep 1 is <100%.

7.2 Predicting response in Age 23 and deriving non-response weights

7.2.1 Introduction

In longitudinal studies like MCS, it is inevitable that some members will drop-out of the study. Attrition means less statistical power but can also introduce bias as respondents often differ systematically from non-respondents. There are several approaches that can be used to handle missing data, including multiple imputation (MI), weighting, and full-information maximum likelihood (FIML). Elsewhere, CLS has developed comprehensive advice on dealing with missing data in its cohort studies (Silverwood et al., 2024). Here we discuss, in abbreviated form, the derivation and implementation of a non-response weight that is provided with the MCS Age 23 Sweep 8 data. More detailed information is provided in Appendix 2.

7.2.2 Target populations and response definition

Not all MCS cohort members were eligible to complete the Age 23 Sweep 8 survey. The target population used to create the weight (the population the weights attempt to recover) was *all children born between 1 September 2000 and 31 August 2001 (for England and Wales), and between 23 November 2000 and 11 January 2002 (for Scotland and Northern Ireland), alive and living in the UK at age nine months, eligible to receive Child Benefit at that age, and for as long as they remain living at the time of sampling*. 19,471 MCS cohort members were covered by this definition.⁵ Note, this differs slightly to the target population in prior sweeps where emigrants

⁵ Note, again: here we are limiting to the singleton and twin sample that is available on the UK Data Service

from the UK were also excluded. The reason for this change is the introduction of a web survey at Age 23 which could be completed outside the UK.

For the purposes of weighting, response was defined as participation in the CAPI, CAWI or proxy versions of the Age 23 questionnaire (n = 9,717).

7.2.3 Derivation of Non-Response Weights

The derivation of the non-response rate proceeded as follows:

1. Within the sample corresponding to the target population, model 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 that can be found on the CLS website](#).
2. Among respondents to the Age 23 survey, calculate the probability of response implied by each model.
3. Calculate the non-response weight as the inverse of the probability of response.
4. Examine the distribution of the non-response weight to decide whether truncation may be desirable, applying truncation, if so.
5. Given MCS used a complex sampling design, multiply the survey non-response weights with initial survey weights (WEIGHT1, WEIGHT2, & WEIGHTGB) to create a set of 'combined weights' to recover the initial target population.
6. Finally, calibrate the combined non-response weights so that that they sum to the number of survey respondents (n = 9,717).

The variables included in the response model in are listed in Table 27. We used the same set of variables as used to derive weights for the MCS Age 17 (Sweep 7) survey, with the addition of two further variables (country at sampling and number of productive COVID-19 sweeps). The code used to clean the data and create the weights can be [viewed on the Open Science Framework website](#).

Table 27. Variables used as predictors in non-response models

Variable	Description
Sex	Categories: Male, Female.
Ethnicity	Categories: White, Mixed, Indian, Pakistani & Bangladeshi, Black or Black British, Other.
MCS Family Type	Categories: Original Family, New Family.
Country @ Sampling	Categories: England, Wales, Scotland, Northern Ireland.
# Productive Mainstage Sweeps	Mean: 5.1, Range: 1 - 7
# Productive COVID Sweeps	Mean: 0.5, Range: 0 - 3
Cognitive Ability @ Age 5y	Mean: 0, Range: -5 - 4.1
Internalising Symptoms @ Age 11y	Mean: 3.2, Range: 0 - 19
Externalising Symptoms @ Age 11y	Mean: 4.4, Range: 0 - 20
Cohort Member Breastfed	Categories: No, Yes.
Maternal Age at First Birth	Mean: 25.4, Range: 13.4 - 48.8
Maternal Mental Health @ 9 months	Mean: 1.7, Range: 0 - 9
# Parents in Housing @ Age 11y	Categories: Two parents/carers, One parent/carers.
Highest Main Parental Education, 0y-11y	Categories: NVQ level 1, NVQ level 2, NVQ level 3, NVQ level 4, NVQ level 5, Overseas Qualification, None.

Variable	Description
Highest Parental NS-SEC, 0y-11y	Categories: Managerial & Professional, Intermediate, Small Employers & Own Account, Lower Supervisory & Technical, Semi-Routine & Routine, Not in Work.
% Sweep in Employment	Mean: 0.5, Range: 0 - 1
Housing Tenure @ Age 11y	Categories: Own outright, Mortgage, Rent from LA, Rent from Housing Association, Rent privately, Other.
Accommodation Type	Categories: House or Bungalow, Other.
% Sweeps without Reporting Income	Mean: 0.1, Range: 0 - 1

Missing values in the above variables were handled using multiple imputation (MI). The imputation procedures were carried out in the eligible sample and the imputation models included the variables above plus a variable for response for Age 23 Survey. Separate imputation models were conducted for each survey weight being derived (forty imputed datasets created using chained equations).

Models for Age 23 survey response were fitted in each imputed dataset. Pooled models are reported in Appendix 2 From these models, the probability of Age 23 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 10 could provide some improvement in precision without undue introduction of bias. The non-response weights were therefore truncated to 10 and were then calibrated so that they summed to the number of Age 23 survey respondents for the relevant survey element by multiplying them by the ratio of the number of responses to the total of the uncalibrated nonresponse weight.

Finally, we multiplied the non-response weight with three sampling weights for single-country (WEIGHT1), UK-wide (WEIGHT2), and Great Britain only (WEIGHTGB)

analyses, respectively, again rescaling so the sum of the weights equalled the number of Age 23 survey respondents. These “combined” weights account for selection into the initial MCS sample and (sweep) non-response at Age 23 to recover the target population. The distributions of the resultant calibrated non-response and combined weights are presented in Table 28 along with the distribution of the non-response weight prior to truncation.

Table 28. Non-Response Weight Distributions

	Non-Response Weight		Combined (Sampling + Non-Response) Weight		
	Pre-	Post-	Single Country	Whole UK	GB Analysis
Mean	1.93	1.00	1.00	1.00	1.00
SD	3.15	0.85	0.96	1.12	1.05
0%	1.01	0.58	0.16	0.13	0.12
5%	1.01	0.58	0.18	0.16	0.18
25%	1.05	0.60	0.50	0.33	0.43
50%	1.29	0.74	0.86	0.71	0.79
75%	1.65	0.94	1.17	1.22	1.20
95%	4.43	2.54	2.41	2.43	2.34
100%	76.43	5.72	11.43	11.52	10.76

7.2.4 Weight Effectiveness

To examine the effectiveness of the derived non-response weights in restoring sample representativeness we conducted several analyses, some of which are presented here (with several more in Appendix 3). We considered the sample means of seven variables – birthweight, mothers age at cohort member’s birth, cohort member’s and mother’s anthropometrics, and behaviour at Age 5 – in the full sample and in the (weighted and unweighted) sample of Age 23 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 weighted sample mean to that in the full sample. The results of this exercise are presented in Figure 3. We observed that there is considerable bias when using the unweighted Age 23 survey data but this is markedly reduced – though not completely – when applying the non-response weights.

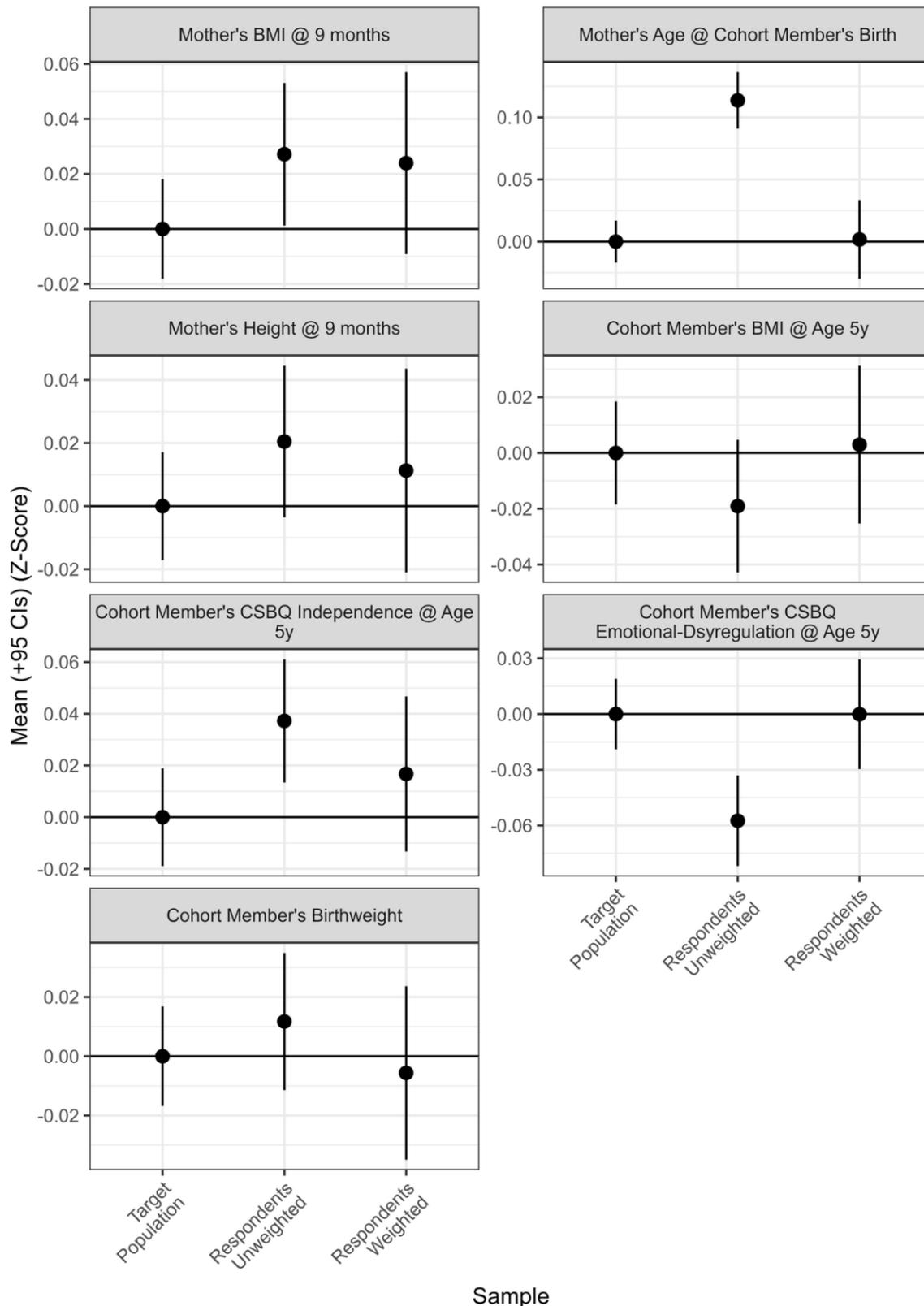


Figure 3: Means for seven variables that were not used to create non-response weights in the target population and in weighted and unweighted Age 23 responding sample. Variables standardised (mean = 0, SD = 1) using the target population mean and SD

Appendix 2 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, in several recovering sample means and proportions close to those in the full sample. Bias remains in several variables however, indicating that either observed information in MCS 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 MCS.

