Next Steps: Core dataset

User guide (Version 1)

October 2026

UCL Centre for Longitudinal Studies

Contact

Data queries: help@ukdataservice.ac.uk

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

Authors

Alison Fang-Wei Wu, Morag Henderson, Martin Danka, David Bann.

How to cite this guide

Wu, A.F.W, Henderson, M, Danka, M, Bann, D. (2025). *Next Steps: Core dataset User guide (Version 1)*. London: UCL Centre for Longitudinal Studies.

Data citation and acknowledgement

You should cite the data and acknowledge CLS following the guidance from Citing our data page on the website.

All outputs using CLS making study easy to use data should cite this user guide.

About UCL Centre for Longitudinal Studies

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.

Next Steps is core funded by:

UKRI Economic and Social Research Council

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

Acknowledgements

This work was supported by the Economic and Social Research Council (ESRC, grant reference number: ES/L000164/1). The CLS cohorts are only possible due to the commitment and enthusiasm of their participants; their time and contribution are gratefully acknowledged. We thank colleagues in the Centre for Longitudinal Studies' Research Data Management Team, Survey, Cohort Maintenance, Administrative, and Communications teams.

Contents

1. About Next Steps 5

2. Introduction to the Next Steps Core dataset 7

3. The Next Steps Core dataset: methodology 8

4. Dataset description and usability 10

5. References 12

6. Appendix 14

Table 1. List of the core dataset variables 14

Table 2. Demographic variables 24

Table 3. Educational attainment variables 29

Table 4. Socioeconomic circumstances variables 35

Table 5. Physical and mental health variables 43

Table 6. Anthropometric variables 47

Table 7. Behaviour variables 49

Table 8. Selected variables with detailed source variables' response category 57

1. About Next Steps

Next Steps is a national cohort study following the lives of around 16,000 people in England, born in 1989-90, since they were in secondary school.

The study collects information about education and employment, economic circumstances, family life, physical and emotional health and wellbeing, social participation and attitudes.

Next Steps, initially launched in 2004 as the Longitudinal Study of Young People in England (LSYPE), follows 15,770 pupils who were in Year 9 in England and born between September 1989 and August 1990. At Sweep 4 (in 2007), there was an ethnic sample boost of 352 pupils, taking the total sample to 16,122.

Conceived to map young people's journeys from compulsory schooling into higher education and employment, it now offers a more holistic view of the UK's "millennial" generation, a cohort positioned between the British Cohort Study (born in 1970) and the Millennium Cohort Study (born in 2001-2) and therefore crucial for tracing social change across time and between cohorts (Wu et al., 2024).

Annual interviews, conducted from ages 14 to 20, captured schooling experiences, aspirations, and early labour-market entry as well as some measures about their life conditions, including health, mental health and the family environment. Subsequent sweeps at ages 25 and 32 have broadened the lens to include health, family, finances, attitudes, and genotyping. The most recent fieldwork, conducted in 2022-23, achieved 7,279 adult interviews. Notably, data for 11 participants were permanently deleted at their request, reducing the overall sample size to 16,111.

Since 2015, Next Steps has been managed by the UCL Centre for Longitudinal Studies and funded by the Economic and Social Research Council.

The study's scientific value rests not only on its repeated surveys but also on rich data linkages to education, health, and administrative records, as well as three CLS-wide COVID-19 surveys and an antibody test, in addition to measures of cognition and genetic data. Oversampling of pupils from deprived schools and ethnic minority backgrounds aids the study of inequality and social mobility.

A vital source of evidence, Next Steps has had a major influence on national education policy and cast light on a wide range of important social issues, including the association between zero-hours contracts and mental health.

More information about the Next Steps cohort study can be found in the sweep user guides through the Next Steps cohort website, and datasets can be accessed through the UK Data Service.

This table presents the issued sample, the achieved sample and the response rate (achieved/issued sample) for each sweep.

Sweep (age)	Year of data collection	Issued	Achieved
1 (14)	2004	21,000	15,770
2 (15)	2005	15,678	13,539
3 (16)	2006	13,525	12,439
4 (17)	2007	12,468	11,449
4 (17) boost sample	2007	600	352
5 (18)	2008	11,793	10,430
6 (19)	2009	11,225	9,799
7 (20)	2010	9,791	8,682
8 (25)	2015	15,531	7,707
9 (32)	2022	13,859	7,279

2. Introduction to the Next Steps Core dataset

CLS is committed to making our rich longitudinal data easier to use. As such, for all our established cohorts, we are creating 'Core' datasets.

These are single datasets containing variables across life spanning multiple domains, including:

- **Demographics** e.g. sex/gender, ethnicity
- **Education attainment**
- **Socioeconomic circumstances** e.g. social class
- **Physical and mental health** e.g. self-rated health wellbeing
- **Anthropometrics** e.g. height and weight
- **Behaviours** e.g. physical activity, alcohol intake

The aim is to help researchers spend less time on data merging and cleaning and more time on substantive analysis.

The code to replicate and extend this work from source survey data is accessible through the CLS GitHub repository, allowing researchers to use, modify, and add to it based on their specific needs.

We envisage that these will be useful for:

- **New users**
- **Established users** e.g. reducing time in the derivation of covariates or other variables of secondary interest, or in speeding up exploratory analyses
- **Teaching** e.g. enabling teaching time to be focused on longitudinal data analysis, rather than recoding/cleaning/merging of data.

These datasets are optimised for within-cohort comparisons; users who wish to conduct comparative analyses across multiple cohorts should ensure that items used are harmonised. We plan to update these datasets in future and welcome user feedback (clsdata@ucl.ac.uk).

3. The Next Steps Core dataset: methodology

Appendix Tables 2-7 show a summary of the variables derived in the Next Steps Core dataset, along with the derivation logic and source datasets/variables. Code to derive these variables is available on GitHub; below shows the overarching process followed. This section describes the overarching process followed to compute the derived variables included the Next Steps Core dataset, including the sequencing of steps, and key decisions made.

The code used to generate all derived variables is openly available on the CLS GitHub syntax repository, allowing users to inspect, reproduce, or adapt the derivation process for their own analytical purposes. The syntax repository includes annotated syntax files covering each derivation step, with references to the relevant source variables. Users are encouraged to review the code if they require a more detailed understanding of any specific derivation.

Tables 2–7 in the Appendix provide a summary of each derived variable, including the derivation logic and the source datasets and variables used. These tables are intended to give users a transparent overview of how each derived variable was constructed and should be consulted alongside the data to aid interpretation.

Overview of Appendix Tables 2-7:

Table	Content
Table 2. Demographics	Sex, ethnicity, language, sexuality, legal marital status, region
Table 3. Education	Own education, parents' education
Table 4. Socioeconomic circumstances	Own economic activity, parents' economic activity, own social class, social class of parents, home ownership, income own, income parents, Index of Multiple Deprivation
Table 5. Physical and mental health	General Health Questionnaire, life satisfaction, self-harm, Generalized Anxiety Disorder 2-item, Patient Health Questionnaire-2, UCLA loneliness, self-rated health, long-standing illness
Table 6. Anthropometrics	Weight, height, Body Mass Index
Table 7. Behaviours	Smoke, alcohol consumption, drug use, exercise, school absences, school suspended, school truancy, police contact, bullying

3.1 Variables included in the core dataset

Variables were selected across core domains: demographics, education attainment, socioeconomic circumstances, physical and mental health, anthropometrics and behaviours. Their inclusion was informed by our knowledge of Next Steps, patterns of data users in prior work, and anticipated future research needs.