7.2.5 Implementation of Non-Response Weights

Data users should use weights that reflect both (a) the population they want their analysis to be representative of and (b) the variables used in their analysis. The weight we have provided are designed to be usable in a number of situations, but data users should always consider whether creating their own bespoke weights is required. This may be the case if, for instance, the analysis to be undertaken only uses data from those who displayed a particular response pattern (e.g., participated in Sweep 1 and Sweep 8), or are otherwise a subset of all Sweep 8 respondents (e.g., those in employment, those with children).

MCS used a complex sampling design to recruit cohort members: families were recruited from stratified samples of wards (small geographical areas). Analyses of MCS should account for this complex sampling design by specifying primary sampling unit (PSU; SPTN00) and stratum (PTTYPE2). In Stata, this can be achieved by first using `svyset` to specify the survey design and then conducting analyses using the `svy` prefix, e.g.:

```
svyset SPTN00 [pweight= HOVWT2], strata(PTTYPE2)
svy: proportion HSSMOKING00 if HSSMOKING00 >= 0, citype(agresti)
```

In R, the `survey` package can be used to specify complex survey design, e.g.:

```
library(survey)
```

```
mcs_svy <- svydesign(id = ~ SPTN00, strata = ~ PTTYPE2, weights = ~  
HOVWT2, data = mcs_23)  
svytable(~ HSSMOKING00 , mcs_svy)
```

See relevant Stata help files (StataCorp 2023) and survey package documentation (Lumley 2011) for more information on using survey data in Stata and R. Users with experience of the tidyverse may want to alternatively use the `srvyr` package, which provides similar functionality to `survey` but within a tidy framework.

Given attrition, in some cases, there may only be one PSU in a stratum, in which case users may consider allocating these observations to the modal stratum. Also, for some commands, users may find that certain functionalities have not been adapted for use with complex survey data. In this case, users may consider using survey weights without declaring complex survey design and noting this in the write up of their analyses. Users can also check whether accounting for complex survey design makes much difference in their particular analysis, by running analyses declaring and not declaring the complex design; as participants may have moved residence since birth, observations at Age 23 may be independent (or at least *less* dependent) at the PSU level.

8. Mode Effects

The Age 23 Sweep of the Millennium Cohort Study used a sequential mixed-mode design, in which cohort members were first invited to complete the survey online, with non-respondents followed up by interviewers. This approach helps reduce fieldwork costs and improve participation rates. However, mixed-mode designs can introduce differences in how respondents answer questions across survey modes. For example, some questions may be understood differently when read on screen compared with being asked by an interviewer, and respondents may report sensitive behaviours more accurately online than face to face. Such differences, known as *mode effects*, can affect the comparability of data collected across modes.

If not accounted for, mode effects can introduce bias into both descriptive and analytical estimates. For example, changes in health or wellbeing measures may partly reflect the mode in which the data were collected rather than true change over time.

It is important to note that simply including a variable for mode of completion in statistical models may not fully adjust for these differences. Because participation in a particular mode is not random, for instance, those who do not respond to initial online invitations are systematically different from early web respondents, observed differences across modes reflect a combination of true mode effects and selection effects.

CLS has published guidance for researchers on understanding and addressing mode effects in its cohort studies (Wright et al., 2024). This guidance outlines conceptual frameworks for considering how mode effects might influence specific research questions and describes a range of recommended analytical approaches. It also provides worked examples in R and Stata and offers practical recommendations for handling mode effects when analysing MCS data.

CLS are currently using data from the soft-launch mode experiment to examine in detail which variables from the Age 23 questionnaire are most likely to be affected by mode of data collection. This work includes comparing response distributions across experimental arms and assessing potential bias. Findings from this study will be

published shortly, and updated guidance will be made available to support researchers in working with the MCS Age 23 data.

In addition, CLS has recently conducted a broader methodological study to assess the extent of mode effects in key measures used in both the MCS Age 23 Survey and the Next Steps Age 32 Survey. A convenience sample of adults aged 20–40 in England (n = 1,800) completed two surveys approximately two weeks apart. Participants were randomly allocated to web, video or in-person modes at both time points, creating nine possible mode sequences. This experimental design enables a robust assessment of whether survey responses differ systematically by mode. Findings from this study, including updated guidance for analysts, will be published shortly.

The Backward Digit Span was one of the measures included in this methodological study. A dedicated research project examining the impact of mode on this assessment is now nearing completion and encouragingly, results are indicating that mode effects are minimal (Tsigaridis et al., 2026).

9. References

- Bethell, C., Jones, J., & Gombojav, N. (2019). Positive childhood experiences and adult mental and relational health in a statewide sample. *JAMA Pediatrics*, 173(11).
- Bush, K., Kivlahan, D.R., McDonell, M.B., et al. (1998). The AUDIT-C alcohol consumption questions. *Archives of Internal Medicine*, 158, 1789–1795.
- Cutrona, C.E., & Russell, D.W. (1987). The provisions of social relationships and adaptation to stress. In W.H. Jones & D. Perlman (Eds.), *Advances in Personal Relationships* (Vol. 1, pp. 37–67). Academic Press.
- Duckworth, A.L., Peterson, C., Matthews, M.D., & Kelly, D.R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92, 1087–1101.
- Duckworth, A.L., & Quinn, P.D. (2009). Development and validation of the Short Grit Scale (Grit–S). *Journal of Personality Assessment*, 91(2), 166–174.
- Gideon, N., Hawkes, N., Mond, J., Saunders, R., Tchanturia, K., & Serpell, L. (2016). Development and psychometric validation of the EDE-QS. *PLoS One*, 11(5).
- Hanzlová, R., & Lynn, P. (2023). Psychometric analysis of SWEMWBS. University of Essex.
- Hughes, M.E., Waite, L.J., Hawkley, L.C., & Cacioppo, J.T. (2004). A short scale for measuring loneliness in large surveys. *Research on Aging*, 26(6), 655–672.
- Kessler, R.C., et al. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, 60, 184–189.
- Kroenke, K., Spitzer, R.L., Williams, J.B.W. (2007). The PHQ-2 and GAD-2 screening tools. *Medical Care*, 45, 259–265.
- Lantos, J., et al. (2024). Short-form validation of EDE-QS and DCQ measures. *UCL Experimental Psychology*.
- Lumley, T. (2011). *Complex Surveys: A Guide to Analysis Using R*. Wiley.

- Mancuso, S.G., et al. (2010). The Dysmorphic Concern Questionnaire. *Psychological Assessment*, 22(3), 711–719.
- McKay, M., & Andretta, J. (2017). The Warwick–Edinburgh Mental Wellbeing Scale at Age 17. *British Journal of Psychiatry*.
- Ng Fat, L., et al. (2017). Population norms for SWEMWBS. *Quality of Life Research*, 26(2).
- ONS (2015). Harmonised standards for long-lasting health conditions. In *Harmonised Concepts and Questions for Social Data Sources Primary Principle* Office for National Statistics.
- ONS (2018). Recommended national indicators of loneliness. Office for National Statistics.
<https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/compendium/nationalmeasurementofloneliness/2018/recommendednationalindicatorsofloneliness>
- Russell, D. (1996). UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, 66, 20–40.
- Silverwood, R., et al. (2024). Missing data handling guidance for cohort studies. UCL.
- Singh, K., et al. (2021). TestMyBrain cognitive toolkit. Harvard University.
- Spitzer, R.L., Kroenke, K., Williams, J.B.W. (2006). The GAD-7 and PHQ-9. *Annals of Internal Medicine*.
- StataCorp. (2023). *Stata Statistical Software: Release 18*. StataCorp LLC.
- Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Tsigaridis, I., et al. (2026). Mode effects in cognitive assessment: Analysis of an experiment using in-person, video and web data collection. UCL.
https://osf.io/preprints/socarxiv/ue4ta_v1.
- Wright, L., et al. (2024). Mixed-mode survey methodological guidance. UCL.

Appendix 1: Derived Variables Guide

This Appendix describes the derived variables for MCS Age 23 Sweep (Sweep 8). These variables are all included in the safeguarded dataset ' mcs8_23y_cm_derived' under End User Licence.

Paradata

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

HDAGEINT - “(Derived) Age in months at interview”

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

Population: All cohort members

Value and Value Labels:

(-8) “Insufficient information”

Derivation description: $(\text{Interview year} - \text{birth year}) * 12 + (\text{Interview month} - \text{birth month})$

HDAGEINTY – “(Derived) Age at interview (year part)”

HDAGEINTM – “(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”

HDSURVEY_MODE – “(Derived) Survey mode of data collection”

Description: The primary mode in which the interview was conducted including interview stage

Population: All cohort members

Value Labels:

(1) “In-person (Soft Launch)”

(2) “Web (Soft-Launch)”

(3) “In-person (Main Stage)”

(4) "Web (Main Stage)"

(5) "Web (Mop-up)"

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

Description: How cohort members completed the section with sensitive questions

Population: All Soft Launch and Main Stage respondents

Value labels:

(-8) "Not enough information"

(-1) "Not applicable"

(1) "Web Completion"

(2) "Self-completion during in person interview"

(3) "Completed with interviewer during in person interview"

(4) "Self-completion refused"

Geography

The geography variables are produced using postcode data collected at interview.

HDCNTRY - "(Derived) Country of residence – CM"

Description: Country of residence at time of interview

Population: All cohort members living in the UK at interview

Value Labels:

(0) "Unknown"

(1) "England"

(2) "Wales"

(3) "Scotland"

(4) "Northern Ireland"

HDRGN - "(Derived) Region of residence -CM"

Description: Region in which interview took place

Population: All cohort members living in the UK at interview

Value Labels:

(0) "Unknown"

(1) "North East"

(2) "North West"

(3) "Yorkshire and the Humber"

- (4) "East Midlands"
- (5) "West Midlands"
- (6) "East of England"
- (7) "London"
- (8) "South East"
- (9) "South West"
- (10) "Wales (pseudo)"
- (11) "Scotland (pseudo)"
- (12) "Northern Ireland (pseudo)"

HDIMD – "(Derived) Index of Multiple Deprivation Decile"

Description: Index of Multiple Deprivation Decile

Population: All cohort members living in the UK at interview

Value Labels:

- (-8) "Not enough information"
- (1) "Most deprived decile"
- (10) "Least deprived decile"

Household Relationships

HDTOTP - "(Derived) Household 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.

HDNUMCH - "(Derived) Number of children in household"

Description: Number of children currently living in same household as cohort member. Children are defined either by relationship to CM, or anyone under the age of 18

Population: All cohort members

Value Labels:

(-8) "Not enough information"

(-1) "Not applicable"

Derivation description: Number of children in household, RELTOCM=3,4,5,6
(Natural, foster, adopted, step child) OR AGE < 18

HDNOCHH - "(Derived) Number of own children in household"

Description: Number of cohort member's own children currently living in same household as cohort member

Population: All cohort members

Value Labels:

(-8) "Not enough information"

(-1) "Not applicable"

Derivation description: Number of own (biological) children in household
RELTOCM=3

HDNOC2A - "(Derived) Number of own children aged 0 to 2 in household or absent"

Description: Number of cohort member's own children aged between 0 and 2

Population: Cohort members with children

Value Labels:

(-8) "Not enough information"

(-1) "Not applicable"

Derivation description: The number of CM's own children (RELTOCM=3) by age
(AGE >= 0 and <= 2).

HDNOC4A - "(Derived) Number of own children aged 3 to 4 in household or absent"

Description: Number of cohort member's own children aged between 3 and 4

Population: Cohort members with children

Value and Value Labels:

(-8) "Not enough information"

(-1) "Not applicable"

Derivation description: The number of CM's own children (RELTOCM=3) by age
(AGE >= 3 and <= 4).

HDNOC11A - "(Derived) Number of own children aged 5 to 11 in household or absent"

Description: Number of cohort member's own children aged between 5 and 11

Population: Cohort members with children

Value Labels:

(-8) "Not enough information"

(-1) "Not applicable"

Derivation description: The number of CM's own children (RELTOCM=3) by age (AGE >= 5 and <= 11).

HDSIBHH - "(Derived) Nuber of CM's siblings in household"

Description: Total number of natural, adopted, foster, step or half siblings of CM living in the same household

Population: All cohort members

Value Labels:

(-8) "Not enough information"

(-1) "Not applicable"

Derivation description: Number of siblings (RELTOCM = 11,12,13,14,15)

HDFINH - "(Derived) CM's natural father in household"

Description: Whether cohort member's natural father currently lives in same household as cohort member

Population: All cohort members

Value Labels:

(-8) "Not enough information"

(0) "No"

(1) "Yes"

Derivation description: Whether the natural father of the cohort member lives in the household. RELTOCM = 7 and SEX=1

HDMINH - "(Derived) CM's natural mother in household"

Description: Whether cohort member's natural mother currently lives in same household as cohort member

Population: All cohort members

Value Labels:

(-8) "Not enough information"

(0) "No"

(1) "Yes"

Derivation description: Whether the natural mother of the cohort member lives in the household. RELTOCM = 7 and SEX=2

HDAPARNO - "(Derived) CNumber of CM's adopted/foster/step/parental figures in household"

Description: Number of non-natural parental figures living with CM (includes partners of parents)

Population: All cohort members

Value Labels:

(-8) "Not enough information"

Derivation description: Total number of parental figures of the cohort member lives in the household. RELTOCM = 8,9,10

HDMS - "(Derived) CM's Legal marital status"

Description: Legal marital status of cohort member

Population: All cohort members

Value Labels:

(-8) "Not enough information"

(-3) "Not asked in fieldwork stage"

(1) "No longer married"

(2) "Married"

(3) "A Civil Partner"

(4) "Single and never married or in a CP" **Derivation description:** The legal marital status of cohort members who are currently single or who are married or in a civil partnership: MARST or household member RELTOCM=1 (spouse/civil partner)

HDCOHAB - "(Derived) Whether CM cohabiting as a couple"

Description: Whether Cohort member's partner living in the same household at interview

Population: All cohort members

Value Labels:

(-8) "Not enough information"

(0) "No"

(1) "Yes"

Derivation description: If any household member RELTOCM=1,2 (spouse/civil partner, partner/cohabitee)

Housing

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

Description: Number of months cohort member has lived at current address. This is not combined with previous sweep data, because this has previously only been asked to parents

Population: All cohort members providing new housing details at current sweep

Value Labels:

(-8) "Not enough information"

(-3) "Not asked in fieldwork stage"

(-1) "Not applicable"

Derivation description: The date the cohort member moved into their current address (MOVINM, MOVINY) subtracted from the date of interview (INTDATE(MONTH), INTDATE(YEAR)).

Activities and Employment

HDECONACT - "(Derived) CM Current economic activity status"

Description: Current activity of cohort member - backcoded

Population: All cohort members

Value Labels:

(-8) "Not enough information"

(-3) "Not asked in fieldwork stage"

(-1) "Not applicable"

- (1) "Employee - in paid work"
- (2) "Self-employed"
- (3) "In unpaid/voluntary work"
- (4) "Unemployed"
- (5) "Education: School/college/university"
- (6) "Apprenticeship"
- (7) "On gov't scheme for employment training"
- (8) "Sick or disabled"
- (9) "Looking after home or family"
- (10) "Something else"

Derivation description: The current activity of the cohort member (ACTIVITY) is recoded as follows: (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (>=10=10)

HDPECONACT - "(Derived) Partner Current economic activity status"

Description: Current activity of cohort member's partner - backcoded

Population: All cohort members with a cohabiting partner

Value Labels:

- (-8) "Not enough information"
- (-3) "Not asked in fieldwork stage"
- (-1) "Not applicable"
- (1) "Employee - in paid work"
- (2) "Self-employed"
- (3) "In unpaid/voluntary work"
- (4) "Unemployed"
- (5) "Education: School/college/university"
- (6) "Apprenticeship"
- (7) "On gov't scheme for employment training"
- (8) "Sick or disabled"
- (9) "Looking after home or family"
- (10) "Something else"

Derivation description: The current activity of the cohort member (PECONACT2) is recoded as follows: (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (>=10=10)

HDDWRKCP - "(Derived) Combined labour market status"

Description: Combined labour market status of cohort member and partner

Population: Cohort members currently living with someone as part of a couple

Value Labels:

- (-8) "Not enough information"
- (-3) "Not asked in fieldwork stage"
- (-1) "Not applicable"
- (1) "Both in work"
- (2) "CM in work; partner not in work"
- (3) "CM not in work; partner in work"
- (4) "Both not in work"
- (5) "CM in work; no partner"
- (6) "CM not in work: no partner"
- (7) "CM status unknown; partner in work"
- (8) "CM status unknown; partner not in work"
- (9) "CM in work; partner status unknown"
- (10) "CM not in work; partner status unknown"
- (11) "CM status unknown; no partner"
- (12) "CM status and partner status unknown"

Derivation description: The working status of the cohort member (ACTIVITY=1,2) and/or partner (PECONACT2=1,2) identifying all possible combinations.

Finance

HDINCB - "(Derived) CM and partner banded weekly income"

Description: Banded weekly income of cohort member and partner, or just CM if no partner

Population: All cohort members

Value Labels:

- (-1) "Not applicable"
- (1) "Less than 30"
- (2) "30 to 50"
- (3) "50 to 90"
- (4) "90 to 110"

- (5) "110 to 140"
- (6) "140 to 210"
- (7) "210 to 270"
- (8) "270 to 310"
- (9) "310 to 350"
- (10) "350 to 400"
- (11) "400 to 500"
- (12) "500 to 550"
- (13) "550 to 650"
- (14) "650 to 800"
- (15) "800 to 950"
- (16) "more than 950"

Derivation description: Cohort member and partner's consolidated banded weekly (TOTINP=1) take-home income. Consolidated bands were assigned as follows:

- (1) less than 30 (TOTINB=1) (2) 30 to 50 (TOTINB=2) (3) 50 to 90 (TOTINB=3) (4) 90 to 110 (TOTINB=4) (5) 110 to 140 (TOTINC=1) (6) 140 to 210 (TOTINC=2) (7) 210 to 270 (TOTINC=3) (8) 270 to 310 (TOTINC=4) (9) 310 to 350 (TOTIND=1) (10) 350 to 400 (TOTIND=2) (11) 400 to 500 (TOTIND=3) (12) 500 to 550 (TOTIND=4) (13) 550 to 650 (TOTINE=1) (14) 650 to 800 (TOTINE=2) (15) 800 to 950 (TOTINE=3) (16) more than 9500 (TOTINE=4)

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

Description: Whether cohort member or partner receives any benefits

Population: All cohort members

Value Labels:

(-8) "Not enough information"

(-1) "Not applicable"

(1) "Yes"

(2) "No" **Derivation description:** Cohort members are categorised as receiving benefits if they reported receiving universal credit (UNCR) or any of the benefits reported in question BENT.