In addition to core variables, the dataset includes non-response weights, sample type (main vs. boost), sweep-level participation flags, and data availability indicators (e.g., permanent deletions). User guides of Sweeps 1-9 provide details on these variables and information on how to incorporate them into the analyses (Adali et al., 2025; Calderwood et al., 2021; Department of Education, 2011). These variables can be used, alongside other strategies (e.g., multiple imputation) to reduce bias due to missing data. CLS provides detailed guidance on handling missing data across its cohort studies (see (R. Silverwood et al., 2024)).

The variable names and labels are in Appendix Table 1.

3.2 Longitudinal consistency

To support analyses across time, we derived consistent response categories where possible:

- When identical response categories existed across ages, we retained them unchanged.
- When categories differed, we collapsed or re-mapped them to a common scheme; we additionally provide the original versions to aid users (e.g., in sensitivity analyses).
- Where using the same response categories was not possible (e.g., substantially different instruments, such as parental income collected with different bands), we provide multiple variables with clearly labelled, age-specific categories.

4. Dataset description and usability

4.1 Licensing and data access

The Next Steps Core dataset is available from the UK Data Service (UKDS) as safeguarded data.

All users of the data need to be registered with the UKDS. Details of how to do this are available on the UK Data Service webpage.

The Next Steps Core dataset can be downloaded from the UKDS once the End User License (EUL) access conditions have been accepted by the user.

4.2 Dataset and data structure

The Next Steps Core dataset (**ns_core**) is a flat (wide format) file. i.e., where one record exists for each cohort member (CM).

The variable names and labels are in Appendix Table 1.

4.3 Code used to derived variables

To support transparency and reproducibility, the code used to derive all variables documented in this guide is available on the CLS GitHub syntax repository. The repository contains the scripts used to clean, recode, and harmonise the source data into the derived variables described here, allowing users to inspect the derivations and, where they have access to the underlying data, reproduce the construction of the dataset.

4.4 Individual identifiers

All Next Steps datasets use a consistent research identifier (NSID) as the primary key across all available cohort data from the UK Data Service.

4.5 Naming conventions

All variable names are short, descriptive, and lowercase (e.g., eth = ethnicity). Age at measurement is appended as a numeric suffix (e.g., eth14 = ethnicity at age 14).

We round-up ages for consistency across sweeps (e.g., the Sweep 1 sample was recruited in Year 9 (aged 13–14); age 14 was used to represent this sweep).

4.6 Missing values

The missing-value codes are coded using minus values, for example:

-1 = Item not applicable

-2 = Script error/information lost

-3 = Not asked at the fieldwork stage/did not participate at specific wave/ was not surveyed

-5 = data not available

-7 = prefer not to say

-9 = Refusal

-8 = Don’t know/insufficient information

This means that users can flag missing values across all variables with a single command (e.g., in stata: “mvdecode _all, mv(-9 -8 -7 -3 -2 -1)”).

5. References

Adali, T., Rihal, S., Wright, L., Silverwood, R., Brown, M., Henderson, M., Sanchez-Galvez, A., & Curran, L. (2025). *Next Steps: Sweep 9 Age 32 User Guide (Version 2)*.

Calderwood, L. , Peycheva, D. , Henderson, M. , Silverwood, R. , Mostafa, T. , & Rihal, S. (2021). *Next Steps Sweep 8-Age 25 Survey User Guide (3rd edition)*.

Carpenter, J. R., Bartlett, J. W., Morris, T. P., Wood, A. M., Quartagno, M., & Kenward, M. G. (2023). *Multiple imputation and its application*. John Wiley & Sons.

Centre for Longitudinal Studies. (2024). COVID-19 Survey in Five National Longitudinal Cohort Studies: Millennium Cohort Study, Next Steps, 1970 British Cohort Study and 1958 National Child Development Study, 2020-2021. In *UK Data Service*.

Department of Education. (2011). *LSYPE User Guide to the Datasets: Wave 1 to Wave 7*.

Goldberg, D. P., & Williams, P. (1988). *A users' guide to the general health questionnaire*. Nfer-Nelson.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2003). The Patient Health Questionnaire-2. *Medical Care*, *41*(11), 1284–1292. https://doi.org/10.1097/01.MLR.0000093487.78664.3C

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety Disorders in Primary Care: Prevalence, Impairment, Comorbidity, and Detection. *Annals of Internal Medicine*, *146*(5), 317–325. https://doi.org/10.7326/0003-4819-146-5-200703060-00004

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

Silverwood, R. J., Calderwood, L., Henderson, M., Sakshaug, J. W., & Ploubidis, G. B. (2024). A data-driven approach to understanding non-response and restoring sample representativeness in the UK Next Steps cohort. *Longitudinal and Life Course Studies*, *15*(2), 227–250. https://doi.org/10.1332/17579597Y2024D000000010

Silverwood, R., Narayanan, M., Dodgeon, B., Katsoulis, M., & Ploubidis, G. (2024). *Handling missing data in the CLS cohort studies User guide*. www.cls.ucl.ac.uk.

Wu, A. F.-W., Henderson, M., Brown, M., Adali, T., Silverwood, R. J., Peycheva, D., & Calderwood, L. (2024). Cohort Profile: Next Steps—the longitudinal study of people in England born in 1989–90. *International Journal of Epidemiology*, *53*(6). https://doi.org/10.1093/IJE/DYAE152.

6. Appendix

Table 1. List of the core dataset variables

Variable name	Variable label
NSID	NSID - cohort member identifier
sex	Sex (age 14-32y)
eth	Ethnicity (age 14-32y)
lang	Language spoken at home (age 14-17y)
sori19	Sexuality (age 19y)
sori20	Sexuality (age 20y)
sori25	Sexuality (age 25y)
sori32	Sexuality (age 32y)
partnr19	Legal marital status simple (age 19y)
partnr25	Legal marital status simple (age 25y)
partnr32	Legal marital status simple (age 32y)
partnradu25	Legal marital status complete (age 25y)
partnradu32	Legal marital status complete (age 32y)
regub15	Region urban (age 15y)
regov15	Region government area (age 15y)
regub16	Region urban (age 16y)
regov16	Region government area (age 16y)
regor25	Region government area (age 25y)
regor32	Region government area (age 32y)
regint32	Living abroad (age 32y)
educaim17	Current qualification studied level 3-category (age 17y)
educaim19	Current qualification studied level 3-category (age 19y)

Variable name	Variable label
educaim20	Current qualification studied level 3-category (age 20y)
educaim25	Current qualification studied level 3-category (age 25y)
educaim32	Current qualification studied level 3-category (age 32y)
educ25	Highest education level 3-category (age 25y)
educ32	Highest education level 3-category (age 32y)
educadtl32	Highest education detail - academic (age 32y)
educvdtl32	Highest education detail - vocational (age 32y)
educma	Highest education level mother 3-category (CM age 14-17y)
educpa	Highest education level father 3-category (CM age 14-17y)
educdtlma	Highest education detail mother (CM age 14-17y)
educdtlpa	Highest education detail father (CM age 14-17y)
ecoact17	Economic activity simple (age 17y)
ecoact18	Economic activity simple (age 18y)
ecoact19	Economic activity simple (age 19y)
ecoact20	Economic activity simple (age 20y)
ecoact25	Economic activity simple (age 25y)
ecoact32	Economic activity simple (age 32y)
ecoactadu25	Economic activity detailed (age 25y)
ecoactadu32	Economic activity detailed (age 32y)
ecoactdtma14	Economic activity mother (CM age 14y)
ecoactdtma15	Economic activity mother (CM age 15y)
ecoactdtma16	Economic activity mother (CM age 16y)
ecoactdtma17	Economic activity mother (CM age 17y)
ecoactdtpa14	Economic activity father (CM age 14y)
ecoactdtpa15	Economic activity father (CM age 15y)