HDFINLIT3 – “(Derived) Financial Literacy – All three questions correct”

Description: Whether cohort member answered all three financial literacy questions correctly

Population: Cohort members who have completed the self-completion section

Value Labels:

- (-8) “Not enough information”
- (-3) “Not asked in fieldwork stage”
- (-1) "Not applicable"
- (1) "Yes"
- (2) "No"

Derivation description: Correct answers for all three of the financial literacy variables: FINLIT1=2, FINLIT2=102, FINLIT3=1

HDFINLITC – “(Derived) Financial Literacy – Number of correct answers”

Description: Number of correct answers to the three financial literacy questions

Population: Cohort members who have completed the self-completion section

Value Labels:

- (-8) “Not enough information”
- (-3) “Not asked in fieldwork stage”
- (-1) "Not applicable"

Derivation description: Sum of the correct answers for the three financial literacy variables below. Refusal/Don’t Know responses are scored as 0.

HDFINLITA – “(Derived) Financial Literacy – Number of questions attempted”

Description: Number of questions attempted for the three financial literacy questions

Population: Cohort members who have completed the self-completion section

Value Labels:

- (-8) “Not enough information”
- (-3) “Not asked in fieldwork stage”
- (-1) "Not applicable"

Derivation description: Sum of the questions attempted for the three financial literacy questions below. Refusal/Don’t Know responses not counted as attempts.

HDFINLITR – “(Derived) Financial Literacy – At least one question refused”

Description: Whether at least one question was refused for the three financial literacy questions

Population: Cohort members who have completed the self-completion section

Value Labels:

- (-8) “Not enough information”
- (-3) “Not asked in fieldwork stage”
- (-1) “Not applicable”
- (1) “Yes”
- (2) “No”

Derivation description: Any refusal (-9) responses from the three financial literacy questions below output 1. Else, output 0.

Education

HDACDQC - “Highest Academic Qualification (Including Currently Studying) in MCS8”

Description: Cohort member’s highest academic qualification either completed or currently studying towards to – self reported

Population: All cohort members with new qualifications in sweep 8

Value Labels:

- (-8) “Insufficient information”
- (-1) “Not applicable”
- (1) “Doctorate (PhD)”
- (2) “Master’s Degree”
- (3) “Post-graduate diplomas and certificates”
- (4) “Bachelor’s Degree”
- (5) “Diploma in higher education”
- (6) “A Level or A2 Level”
- (7) “AS Level”
- (8) “Advanced Higher (Scotland)”
- (9) “Higher (Scotland)”
- (10) “Cambridge Pre-U Certificate”
- (11) “Extended Project Qualification”

- (12) "Skills challenge certificate (Wales)"
- (13) "GCSE"
- (14) "iGCSE"
- (15) "National 5 (Scotland)"
- (16) "National 2, 3, 4 (Scotland)"
- (95) "Non-UK qualifications"
- (96) "None of these qualifications"

HDACDQ - "Highest Academic Qualification obtained in MCS8"

Description: Cohort member's completed highest academic qualification— self reported

Population: All cohort members with new qualification

Value Labels:

- (-8) "Insufficient information"
- (-1) "Not applicable"
- (1) "Doctorate (PhD)"
- (2) "Master's Degree"
- (3) "Post-graduate diplomas and certificates"
- (4) "Bachelor's Degree"
- (5) "Diploma in higher education"
- (6) "A Level or A2 Level"
- (7) "AS Level"
- (8) "Advanced Higher (Scotland)"
- (9) "Higher (Scotland)"
- (10) "Cambridge Pre-U Certificate"
- (11) "Extended Project Qualification"
- (12) "Skills challenge certificate (Wales)"
- (13) "GCSE"
- (14) "iGCSE"
- (15) "National 5 (Scotland)"
- (16) "National 2, 3, 4 (Scotland)"
- (95) "Non-UK qualifications"
- (96) "None of these qualifications"

HDHACDQC - “Highest Academic Qualification (Including Currently Studying) to Age 23”

Description: Cohort member’s highest academic qualification either completed or currently studying towards up to age 23– self reported

Population: All cohort members

Value Labels:

- (-8) “Insufficient information”
- (-1) “Not applicable”
- (1) “Doctorate (PhD)”
- (2) “Master’s Degree”
- (3) “Post-graduate diplomas and certificates”
- (4) “Bachelor’s Degree”
- (5) “Diploma in higher education”
- (6) “A Level or A2 Level”
- (7) “AS Level”
- (8) “Advanced Higher (Scotland)”
- (9) “Higher (Scotland)”
- (10) “Cambridge Pre-U Certificate”
- (11) “Extended Project Qualification”
- (12) “Skills challenge certificate (Wales)”
- (13) “GCSE”
- (14) “iGCSE”
- (15) “National 5 (Scotland)”
- (16) “National 2, 3, 4 (Scotland)”
- (95) “Non-UK qualifications”
- (96) “None of these qualifications”

Derivation description:

This variable indicates the highest academic qualification (either achieved or studying towards to) up to sweep 8 and including in sweep 8. This variable has been derived by merging qualifications reported in sweep 7

HDHACDQ - "Highest Academic Qualification obtained to Age 23"

Description: Cohort member's highest academic qualification attained up to age 23– self reported

Population: All cohort members

Value and Value Labels:

- (-8) "Insufficient information"
- (-1) "Not applicable"
- (1) "Doctorate (PhD)"
- (2) "Master's Degree"
- (3) "Post-graduate diplomas and certificates"
- (4) "Bachelor's Degree"
- (5) "Diploma in higher education"
- (6) "A Level or A2 Level"
- (7) "AS Level"
- (8) "Advanced Higher (Scotland)"
- (9) "Higher (Scotland)"
- (10) "Cambridge Pre-U Certificate"
- (11) "Extended Project Qualification"
- (12) "Skills challenge certificate (Wales)"
- (13) "GCSE"
- (14) "iGCSE"
- (15) "National 5 (Scotland)"
- (16) "National 2, 3, 4 (Scotland)"
- (95) "Non-UK qualifications"
- (96) "None of these qualifications"

Derivation description: This variable indicates the highest academic qualification achieved up to sweep 8 and including in sweep 8. This variable has been derived by merging qualifications reported in sweep 7

HDDEG- “DV: Whether achieved first degree or higher”

Description: Whether cohort member achieved a first degree, PGCE, master’s or PhD

Population: All cohort members in Age 23

Value Labels:

(-8) “Insufficient information”

(-1) “Not applicable”

(0) “No”

(1) “Yes”

Derivation description: Whether the cohort member achieved at least a first degree

HDPDEG- “DV: Whether achieved first post grad degree or higher”

Description: Whether cohort member achieved a PGCE, master’s or PhD

Population: All cohort members in Age 23

Value Labels:

(-8) “Insufficient information”

(-1) “Not applicable”

(0) “No”

(1) “Yes”

Derivation description: Whether the cohort member achieved at least a first degree

Health

HDDISEQ - “(Derived) Disability classification Equality act (2010)”

Description: Whether cohort member is disabled according to the Equality act (2010) definition

Population: All cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

(0) “Not disabled (Equality act)”

(1) “Disabled (Equality act)”

Derivation description: According to this definition, a cohort member is considered disabled if they report a longstanding illness (CLSI) and have a reduced ability to carry out day-to-day activities as a result of their illness (CLSL).

HDDISLS - “(Derived) Disability classification EU-SILC”

Description: Whether cohort member is disabled according to the European Statistics on Income and living conditions (EU-SILC) definition

Population: All cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

(0) “No EU-SILC long-standing health condition”

(1) “EU-SILC classification to some extent”

(2) “EU-SILC classification severely hampered”

Derivation description: According to this definition, a cohort member is considered disabled if they report a longstanding illness (CLSI), have a reduced ability to carry out day-to-day activities as a result of their illness (CLSL), and this reduced ability has lasted for more than 6 months (CLSP). This variable also distinguishes between those that are disabled to some extent, and those that are severely hampered.

HDHGHTM - “(Derived) Height in metres – self reported”

Description: Self-reported height in metres.

Population: All cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: Height in metres reported at Age 23 Metric values are combined (HTMETRES + (HTCMS/100), Imperial measures are converted to metric using the formula: $((HTFEET*12)+ HTINCHES)/39.370$

HDMHGHTM - “(Derived) Height in metres – Measured and self reported combined”

Description: Measured height if CM had height measured at Age 23, and self-reported height if not.

Population: All cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: HTCM, or HDHGHTM if missing

HDWGHTK - ‘(Derived) Weight in kilograms – self reported’

Description: Self-reported weight in kilos.

Population: All cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: Weight in kilos reported at Age 23. Metric weight (WTKILOS), and Imperial measures are converted to metric using the formula:
 $14 * WTSTONE + WTPOUNDS) * 0.453592$

HDMWGHTK - ‘(Derived) Weight in kilograms – Measured and self reported combined’

Description: Measured weight if CM had weight measured at Age 23, and self-reported weight if not.

Population: All cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: WTCM, or HDWGHTK if missing

HDBMI - “(Derived) Body mass index – Based on self-report”

Description: Body mass index of cohort member based on self reported values

Population: All cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: Body mass index (BMI) kg/m² of cohort member derived from height (HDHGHTM) and weight (HDWGHTK).

HDMBMI - “(Derived) Body mass index – Based on measure and self-report combined”

Description: Body mass index of cohort member based on measured and self reported values combined

Population: All cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: Body mass index (BMI) kg/m² of cohort member derived from height (HDMHGHTM) and weight (HDMWGHTK).

HDBMIC - “(Derived) Body mass index classification – Based on self-report”

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

Population: All cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

(1) “Underweight”

(2) “Healthy weight”

(3) “Overweight”

(4) “Obese”

Derivation description: Whether the cohort member was underweight, healthy weight, overweight, or obese, according to World Health Organization (WHO) BMI thresholds. Reference: “BMI Classification” Global Database on Body Mass Index. World Health Organization. 2006.

www.who.int/bmi/index.jsp?introPage=intro_3.html. Retrieved March 22, 2017.

HDMBMIC - “(Derived) Body mass index classification – Based on measure and self-report combined”

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

Population: All cohort members

Value Labels:

- (-8) “Not enough information”
- (-3) “Not asked in fieldwork stage”
- (1) “Underweight”
- (2) “Healthy weight”
- (3) “Overweight”
- (4) “Obese”

Derivation description: Whether the cohort member was underweight, healthy weight, overweight, or obese, according to World Health Organization (WHO) BMI thresholds. Reference: “BMI Classification” Global Database on Body Mass Index. World Health Organization. 2006.

www.who.int/bmi/index.jsp?introPage=intro_3.html. Retrieved March 22, 2017.

HDSMOKE – “(Derived) Smoking habits”

Description: Cohort member smoking habits

Population: All cohort members

Value Labels:

- (-8) “Not enough information”
- (-3) “Not asked in fieldwork stage”
- (-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: Banded number of cigarettes per day for daily smokers.

Indicators for occasional/ex-/non-smokers.

If SMOKING in {1,2,3}, recode SMOKING (1=0,2=1,3=2).

If SMOKING = 4:

If $1 \leq \text{NOFCIGS} \leq 10$: HDSMOKE = 3

If $11 \leq \text{NOFCIGS} \leq 20$: HDSMOKE = 4

If $\text{NOFCIGS} > 20$: HDSMOKE = 5

If NOFCIGS in {-8, -9}: HDSMOKE = 6

HDAUDIT3 - "(Derived) Total AUDIT-C Score"

Description: Cohort member's score on the Alcohol Use Disorders Identification Test Consumption (AUDIT-C) scale.

Population: Cohort members who drink alcohol

Value Labels:

(-8) "Not enough information"

(-3) "Not asked in fieldwork stage"

(-1) "Not applicable"

Derivation description: The responses to the AUDIT-C questions (ALDR, AUDIT2, AUDIT3) were scored as follows: (1=0) (2=1) (3=2) (4=3) (5=4). ALDR is reverse coded. The derived variable is the sum of these scores. The score does not need to be adjusted for different limits for male and female respondents as the values in questions are changed to account for this

HDAUDG3 - "(Derived) AUDIT-C Group"

Description: Whether the Cohort Member's drinking habits are high risk according to the Alcohol Use Disorders Identification Test Consumption (AUDIT-C) scale.

Population: Cohort members who drink alcohol

Value Labels:

(-8) "Not enough information"

(-3) "Not asked in fieldwork stage"

(-1) "Not applicable"

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

(2) “Increasing or high risk drinking (5+)”

Derivation description: The responses to the AUDIT-C grouped into either high risk (HDAUDIT3 \geq 5) or low risk (HDAUDIT3 $<$ 5)

Mental Health

HDKESSLER6 - “(Derived) Kessler K6 scale of non-specific psychological distress score”

Description: Total score of the 6 item Kessler scale from 0-24 with higher scores indicating higher levels of psychological distress.

Population: All Cohort members

Value Labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: The sum of the 6-item Kessler scale, all reverse coded ((5-PHDE), (5-PHHO), (5-PHRF), (5-PHEE), (5-PHWO), (5-PHNE)). If any question is refused or missing then –8

HDWEMWB - “(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

Value labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: The sum of 7 items, each on a 5-point scale (WWOP, WWUS, WWRE, WWDE, WWTH, WWCL, WWMN)

HDLONELINESS – “(Derived) UCLA Loneliness 3-item”

Description: Measurement of loneliness

Population: All cohort members

Value labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: Addition of the scores of three questions (LNLY1, LNLY2, LNLY3). Only computed if all questions have valid responses.

HDGAD2 – “(Derived) Generalised Anxiety Disorder 2-item”

Description: Measurement of the frequency of feeling nervous, anxious or on edge. Scores range from 0-6 with higher scores indicating higher anxiety levels

Population: All cohort members who completed the self-completion

Value labels:

(-8) “Not enough information”

(-3) “Not asked in fieldwork stage”

Derivation description: Addition of the scores of two questions (GAD1, GAD2). Only computed if all questions have valid responses.

HDPHQ2 – “(Derived) Patient Health Questionnaire 2-item”

Description: Measurement of the frequency of depressed mood over the past 2 weeks. Scores range from 0-6 with higher scores indicating higher levels of depression

Population: All cohort members

Value labels:

(-8) “Not enough information”

Derivation description: Addition of the scores of two questions (PHQ1, PHQ2). Only computed if all questions have valid responses.

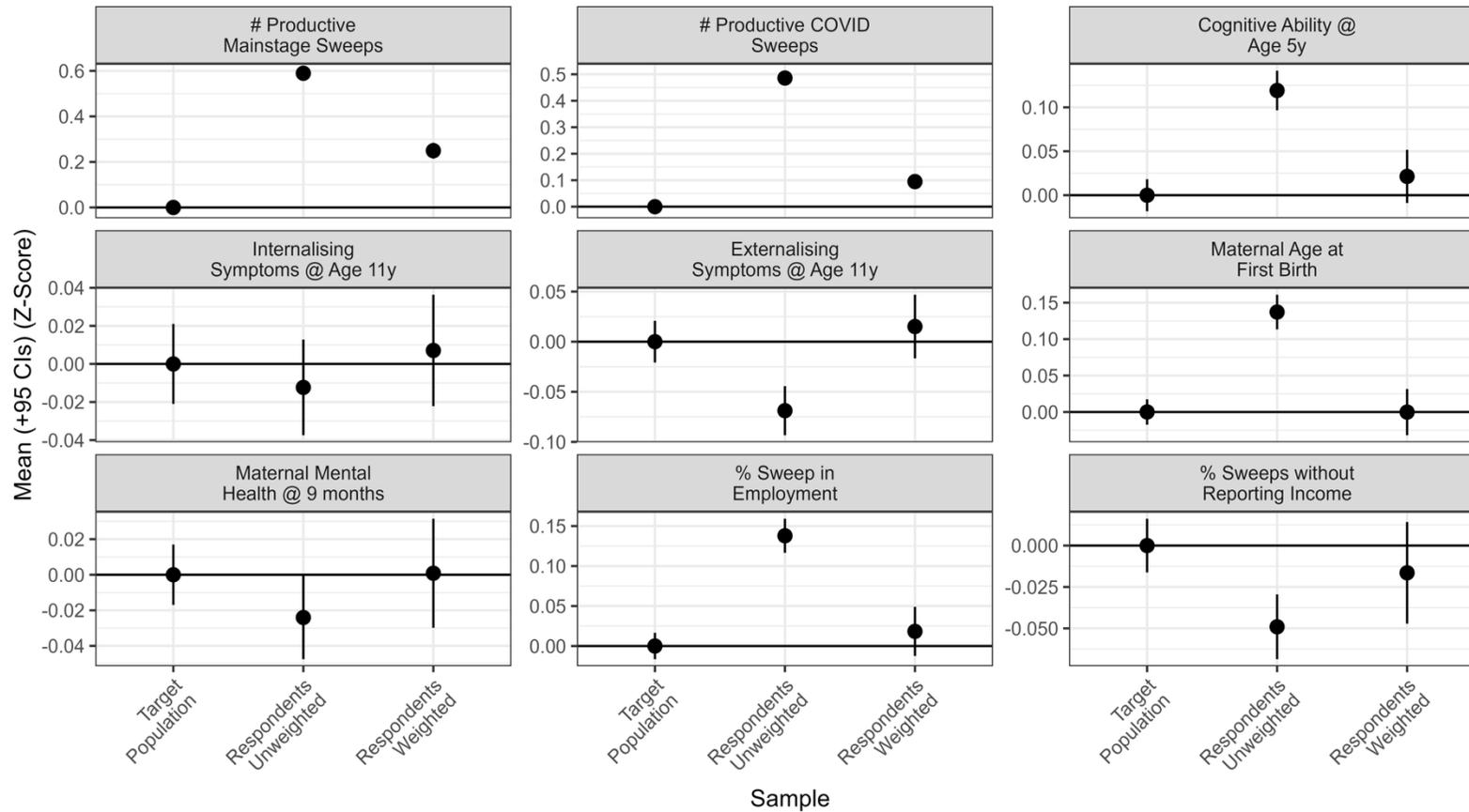
Appendix 2: Response Models

	Variable	OR (+ 95% CI)
Sex (Ref: Male)	Female	1.39 (1.28, 1.5)
	Mixed	1.21 (0.99, 1.46)
Ethnicity (Ref: White)	Indian	1.3 (1.01, 1.69)
	Pakistani & Bangladeshi	1.53 (1.29, 1.83)
	Black or Black British	1.27 (1.02, 1.57)
	Other	1.23 (0.94, 1.61)
MCS Family Type (Ref: Original Family)	New Family	1.86 (1.52, 2.28)
Country @ Sampling (Ref: England)	Wales	0.97 (0.86, 1.08)
	Scotland	0.74 (0.65, 0.84)
	Northern Ireland	1.1 (0.96, 1.26)
	# Productive Mainstage Sweeps	2.11 (2.05, 2.17)
	# Productive COVID Sweeps	3.43 (3.18, 3.7)