Variable name	Variable label
ecoactdtpa16	Economic activity father (CM age 16y)
ecoactdtpa17	Economic activity father (CM age 17y)
nssec17	NS-SEC 17-category (age 17y)
nssec18	NS-SEC 17-category (age 18y)
nssec19	NS-SEC 17-category (age 19y)
nssec20	NS-SEC 17-category (age 20y)
nssec25	NS-SEC 17-category (age 25y)
nssec32	NS-SEC 17-category (age 32y)
nssecma14	NS-SEC mother 17-category (CM age 14y)
nssecpa14	NS-SEC father 17-category (CM age 14y)
nssecma15	NS-SEC mother 17-category (CM age 15y)
nssecpa15	NS-SEC father 17-category (CM age 15y)
nssecma16	NS-SEC mother 17-category (CM age 16y)
nssecpa16	NS-SEC father 17-category (CM age 16y)
nssecma17	NS-SEC mother 17-category (CM age 17y)
nssecpa17	NS-SEC father 17-category (CM age 17y)
nssecma18	NS-SEC mother 17-category (CM age 18y)
nssecpa18	NS-SEC father 17-category (CM age 18y)
hown14	Household ownership (age 14y)
hown15	Household ownership (age 15y)
hown16	Household ownership (age 16y)
hown17	Household ownership (age 17y)
hown18	Household ownership (age 18y)
hown19	Household ownership (age 19y)
hown20	Household ownership (age 20y)

Variable name	Variable label
hown25	Household ownership (age 25y)
hown32	Household ownership (age 32y)
hownteen14	Household ownership detailed (age 14y)
hownteen15	Household ownership detailed (age 15y)
hownteen16	Household ownership detailed (age 16y)
hownteen17	Household ownership detailed (age 17y)
hownteen18	Household ownership detailed (age 18y)
hownteen19	Household ownership detailed (age 19y)
hownteen20	Household ownership detailed (age 20y)
inc25	Banded weekly income of cohort member and partner (pounds, age 25y)
inc32	Banded weekly income of cohort member and partner (pounds, age 32y)
incwhh14	Banded weekly parents' gross salary (pounds, CM age 14y)
incwhh15	Banded weekly parents' gross salary (pounds, CM age 15y)
incwhhcnt14	Weekly parents' gross salary (pounds, continuous, CM age 14y)
incwhhcnt15	Weekly parents' gross salary (pounds, continuous, CM age 15y)
incwhh16	Banded weekly parents' gross salary (pounds, CM age 16y)
incwhh17	Banded weekly parents' gross salary (pounds, CM age 17y)
imd15	2004 Index of Multiple Deprivation (IMD) (age 15y)
imd16	2004 Index of Multiple Deprivation (IMD) (age 16y)
imd32	2019 Index of Multiple Deprivation (IMD) (age 32y)
ghq15	General Health Questionnaire (GHQ12) symptom count (age 15y)
ghq17	General Health Questionnaire (GHQ12) symptom count (age 17y)
ghq25	General Health Questionnaire (GHQ12) symptom count (age 25y)
ghq32	General Health Questionnaire (GHQ12) symptom count (age 32y)
ghqtl15	General Health Questionnaire (GHQ12) score (age 15y)

Variable name	Variable label
ghqtl17	General Health Questionnaire (GHQ12) score (age 17y)
ghqtl25	General Health Questionnaire (GHQ12) score (age 25y)
ghqtl32	General Health Questionnaire (GHQ12) score (age 32y)
lsat20	Life satisfaction (age 20y)
lsat25	Life satisfaction (age 25y)
lsat32	Life satisfaction (age 32y)
sharm25	Self-harm (age 25y)
gad32	GAD 2-item score (age 32y)
phq32	PHQ 2-item score (age 32y)
lon32	UCLA Loneliness 3-item (age 32y)
wt0	Weight (kilograms, parent reported, age 0y)
wt25	Weight (kilograms, age 25y)
wt32	Weight (kilograms, age 32y)
ht25	Height (metres, age 25y)
ht32	Height (metres, age 32y)
ht25_32	Height (metres, age 25y or 32y)
bmi25	Body Mass Index (BMI) (age 25y)
bmi32	Body Mass Index (BMI) (age 32y)
gheateen15	Self-rated general health adolescent 4-point (age 15y)
gheateen16	Self-rated general health adolescent 4-point (age 16y)
gheateen17	Self-rated general health adolescent 4-point (age 17y)
gheaadu25	Self-rated general health adult 5-point (age 25y)
gheaadu32	Self-rated general health adult 5-point (age 32y)
ghea15	Self-rated general health binary (age 15y)
ghea16	Self-rated general health binary (age 16y)

Variable name	Variable label
ghea17	Self-rated general health binary (age 17y)
ghea25	Self-rated general health binary (age 25y)
ghea32	Self-rated general health binary (age 32y)
lsi14_15	Long-term illness binary (age 14-15y)
lsi17	Long-term illness binary (age 17y)
lsi19	Long-term illness binary (age 19y)
lsi20	Long-term illness binary (age 20y)
lsi25	Long-term illness binary (age 25y)
lsi32	Long-term illness binary (age 32y)
smknw14	Smoking now binary (age 14y)
smknw15	Smoking now binary (age 15y)
smknw16	Smoking now binary (age 16y)
smknw25	Smoking now binary (age 25y)
smknw32	Smoking now binary (age 32y)
smk14	Smoking ever and frequency (age 14y)
smk15	Smoking ever and frequency (age 15y)
smk16	Smoking ever and frequency (age 16y)
smk25	Smoking ever and frequency (age 25y)
smk32	Smoking ever and frequency (age 32y)
alcfst	First time reported having alcohol (age 14-32y)
alcfreq14	Alcohol frequency (age 14y)
alcfreq15	Alcohol frequency (age 15y)
alcfreq16	Alcohol frequency (age 16y)
alcfreq17	Alcohol frequency (age 17y)
alcfreq19	Alcohol frequency (age 19y)

Variable name	Variable label
alcfreq20	Alcohol frequency (age 20y)
audita25	AUDIT A (age 25y)
audita32	AUDIT A (age 32y)
auditb25	AUDIT B (age 25y)
auditb32	AUDIT B (age 32y)
auditc25	AUDIT C (age 25y)
auditc32	AUDIT C (age 32y)
drgcnbevr	Cannabis ever used (age 14-32y)
drgcnbfst	First time reported ever used cannabis (age 14-32y)
drgcnbnw19	Cannabis current use (age 19y)
drgcnbnw20	Cannabis current use (age 20y)
drgcnbnw25	Cannabis current use (age 25y)
drgcnbnw32	Cannabis current use (age 32y)
drgothevr	Other drugs ever used (age 19-32y)
drgothfst	First time reported ever used other drugs (age 19-32y)
drgothnw19	Other drugs current use (age 19y)
drgothnw20	Other drugs current use (age 20y)
drgothnw25	Other drugs current use (age 25y)
drgothnw32	Other drugs current use (age 32y)
spt14	Exercise frequency (age 14y)
spt15	Exercise frequency (age 15y)
spt17	Exercise frequency (age 17y)
spt19	Exercise frequency (age 19y)
spt20	Exercise frequency (age 20y)
spt25	Exercise frequency (age 25y)