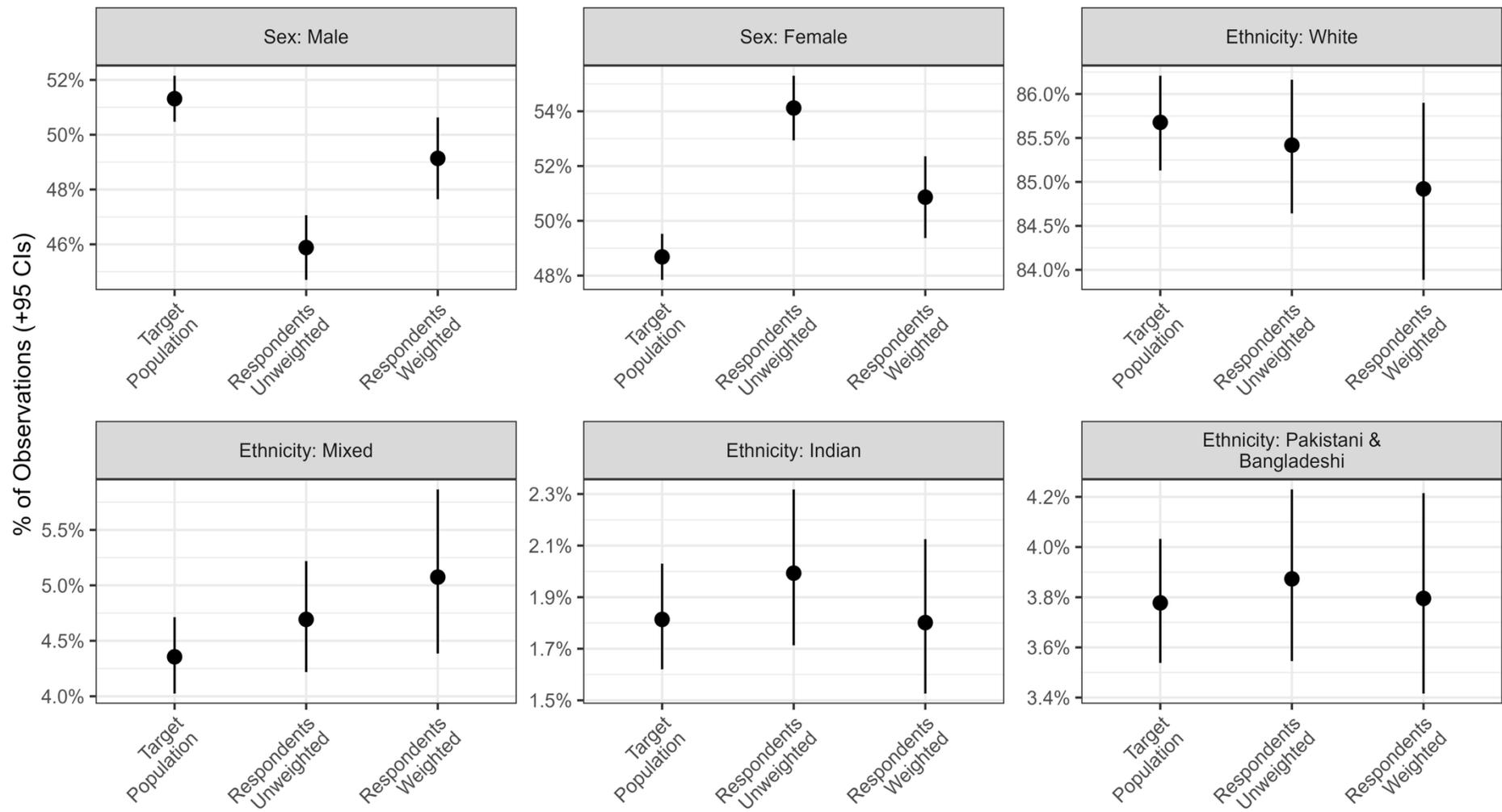
	Variable	OR (+ 95% CI)
	Cognitive Ability @ Age 5y	1.05 (1.01, 1.09)
	Internalising Symptoms @ Age 11y	1.02 (1, 1.04)
	Externalising Symptoms @ Age 11y	0.99 (0.97, 1)
Cohort Member Breastfed (Ref: No)	Yes	1.05 (0.95, 1.15)
	Maternal Age at First Birth	0.99 (0.98, 1)
	Maternal Mental Health @ 9 months	1.01 (0.98, 1.03)
# Parents in Housing @ Age 11y (Ref: Two parents/carers)	One parent/carers	1.01 (0.91, 1.12)
	NVQ level 2	1.07 (0.93, 1.23)
	NVQ level 3	1.01 (0.86, 1.19)
	NVQ level 4	1.1 (0.94, 1.29)
Highest Main Parental Education, 0y-11y (Ref: NVQ level 1)	NVQ level 5	1.31 (0.99, 1.72)
	Overseas Qualification	1.06 (0.82, 1.36)
	None	1.09 (0.92, 1.29)

	Variable	OR (+ 95% CI)
	Intermediate	0.93 (0.82, 1.06)
	Small Employers & Own Account	0.9 (0.74, 1.09)
Highest Parental NS-SEC, 0y-11y (Ref: Managerial & Professional)	Lower Supervisory & Technical	0.96 (0.76, 1.2)
	Semi-Routine & Routine	1.01 (0.88, 1.16)
	Not in Work	1.08 (0.89, 1.31)
	% Sweep in Employment	0.99 (0.82, 1.19)
	Mortgage	0.97 (0.82, 1.15)
	Rent from LA	0.99 (0.8, 1.21)
Housing Tenure @ Age 11y (Ref: Own outright)	Rent from Housing Association	0.95 (0.76, 1.2)
	Rent privately	0.84 (0.68, 1.04)
	Other	0.92 (0.66, 1.29)
Accommodation Type (Ref: House or Bungalow)	Other	1.08 (0.93, 1.25)
	% Sweeps without Reporting Income	0.81 (0.68, 0.97)

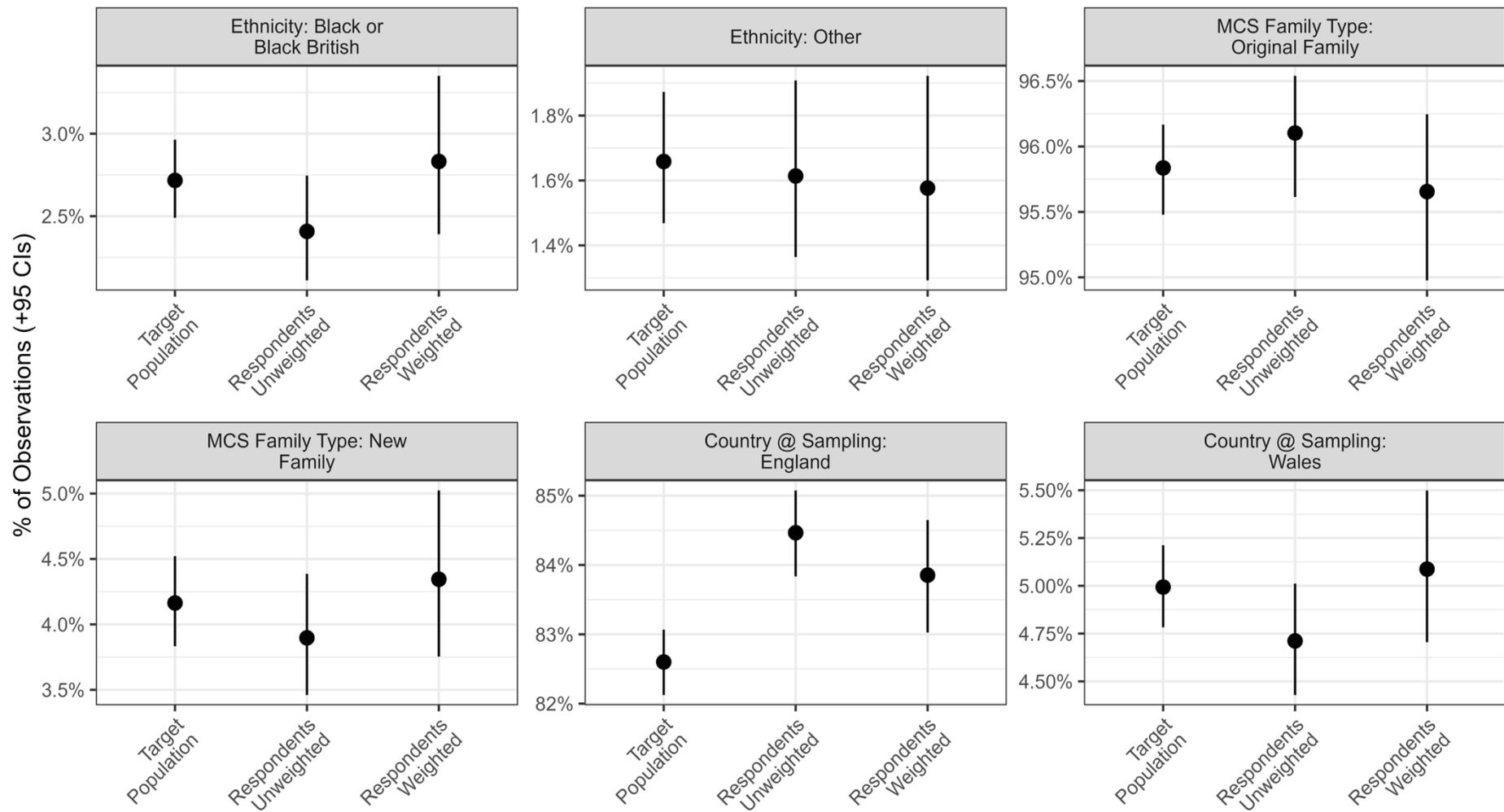
Appendix 3: Performance of Non-Response Weights