Variable name	Variable label
spt32	Exercise frequency (age 32y)
abs1m14	Absence from school over 1 month in the past 12 months (age 14y)
abs1m15	Absence from school over 1 month in the past 12 months (age 15y)
abs1m16	Absence from school over 1 month in the past 12 months (age 16y)
susp14	Suspended from school (age 14y)
susp15	Suspended from school (age 15y)
susp16	Suspended from school (age 16y)
susp17	Suspended from school (age 17y)
expl14	Expelled from school (age 14y)
expl15	Expelled from school (age 15y)
expl16	Expelled from school (age 16y)
expl17	Expelled from school (age 17y)
trua14	Truancy within past three years (age 14y)
trua15	Truancy in the past year (age 15y)
trua16	Truancy in the past year (age 16y)
trua17	Truancy in the past year (age 17y)
bul14	Bullying victimisation (age 14y)
bul15	Bullying victimisation (age 15y)
bul16	Bullying victimisation (age 16y)
bul17	Bullying victimisation (age 17y)
bul20	Bullying victimisation (age 20y)
bul25	Bullying victimisation (age 25y)
pol14	Police contact binary (age 14y)
pol15	Police contact binary (age 15y)
pol16	Police contact binary (age 16y)

Variable name	Variable label
pol17	Police contact binary (age 17y, only boost sample)
polcnt14	Police contact count (age 14y)
polcnt15	Police contact count (age 15y)
polcnt16	Police contact count (age 16y)
polcnt17	Police contact count (age 17y, only boost sample)
polwrn25	Police warning (age 25y)
polwrn32	Police warning (age 32y)
polars25	Police arrest (age 25y)
polars32	Police arrest (age 32y)
polcau25	Police caution (age 25y)
polcau32	Police caution (age 32y)
polglt25	Police found guilty (age 25y)
polglt32	Police found guilty (age 32y)
polpnd25	Given a Penalty Notice for Disorder (age 25y)
polpnd32	Given a Penalty Notice for Disorder (age 32y)
samppsu	Sampling: School (primary sampling unit)
sampstratum	Sampling: Stratum
designweight	Weight: Design weight
mainboost	Whether main or boost sample
w1outcome	Fieldwork outcome at Wave 1
w1finwt	Weight: Design weight * NR weights (trimmed for independent schools)
w2outcome	Fieldwork outcome at Wave 2
w2finwt	Weight: W2 final weight (W1 final weight * W2 non-response weight)
w3outcome	Fieldwork outcome at Wave 3
w3finwt	Weight: W3 final weight (W2 final weight * W3 non-response weight)

Variable name	Variable label
w2tow3nrwt	Weight: Non-response weight from Wave 2 to Wave 3
w4outcome	Fieldwork outcome at Wave 4
w4weight_main_boost	Weight: Cross-sectional weight for main and boost sample respondents
w4weight_main	Weight: Longitudinal weight for W1 to W4 respondents
w5outcome	Fieldwork outcome at Wave 5
w5finwt_cross	Weight: W5 cross-sectional weight including boost respondents
w6outcome	Fieldwork outcome at Wave 6
w6finwt_cross	Weight: W6 cross-sectional weight all respondents
w6lsypewt	Weight: Respondents who took part in Wave 5 and Wave 6
w6lsypew4wt	Weight: Respondents who took part in Wave 4 and Wave 6, but not Wave 5
w7outcome	Fieldwork outcome at Wave 7
w7_lsype_wt	Weight: Respondents who took part in Wave 5, Wave 6 and Wave 7
w7finwt	Weight: W7 final weight (W6 final weight * W7 non-response weight)
w7_lsype_wt_skiponly	Weight: Respondents who took part in Wave 6 and Wave 7, but not Wave 5
w8outcome	Fieldwork outcome at Wave 8
w8finwt	Weight: W8 final weight
w9outcome	Fieldwork outcome at Wave 9
w9finwt	Weight: W9 final weight
data_availability	Whether the data is available for research use

Table 2. Demographic variables

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Sex	**sex** **"Sex (age 14-32y)"** 0= male 1= female	Sex in Sweeps 1 to 4 were recorded by the survey interviewer, while sex in Sweep 9 (age 32) was collected directly from the cohort members. Thus, sex at age 32 was used (reported from cohort members). If not available, used the earliest response. If this was missing across sweeps, we used the earliest available observed value.	**W1sexYP** (wave_one_lsype_young_person_2020) **W2SexYP** (wave_two_lsype_young_person_2020) **W3sexYP** (wave_three_lsype_young_person_2020) **W4SexYP** (wave_four_lsype_young_person_2020) **W5SexYP** (wave_five_lsype_young_person_2020) **W6Sex** (wave_six_lsype_young_person_2020) **W7Sex** (wave_seven_lsype_young_person_2020) **W8CMSEX** (ns8_2015_main_interview) **W9DSEX** (ns9_2022_main_interview)	**W1sexYP; W2SexYP; W3sexYP; W4SexYP; W5SexYP; W6Sex; W7Sex; W8CMSEX; W9DSEX** 1= Male 2= Female
Ethnicity	**eth** **"Ethnicity (age 14-32y)"** 1= White-British 2= White-Irish 3= Any other White background 4= Mixed-White and Black Caribbean 5= Mixed-White and Black African 6= Mixed-White and Asian 7= Any other mixed background 8= Asian/Asian British – Indian	Ethnicity was selected from the earliest valid self-reported response across the available ethnicity measures in Sweeps 1, 2, 4, 8, and 9. If this was missing across those sweeps, we used the earliest available observed value.	**W1ethnic2YP** (wave_one_lsype_young_person_2020) **W2ethnicYP** (wave_two_lsype_young_person_2020) **w4ethnic2YP** (wave_four_lsype_young_person_2020) **W8DETHN15** (ns8_2015_derived) **W9DETHN15** (ns9_2022_derived_variables)	**W1ethnic2YP; W2ethnicYP; w4ethnic2YP; W8DETHN15; W9DETHN15** 1= White-British 2= White-Irish 3= Any other White background 4= Mixed-White and Black Caribbean 5= Mixed-White and Black African 6= Mixed-White and Asian 7= Any other mixed background 8= Asian/Asian British – Indian

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	9= Asian/Asian British Pakistani 10= Asian/Asian British – Bangladeshi 11= Any other Asian background 12= Black/Black British -Caribbean 13= Black/Black British – African 14= Any other Black background 15= Chinese 16= Any other background			9= Asian/Asian British Pakistani 10= Asian/Asian British – Bangladeshi 11= Any other Asian background 12= Black/Black British -Caribbean 13= Black/Black British – African 14= Any other Black background 15= Chinese 16= Any other background
Language	**lang** **"Language spoken at home (age 14-17y)"** 1= English only 2= English first/main and speaks other languages 3= Another language is respondent's first or main language 4= Bilingual	Language was selected from the earliest valid response across Sweeps 1 to 4 for the original sample and Sweep 4 for the boost sample, using young-person reports in Sweeps 1 and 2 and household reports in Sweeps 3 and 4. If this was missing across the available sweeps, we used the earliest available observed value.	**W1englangYP** (wave_one_lsype_young_person_2020) **W2EnglangYP** (wave_two_lsype_young_person_2020) **W3englangHH** (wave_three_lsype_family_background_2020) **W4EngLangHH** (wave_four_lsype_family_background_2020)	**W1englangYP; W2EnglangYP** 1= Yes-English only 2= Yes-English first/main and speaks other languages 3= No, another language is respondent's first or main language 4= Respondent is bilingual **W3englangHH; W4EngLangHH** 1= Yes-English only 2= Yes-English first/main and speaks other languages 3= No, another language is household's first or main language 4= household is bilingual
Sexuality	**sori19; sori20; sori25; sori32** **"Sexuality (age yy)"** 1= Heterosexual/straight 2= Gay/lesbian 3= Bisexual 4= Other	Original values were retained, except for one response category in Sweep 9 (5 = prefer not to say), which was recoded to the standard missing value code for prefer not to say.	**W6SexualityYP** (wave_six_lsype_young_person_2020) **W7SexualityYP** (wave_seven_lsype_young_person_2020) **W8SEXUALITY** (ns8_2015_self_completion) **W9SORI** (ns9_2022_main_interview)	**W6SexualityYP; W7SexualityYP; W8SEXUALITY** 1= Heterosexual/straight 2= gay/lesbian 3= bisexual 4= other