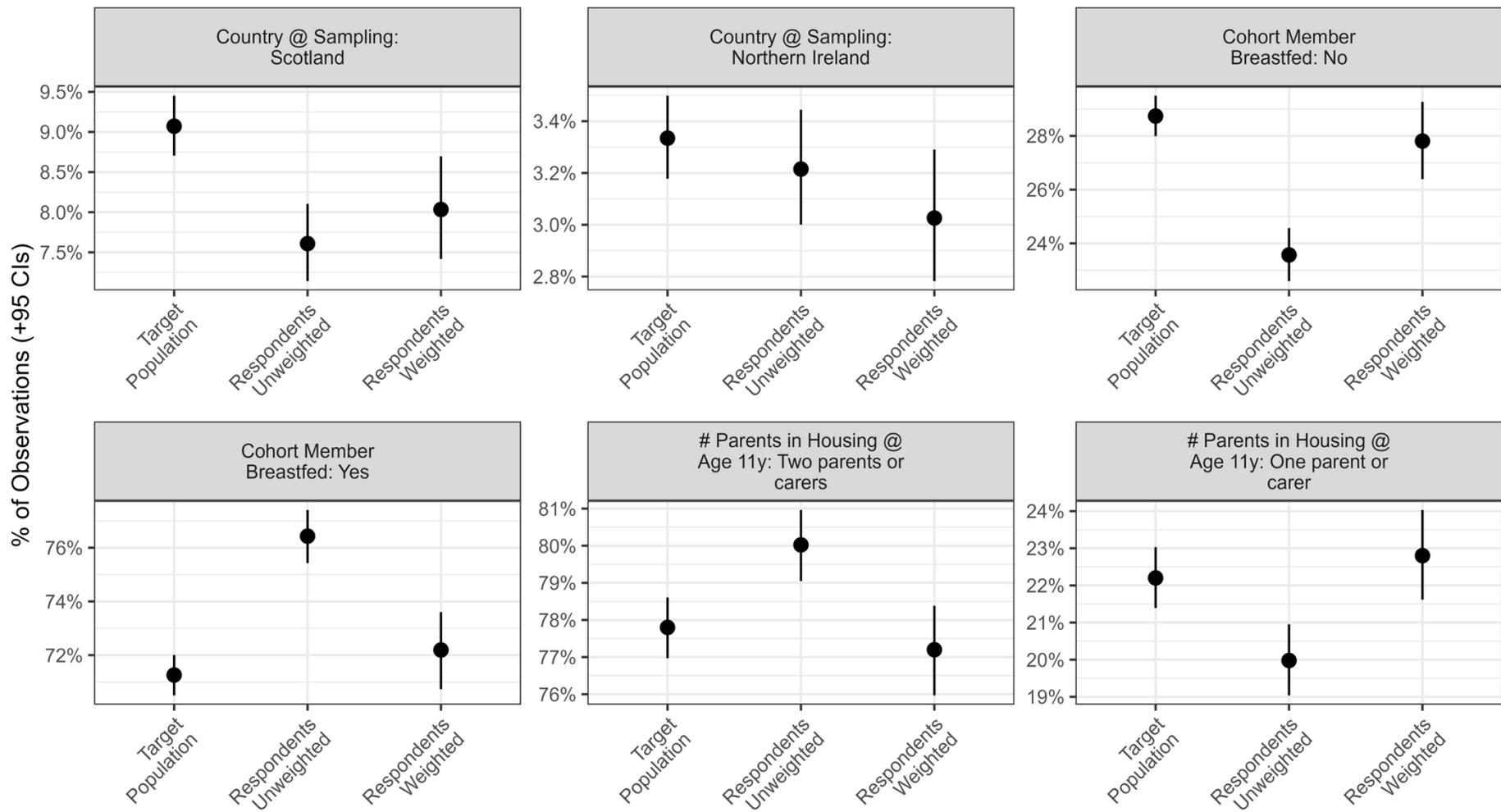
Appendix 3 Figure 1. Means for continuous variables used in the response prediction models – target population and Age 23 respondents, weighted and unweighted (XX-VAR).



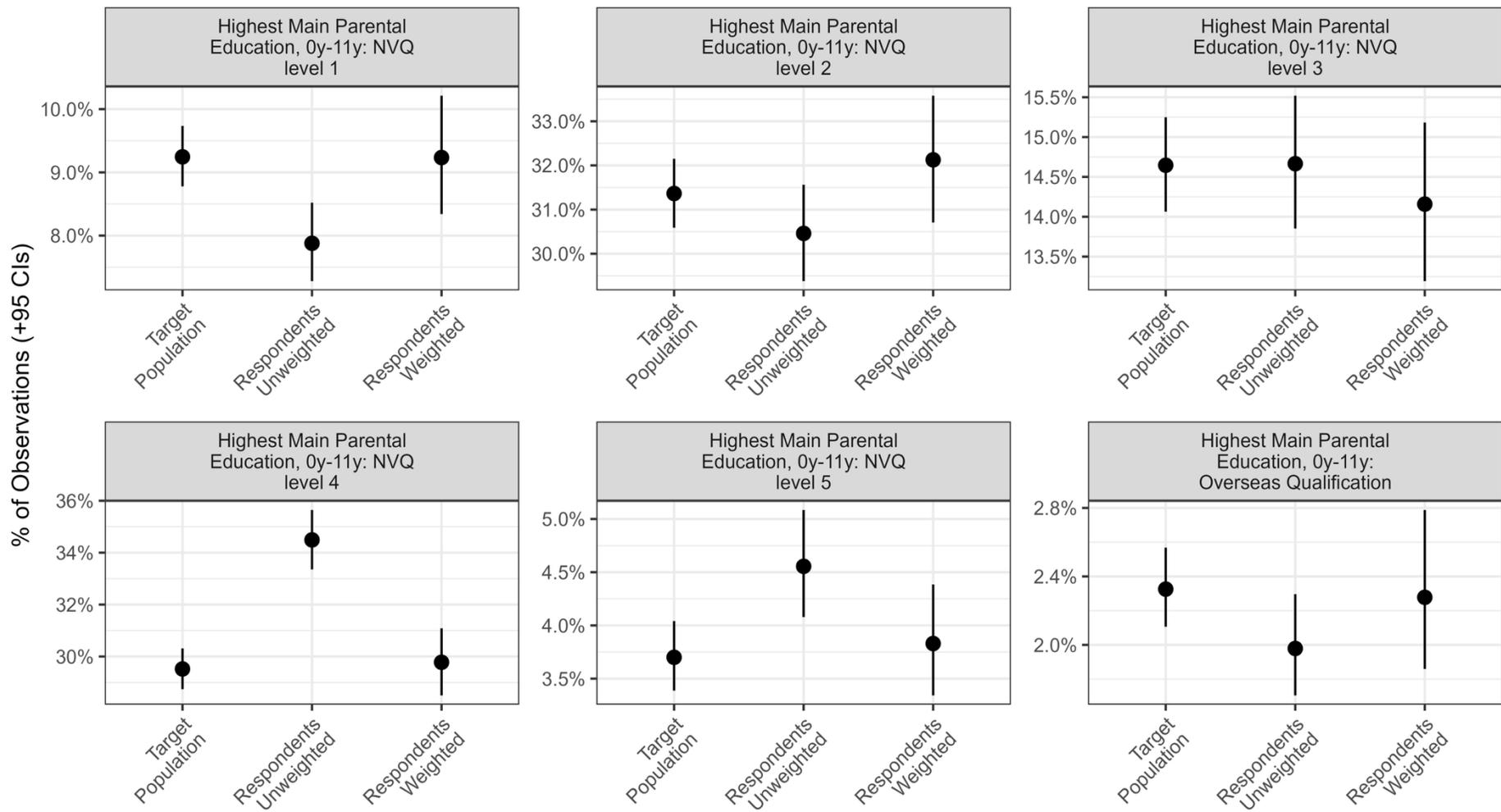
Appendix 3 Figure 2. Proportions for categorical variables used in the response prediction models – target population and Age 23 respondents, weighted and unweighted (XX-VAR). See also Appendix 3 Figures 3-8.



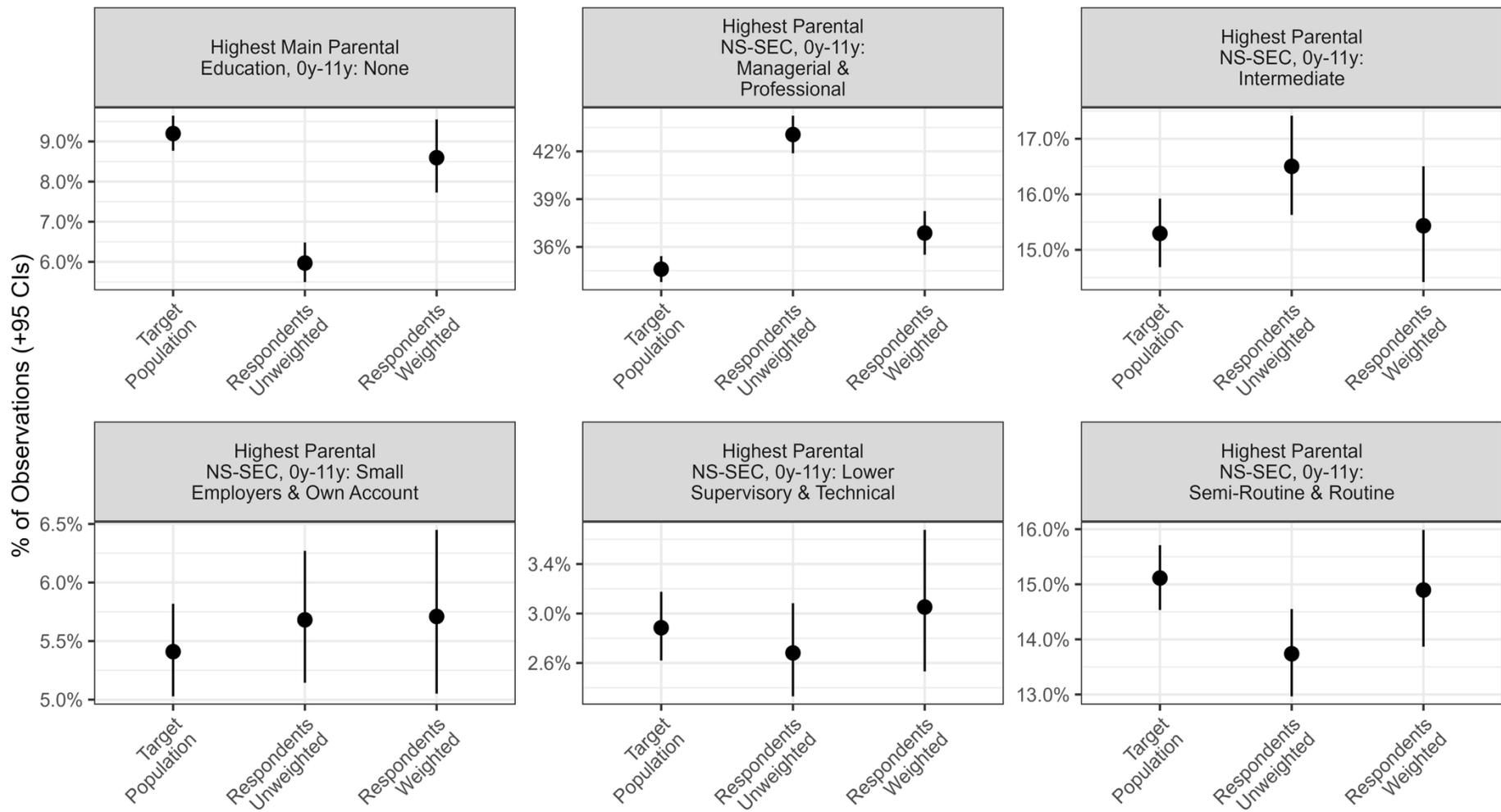
Appendix 3 Figure 3. Proportions for categorical variables used in the response prediction models – target population and Age 23 respondents, weighted and unweighted (XX-VAR). See also Appendix 3 Figures 2 & 4-8.