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				W9SORI 1= Heterosexual/straight 2= gay/lesbian 3= bisexual 4= other 5= prefer not to say
Legal marital status	**partnr19; partnr25; partnr32** **"Legal marital status simple (age yy)"** 0= Single and never married or in a CP 1= Married/civil partner 2= Separated but still legally married/in a CP 3= Divorced/former CP 4= Widowed/surviving CP **partnradu25; partnradu32** **"Legal marital status complete (age yy)"** 0= Single and never married or in a CP 1= Married 2= Separated but still legally married 3= Divorced 4= Widowed 5= A civil partner 6= Separated but still legally in a CP 7= A former civil partner 8= A surviving civil partner	Two types of variables were created. A simple version for all sweeps allows comparison across sweeps. A more detailed version for Sweeps 8 and 9 where additional information is available, ensuring no loss of detail.	**W6MarStatYP** (wave_six_lsype_young_person_2020) **W8DMARSTAT** (ns8_2015_derived) **W9DMARSTAT** (ns9_2022_derived_variables)	**W6MarStatYP** 1= single, that is never married 2= married 3= Separated 4= Divorced 5= Widowed **W8DMARSTAT; W9DMARSTAT** 1= Single and never married or in a CP 2= married 3= separated but still legally married 4= divorced 5= widowed 6= a civil partner 7= separated but still legally in a CP 8= a former civil partner 9= a surviving civil partner

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Region	**regub15; regub16** **"Region urban (age yy)"** 1= Urban >=10k – sparse 2= Town & Fringe – sparse 3= Village – sparse 4= Hamlet and Isolated Dwelling – sparse 5= Urban >= 10k - less sparse 6= Town & Fringe - less sparse 7= Village - less sparse 8= Hamlet and Isolated Dwelling – less sparse **regov15; regov16** **"Region government area (age yy)"** 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 **regor25; regor32** **"Region government area (age yy)"** 1= North East 2= North West 3= Yorkshire and The Humber	Original values were retained except for regint32 (see below). Notably, although Sweeps 2, 3, 8 and 9 had a region government area variable, the response categories were different between Sweeps 2/3 and Sweeps 8/9, with Sweeps 8/9 having more areas. To maximise the information, we allowed varied response categories between these two sets of variables to ensure no loss of detail. regint32 was derived by collapsing W9NATIONRES into In the UK versus Outside of UK or unknown.	**urbind; gor (age 15)** (wave_two_lsype_family_background_2020) **urbind; gor (age 16)** (wave_three_lsype_family_background_2020) **W8DGOR** (ns8_2015_derived) **W9DRGN** (ns9_2022_derived_variables) **W9NATIONRES** (ns9_2022_main_interview)	**urbind** 1= Urban >=10k – sparse 2= Town & Fringe – sparse 3= Village – sparse 4= Hamlet and Isolated Dwelling – sparse 5= Urban >= 10k - less sparse 6= Town & Fringe - less sparse 7= Village - less sparse 8= Hamlet and Isolated Dwelling – less sparse **gor** 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 **W8DGOR; W9DRGN** 1= North East 2= North West 3= Yorkshire and the Humber 4= East Midlands 5= West Midlands 6= East of England

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	4= East Midlands 5= West Midlands 6= East of England 7= London 8= South East 9= South West 10= Wales 11= Scotland 12= Northern Ireland **regint32** **"Living abroad (age 32y)"** 1= In the UK 2= Abroad or unknown			7= London 8= South East 9= South West 10= Wales 11= Scotland 12= Northern Ireland 13= Unknown due to faulty **W9NATIONRES** 1= England 2= Scotland 3= Wales 4= Northern Ireland 5= Outside of UK or unknown

Table 3. Educational attainment variables

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Current qualification studied, own	**educaim17; educaim19; educaim20; educaim25; educaim32** **"Current qualification studied level 3-category (age yy)"** 0= NVQ 4-5 1= NVQ 1-3 2= None/entry 3= Other 4= None of these qualifications 5= Not studying **Key differences between 2, 3, and 4 categories: ▪ None/entry (2): No qualifications or only entry-level. ▪ Other (3): Matches the "Other" category in the original survey; respondent has qualifications outside the predefined list. ▪ None of these qualifications (4): Matches the "None of these qualifications" category in the original survey; respondent may have qualifications but explicitly indicated none fit the listed options.	To allow cross-sweep comparison, the new response categories followed the categorisation of the CLS COVID-19 cross-cohort survey (Centre for Longitudinal Studies, 2024). For ages 25 and 32, activity variables were also used to identify respondents who were not studying.	**w4saim** (wave_four_lsype_young_person_2020) **W6Saim** (wave_six_lsype_young_person_2020) **W7SAim** (wave_seven_lsype_young_person_2020) **W8ACTIVITY05; W8ACQUC0A-0Q; W8VCQUC0A-0R** (ns8_2015_main_interview) **W9ECONACT2; W9ACQUC0A-0V; W9VCQUC0A-AI** (ns9_2022_main_interview)	See Table 8.

Education, own	**educ25; educ32** **"Highest education level 3-category (age yy)"** 0= NVQ 4-5 1= NVQ 1-3 2= None/entry 3= Other 4= None of these qualifications **Key differences between 2, 3, and 4 categories: ▪ None/entry (2): No qualifications or only entry-level. ▪ Other (3): Matches the "Other" category in the original survey; respondent has qualifications outside the predefined list. ▪ None of these qualifications (4): Matches the "None of these qualifications" category in the original survey; respondent may have qualifications but explicitly indicated none fit the listed options. **educadtl32** **"Highest education detail - academic (age 32)"** 1= Doctorate or equivalent 2= Masters or equivalent 3= Undergraduate or equivalent 4= Post-graduate Diplomas and Certificates 5= Diplomas in higher education and other higher education qualifications	To allow cross-sweep comparison, the new response categories followed the categorisation of the CLS COVID-19 cross-cohort survey (Centre for Longitudinal Studies, 2024).	**W8VCQU0A-0R** (ns8_2015_main_interview) **W8DHANVQH** (ns8_2015_derived) **W9DANVQH; W9DVNVQH** (ns9_2022_derived_variables) **W9ACQU0A-0V; W9VCQU0A-AI** (ns9_2022_main_interview)	See Table 8.

	6= Teaching qualifications for schools or further education 7= A/AS Levels or equivalent 8= GCSE - Grade A-C, Level 4-9 9= GCSE - Grade D-G, Level 1-3 10= Scottish Qualifications - SCE Higher 11= Scottish Qualifications - Scottish Certificate Sixth Year Studies 12= Scottish Qualifications - SCE Standard 13= Scottish Qualifications - National 4 and 5 14= Scottish Qualifications - National 2 and 3 15= Irish Qualifications - Leaving Certificate 16= Irish Qualifications - Junior Certificate grade A-C 17= Irish Qualifications - Junior Certificate grade D and below 18= Other academic qualifications (including overseas) 19= None of these qualifications **educvdtl32** **“Highest education detail - vocational (age 32)”** 1= Professional qualifications at degree level 2= Nursing or other medical qualifications 3= NVQ or SVQ - Level 4 or 5 4= NVQ or SVQ - Level 3 5= NVQ or SVQ - Level 2 6= NVQ or SVQ - Level 1			

	7= GNVQ - GNVQ Advanced 8= GNVQ - GNVQ Intermediate 9= GNVQ - Level 3 10= GNVQ - Level 2 11= GNVQ - Level Foundation 12= City & Guilds - advanced craft, Part III 13= City & Guilds - Craft, Part II 14= City & Guilds - Craft, Part I 15= City & Guilds - Level 3 16= City & Guilds - Level 2 17= City & Guilds - Level 1 18= RSA - Advanced Diploma 19= RSA - Higher Diploma 20= RSA - RSA Diploma 21= RSA - RSA Stage I, II,III 22= BTEC - Higher Level BTEC 23= BTEC - BTEC National 24= BTEC - BTEC First 25= SCOTVEC - SCOTVEC National Certificate 26= SCOTVEC - SCOTVEC first or general diploma 27= SCOTVEC - SCOTVEC general diploma 28= SCOTVEC - SCOTVEC modules 29= HND or HNC 30= OND or ONCM 31= Junior certificate 32= Other vocational qualifications (including some overseas) 33= None of these qualifications			

Education, parents	**educpa** **"Highest education level 3-category father (CM age 14-17y)"** **educma** **"Highest education level 3-category mother (CM age 14-17y)"** 0= NVQ 4-5 1= NVQ 1-3 2= None/entry 3= other 4= None of these qualifications **Key differences between 2, 3, and 4 categories: ▪ None/entry (2): No qualifications or only entry-level. ▪ Other (3): Matches the "Other" category in the original survey; respondent has qualifications outside the predefined list. ▪ None of these qualifications (4): Matches the "None of these qualifications" category in the original survey; respondent may have qualifications but explicitly indicated none fit the listed options. **educdtlpa** **"Highest education detail father (CM age 14-17y)"** **educdtlma** **"Highest education detail mother (CM age 14-17y)"**	Parental education was selected from the earliest valid parents' reported response across the available sweeps (1, 2, and 4) for the original sample and Sweep 4 for the boost sample. To allow cross-sweep and parent-children comparison, the new response categories followed the categorisation of the CLS COVID-19 cross-cohort survey (Centre for Longitudinal Studies, 2024).	**W1hiqualmum; W1hiqualdad** (wave_one_lsype_family_background_2020) **W2hiqualmum; W2hiqualdad** (wave_two_lsype_family_background_2020) **w4hiqualmum; w4hiqualdad** (wave_four_lsype_family_background_2020)	See Table 8.

	1= Higher Degree 2= First Degree 3= HE Diploma 4= HNC/HND/NVQ4 5= Teaching qualification, non-degree 6= Nursing qualification, non-degree 7= A Levels 8= OND/ONC 9= City and guilds part III, NVQ3 10= CSYS 11= Scottish Higher Grade 12= AS Level 13= Trade apprenticeship 14= City and guilds part II, NVQ2 15= GCSE grade A-C and equivalent 16= GCSE grade D-E and equivalent 17= City and guilds part I, NVQ1 18= Youth training, skill seekers 19= Qualification, level unspecified 20= No qualification mentioned			

Table 4. Socioeconomic circumstances variables

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Economic activity, own	**ecoact17; ecoact18; ecoact19; ecoact20; ecoact25; ecoact32** **"Economic activity simple (age yy)"** 1= In paid work 2= Apprenticeship/government training scheme/training 3= Education 4= Unemployed 5= Looking after home 6= Other **ecoactadu25; ecoactadu32** **"Economic activity detailed (age yy)"** 1= employee – in paid work 2= self employed 3= voluntary work 4= unemployed 5= education 6= apprenticeship 7= Government employment scheme 8= sick/disabled 9= looking after home 10= something else	Two types of economic activity variables were created. A simple version for all sweeps allows comparison across sweeps. A more detailed version for Sweeps 8 and 9 where additional information is available, ensuring no loss of detail.	**W4empsYP** (wave_four_lsype_young_person_2020) **W5mainactYP** (wave_five_lsype_young_person_2020) **W6TCurrentAct** (wave_six_lsype_young_person_2020) **W7TCurrentAct** (wave_seven_lsype_young_person_2020) **W8DACTIVITYC** (ns8_2015_derived) **W9DACTIVITYC** (ns9_2022_derived_variables)	See Table 8.

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Economic activity, parents	**ecoactdtpa14; ecoactdtpa15; ecoactdtpa16; ecoactdtpa17** **"Economic activity father (CM age yy)"** **ecoactdtma14; ecoactdtma15; ecoactdtma16; ecoactdtma17** **"Economic activity mother (CM age yy)"** 1= FT paid work 2= PT paid work 3= Unemployed 4= training 5= education 6= looking after the family 7= Retired from work altogether 8= Sick/disabled 9= other	Original values were retained.	**W1empsmum; W1empsdad** (wave_one_lsype_family_background_2020) **W2empsmum; W2empsdad** (wave_two_lsype_family_background_2020) **W3empsmum; W3empsdad** (wave_three_lsype_family_background_2020) **W4empsmum; W4empsdad** (wave_four_lsype_family_background_2020)	**W1empsmum; W1empsdad; W2empsmum; W2empsdad; W3empsmum; W3empsdad; W4empsmum; W4empsdad** 1= Doing paid work for 30 or more hours 2= Doing paid work for fewer than 30 hours a week 3= Unemployed/looking for a job 4= On a training course or scheme 5= In full-time education/at school 6= looking after the family/household 7= Retired from work altogether 8= Sick/disabled 9= other

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Social class (National Statistics Socio-economic classification, NS-SEC) Own	**nssec17; nssec18; nssec19; nssec20; nssec25; nssec32** **"NS-SEC 17-category (age yy)"** 1= Employers in large organisations 2= Higher managerial and administrative occupations 3= Higher professional occupations 4= Lower professional and higher technical occupations 5= Lower managerial and administrative occupations 6= Higher supervisory occupations 7= Intermediate occupations 8= Employers in small establishments 9= Own account workers 10= Lower supervisory occupations 11= Lower technical occupations 12= Semi-routine occupations 13= Routine occupations 14= Never worked and Long-term unemployed 15= Full-time students 16= Not classified or inadequately stated 17= Not classifiable for other reasons	To allow comparison across sweeps, 17 categories were used. At ages 25 and 32, activity variables were also used to identify full-time students.	**W4nsseccatYP** (wave_four_lsype_young_person_2020) **W5nsseccatYP** (wave_five_lsype_young_person_2020) **w6nsseccatYP** (wave_six_lsype_young_person_2020) **W7NSSECCat** (wave_seven_lsype_young_person_2020) **W8DNSSEC17, W8DACTIVITYC** (ns8_2015_derived) **W9NSSEC** (ns9_2022_main_interview) **W9DACTIVITYC** (ns9_2022_derived_variables)	See Table 8.

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Social class of parents (National Statistics Socio-economic classification, NS-SEC)	**nssecpa14; nssecpa15; nssecpa16; nssecpa17; nssecpa18** **"NS-SEC father 17-category (CM age yy)"** **nssecma14; nssecma15; nssecma16; nssecma17; nssecma18** **"NS-SEC mother 17-category (CM age yy)"** 1= Employers in large organisations 2= Higher managerial and administrative occupations 3= Higher professional occupations 4= Lower professional and higher technical occupations 5= Lower managerial and administrative occupations 6= Higher supervisory occupations 7= Intermediate occupations 8= Employers in small establishments 9= Own account workers 10= Lower supervisory occupations 11= Lower technical occupations 12= Semi-routine occupations 13= Routine occupations 14= Never worked and Long-term unemployed 15= Full-time students 16= Not classified or inadequately stated 17= Not classifiable for other reasons	To allow comparison across sweeps and between parents and cohort members, 17 categories were used.	**W1nsseccatmum; W1nsseccatdad** (wave_one_lsype_family_background_2020) **W2nsseccatmum; W2nsseccatdad** (wave_two_lsype_family_background_2020) **W3cnsseccatmum; W3cnsseccatdad** (wave_three_lsype_family_background_2020) **w4cnsseccatmum; w4cnsseccatdad** (wave_four_lsype_family_background_2020) **w5Cnsseccatmum; w5Cnsseccatdad** (wave_five_lsype_family_background_2020)	See Table 8.