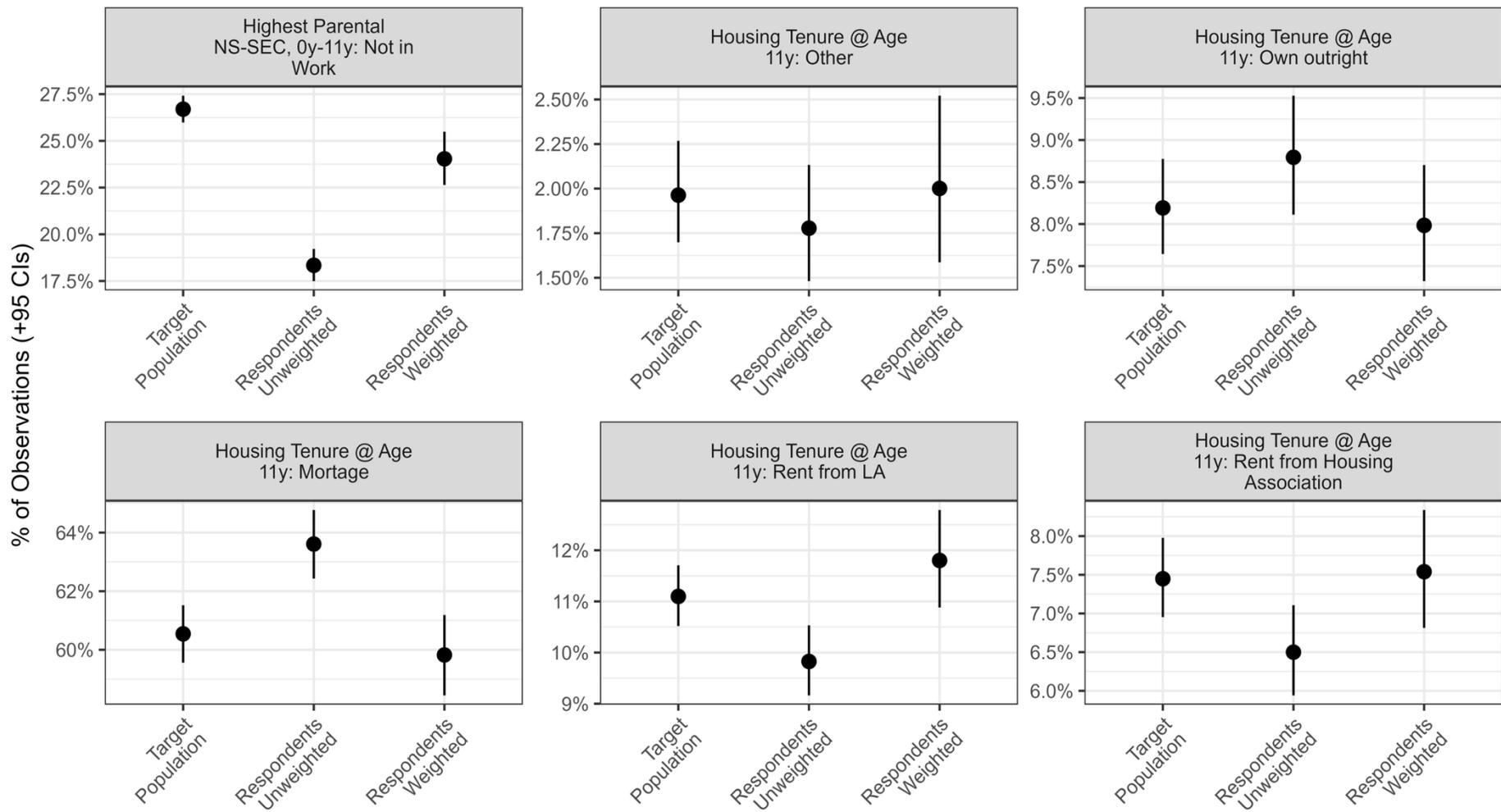
Appendix 3 Figure 4. Proportions for categorical variables used in the response prediction models – target population and Age 23 respondents, weighted and unweighted (XX-VAR). See also Appendix 3 Figures 2-3 & 5-8.



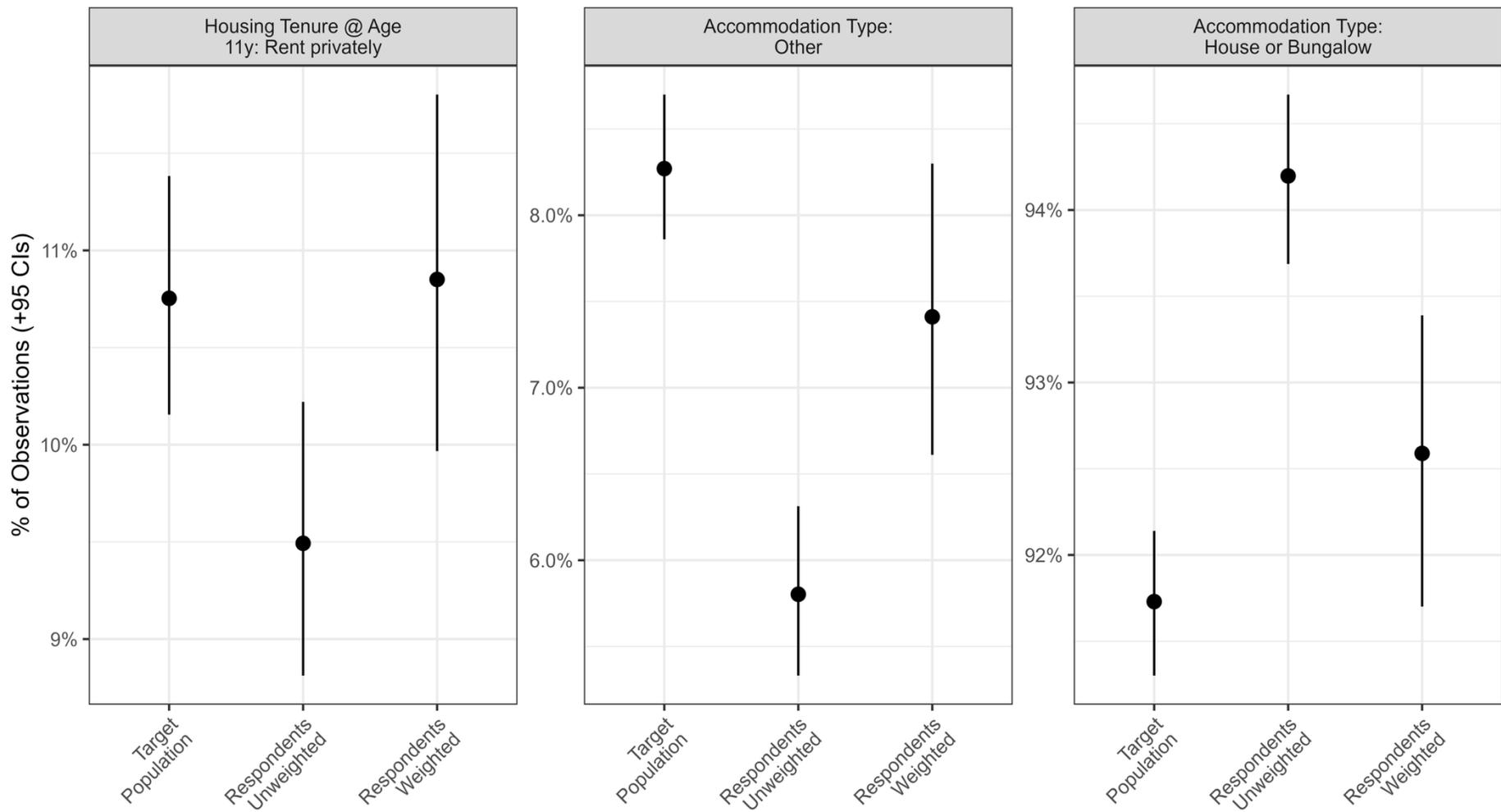
Appendix 3 Figure 5. Proportions for categorical variables used in the response prediction models – target population and Age 23 respondents, weighted and unweighted (XX-VAR). See also Appendix 3 Figures 2-4 & 6-8.



Appendix 3 Figure 6. Proportions for categorical variables used in the response prediction models – target population and Age 23 respondents, weighted and unweighted (XX-VAR). See also Appendix 3 Figures 2-5 & 7-8.



Appendix 3 Figure 7. Proportions for categorical variables used in the response prediction models – target population and Age 23 respondents, weighted and unweighted (XX-VAR). See also Appendix 3 Figures 2-6 & 8.



Appendix 3 Figure 8. Proportions for categorical variables used in the response prediction models – target population and Age 23 respondents, weighted and unweighted (XX-VAR). See also Appendix 3 Figures 2-7.