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Home ownership	**hown14; hown15; hown16; hown17; hown18; hown19; hown20; hown25; hown32** **"Household ownership (age yy)"** 1= Own outright 2= Own, buying with help of mortgage/loan 3= Part rent, part mortgage 4= Rent it 5= Live rent-free 6= Other **hownteen14; hownteen15; hownteen16; hownteen17; hownteen18; hownteen19; hownteen20** **"Household ownership detailed (age yy)"** 1= Owned outright 2= Being bought on a mortgage/bank loan 3= Shared ownership (owns & rents property) 4= Rented from a Council or New Town 5= Rented from a Housing Association, 6= Rented privately, 7= Rent free, 8= Some other arrangement.	Two types of variables were created. A simple version for all sweeps allows comparison across sweeps. A more detailed version for Sweeps 1-7 where additional information is available, ensuring no loss of detail.	**W1hous12HH** (wave_one_lsype_family_background) **W2Hous12HH** (wave_two_lsype_family_background_2020) **W3hous12HH** (wave_three_lsype_family_background_2020) **W4Hous12HH** (wave_four_lsype_family_background_2020) **W5Hous12HH; W5Hous12BHH; W5Hous12CHH** (wave_five_lsype_family_background_2020) **W6Hous12YP; W6Hous12bYP; W6Hous12cYP** (wave_six_lsype_young_person_2020) **W7Hous12YP; W7Hous12bYP; W7Hous12cYP** (wave_seven_lsype_young_person_2020) **W8TENURE** (ns8_2015_main_interview) **W9DTENURE** (ns9_2022_derived_variables)	See Table 8.
Income own	**inc25; inc32** **"Banded weekly take-home income of cohort member and partner (pounds, age yy)"** 1= less than £25 per week 2= 25-50 3= 50-90	Original values were retained.	**W8DINCB** (ns8_2015_derived) **W9DINCB** (ns9_2022_derived_variables)	**W8DINCB; W9DINCB** 1= less than 25 2= 25-50 3= 50-90 4= 90-140 5= 140-240 6= 240-300

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	4= 90-140 5= 140-240 6= 240-300 7= 300-350 8= 350-400 9= 400-500 10= 500-600 11= 600-700 12= 700-800 13= 800-900 14= 900-1200 15= 1200-1400 16= more than 1400			7= 300-350 8= 350-400 9= 400-500 10= 500-600 11= 600-700 12= 700-800 13= 800-900 14= 900-1200 15= 1200-1400 16= more than 1400
Income parents	**incwhh14; incwhh15** **"Banded weekly parents' gross salary (pounds, CM age yy)"** 1= less than 25 per week 2= 25-50 3= 50-90 4= 90-140 5= 140-240 6= 240-300 7= 300-350 8= 350-400 9= 400-500 10= 500-600 11= 600-700 12= 700-800	To allow parent-child comparison, the parents' income in Sweeps 1 and 2 was banded in the way of cohort member's. However, the parents' income in Sweeps 3 and 4 were banded and not able to be transformed. Therefore, the original values were retained.	**W1GrsswkHH** (wave_one_lsype_family_background_2020) **W2GrsswkHH** (wave_two_lsype_family_background_2020) **W3incestw** (wave_three_lsype_family_background_2020) **w4IncEstW** (wave_four_lsype_family_background_2020)	**W1GrsswkHH; W2GrsswkHH** (continuous) **W3incestw; w4IncEstW** 1= up to 49 2= 50-99 3= 100-199 4= 200-299 5= 300-399 6= 400-499 7= 500-599 8= 600-699

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	13= 800-900 14= 900-1200 15= 1200-1400 16= more than 1400 **incwhhcnt14; incwhhcnt15** **"Weekly parents' gross salary (pounds, continuous, CM age yy)"** (continuous) **incwhh16; incwhh17** **"Banded weekly parents' gross salary (pounds, CM age yy)"** 1= up to 49 2= 50-99 3= 100-199 4= 200-299 5= 300-399 6= 400-499 7= 500-599 8= 600-699 9= 700-799 10= 800-899 11= 900-999 12= 1000 or more			9= 700-799 10=800-899 11= 900-999 12= 1000 or more

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Index of Multiple Deprivation (IMD)	**imd15; imd16** **"2004 Index of Multiple Deprivation (IMD) (age yy)"** (continuous, higher scores indicates greater deprivation.) **imd32** **"2019 Index of Multiple Deprivation (IMD) (age 32)"** (ordinal decile, 1= Most deprived decile, 10= Least deprived decile)	Original values were retained. Notably, the Index of IMD 2004 and IMD 2019 share the same overall framework but differ in key ways that limit direct comparability. While both measure relative deprivation across seven domains, the underlying indicators, data sources, and some domain definitions have been updated to reflect policy changes and data availability. Additionally, changes in geographical boundaries (LSOAs) between the 2001 and 2011 Censuses affect area-level comparisons. As a result, IMD scores from 2004 and 2019 should not be directly compared; any analysis of change over time should rely on relative rankings.	**IMDRSCORE (age 15)** (wave_two_lsype_family_background_2020) **IMDRSCORE (age 16)** (wave_three_lsype_family_background_2020) **W9DIMDD** (ns9_2022_derived_variables)	**IMDRSCORE (age 15/16)** Scores derived from geographical information using postcode lookup. **W9DIMDD** Rank derived from geographical information using postcode lookup.

Table 5. Physical and mental health variables

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
General Health Questionnaire (GHQ) (Goldberg & Williams, 1988)	**ghq15; ghq17; ghq25; ghq32** **"GHQ-12 12-item score (0-12) (age yy)"** (continuous, higher scores indicate poorer mental health) **ghqtl15; ghqtl17; ghqtl25; ghqtl32** **"GHQ-12 sum score of each item (12-48) (age yy)"** (continuous, higher scores indicate poorer mental health)	Two types of variables were created. First, each item was recoded into 0 (no to seldom having the symptom) or 1 (usually to always having the symptom) and then was summed up to reflect the number of psychological distress symptoms. The second set is the GHQ Likert score (12–48), capturing the severity of symptoms. Both scoring methods are commonly used in the literature and provide complementary insights into mental health status.	**W2concenYP; W2nosleepYP; W2usefulYP; W2decideYP; W2strainYP; W2difficYP; W2activYP; W2probsYP; W2depressYP; W2noconfYP; W2wthlessYP; W2happyYP; W2ghq12scr** (wave_two_lsype_young_person_2020) **W4ConcenYP; W4NoSleepYP; W4UsefulYP; W4DecideYP; W4StrainYP; W4DifficYP; W4ActivYP; W4ProbsYP; W4DepressYP; W4NoConfYP; W4WthlessYP; W4HappyYP; W4ghq12scr** (wave_four_lsype_young_person_2020) **W8GHQ12_1:12** (ns8_2015_self_completion) **W9GHQ12_1:12** (ns9_2022_main_interview) **W8DGHQSC** (ns8_2015_derived) **W9DGHQSC** (ns9_2022_derived_variables)	See Table 8.
Life satisfaction	**lsat20; lsat25; lsat32** **"Life satisfaction (age yy)"** 1= very satisfied 2= fairly satisfied 3= neither satisfied nor dissatisfied 4= fairly dissatisfied 5= very dissatisfied	Original values were retained.	**W7OSatisYP** (wave_seven_lsype_young_person_2020) **W8OSATIS** (ns8_2015_self_completion) **W9OSatis** (ns9_2022_main_interview)	**W7OSatisYP; W8OSATIS; W9OSatis** 1= very satisfied 2= fairly satisfied 3= neither satisfied nor dissatisfied 4= fairly dissatisfied 5= very dissatisfied

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Self-harm	**sharm25** **"Self-harm (age 25y)"** 0= never 1= less often 2= once a year 3= several times a year 4= once a month 5= two or more times a month 6= once a week 7= several times a week 8= once a day 9= more than once a day	Self-harm was only measured at age 25. We combined the two measures (W8HARM: whether self-harmed and W8HARM2: the frequency of self-harm in the past year) in to a single variable and reverse the scores with higher scores indicating greater severity.	**W8HARM; W8HARM2** (ns8_2015_self_completion)	**W8HARM** 1= yes 2= no **W8HARM2** 1= more than once a day 2= once a day 3= several times a week 4= once a week 5= two or more times a month 6= once a month 7= several times a year 8= once a year 9= less often
Generalized Anxiety Disorder 2-item (GAD-2) (Kroenke et al., 2007)	**gad32** **"GAD 2-item score (age 32y)"** (continuous, higher scores indicate more anxious)	GAD-2 was only measured at age 32. Original values were retained.	**W9DGAD2** (ns9_2022_derived_variables)	See Table 8.
Patient Health Questionnaire-2 (PHQ-2) (Kroenke et al., 2003)	**phq32** **"PHQ 2-item score (age 32y)"** (continuous, higher scores indicate more depressed)	PHQ-2 was only measured at age 32. Original values were retained.	**W9DPHQ2** (ns9_2022_derived_variables)	See Table 8.
UCLA loneliness (Russell, 1996)	**lon32** **"UCLA Loneliness 3-item (age 32y)"** (continuous, higher scores indicate lonelier)	UCLA Loneliness scale was only measured at age 32. Original values were retained.	**W9DLONELINESS** (ns9_2022_derived_variables)	See Table 8.

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Self-rated health	**ghea15; ghea16; ghea17; ghea25; ghea32** **"Self-rated general health binary (age yy)"** 1= Poor/not good at all/not very good/fair 0= Good/fairly good/very good/excellent **gheateen15; gheateen16; gheateen17** **"Self-rated general health adolescent 4-point (age yy)"** 1= Very good 2= Fairly good 3= Not very good 4= Not good at all **gheaadu25; gheaadu32** **"Self-rated general health adult 5-point (age yy)"** 1= Excellent 2= Very good 3= Good 4= Fair 5= Poor	We included three sets of variables in the core data. The first set of variables are binary variables, allowing cross-sweep comparisons. We also kept the original values as the other two sets (ages 15-17 using a 4-point scale; ages 25 and 32 using a 5-point scale) to ensure no loss of detail.	**W2hea1cYP** (wave_two_lsype_young_person_2020) **W3hea1cYP** (wave_three_lsype_young_person_2020) **W4Hea1CYP** (wave_four_lsype_young_person_2020) **W8GENA** (ns8_2015_main_interview) **W9HLTHGEN** (ns9_2022_main_interview)	**W2hea1cYP; W3hea1cYP; W4Hea1CYP** 1= Very good 2= Fairly good 3= Not very good 4= Not good at all **W8GENA; W9HLTHGEN** 1= Excellent 2= Very good 3= Good 4= Fair 5= Poor
Long-standing illness	**lsi14_15** **"Long-term illness binary (age 14-15y)"** **lsi17; lsi19; lsi20; lsi25; lsi32** **"Long-term illness binary (age yy)"** 0= No 1= Yes	Original values were reversely recoded. Notably, at age 15, the original survey only asked those who did not report whether having long-term illness at age 14. Thus, we integrated ages 14 and 15 responses.	**W1chea1HS** (wave_one_lsype_young_person_2020) **W2chea1HS** (wave_two_lsype_young_person_2020) **W4Hea2YP** (wave_four_lsype_young_person_2020) **W6HealthYP** (wave_six_lsype_young_person_2020)	**W1chea1HS; W2chea1HS; W4Hea2YP; W6HealthYP; W7HealthYP; W8LOIL; W9LOIL** 1= yes 2= no

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
			W7HealthYP (wave_seven_lsype_young_person_2020) **W8LOIL** (ns8_2015_main_interview) **W9LOIL** (ns9_2022_main_interview)	

Table 6. Anthropometric variables

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Weight	**wt0** **"Weight (kilograms, parent reported, age 0y)"** **wt25; wt32** **"Weight (kilograms, age yy)"** (continuous, kilograms)	Original values were retained.	**bwtkg** (lsype_history_file_wave_one_and_wave_two_june_2008) **W4bwtkgYP** (wave_four_lsype_history_2020) **W8WEIGHT** (ns8_2015_main_interview) **W9WEIGHT** (ns9_2022_main_interview)	**Bwtkg; W4bwtkgYP** Birth weight Kilos (edited), Source: Wave 1 and 2, History respondent (no min/max); Wave 4, Main parent (0-7.3kg) **W8WEIGHT** Weight in metric from answers of WTKILOS, WTSTONE and WTPOUNDS (no min/max) **W9WEIGHT** Weight in metric from answers of WTKILOS (20-300 kg), WTSTONE (4-50 stones) and WTPOUNDS (0-13 pounds)
Height	**ht25; ht32** **"Height (metres, age yy)"** **ht25_32** **"Height (metres, age 25y or 32y)"** (continuous, metres)	Height may be stable when reaching adulthood, so at age 32, those who reported their heights at age 25 did not need to response. Also, we included an integrated variable that combined both ages 25 and 32 height. We also kept the original measures at these two ages to ensure no loss of detail.	**W8HEIGHT** (ns8_2015_main_interview) **W9HEIGHT** (ns9_2022_main_interview)	**W8HEIGHT** Height in metric from answers of HTMETRES + HTCMS or HTFEET + HTINCHES (no min/max) **W9HEIGHT** Height in metric from answers of HTMETRES (1-2 metres) + HTCMS (0-99 centimetres) or HTFEET (3-8 feet) + HTINCHES (0-11 inches)

Body Mass Index (BMI)	**bmi25; bmi32** **" Body Mass Index (BMI) (age yy)"** (continuous, kg/m^2)	Original values were retained.	W8DBMI (ns8_2015_derived) **W9DBMI** (ns9_2022_derived_variables)	**W8DBMI** Body mass index (BMI) kg/m2 of cohort member derived from height (W8HEIGHT) and weight (W8WEIGHT) (no min/max) **W9DBMI** Body mass index (BMI) kg/m2 of cohort member derived from height (W9DHGHTM) and weight (W9DWGHTK). (no min/max)

Table 7. Behaviour variables

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Smoke	**smk14; smk15; smk16; smk25; smk32** **"Smoking ever and frequency (age yy)"** 0= Never 1= Used to, don't at all now 2= Smoke cigs occasionally – not every day 3= Smoke almost every day **smknw14; smknw15; smknw16; smknw25; smknw32** **"Smoking now binary (age yy)"** 0= no 1= yes	We included two sets of variables in the core data. The first set indicates the frequency of having cigarettes (no specified period), and we used the four response categories to allow cross sweep comparison. The second set indicates whether the cohort member smoked now.	**W1cignowYP; W1cigfreqYP** (wave_one_lsype_young_person_2020) **W2cignowYP; W2cigfreqYP** (wave_two_lsype_young_person_2020) **W3cignowYP; W3cigfreqYP** (wave_three_lsype_young_person_2020) **W8SMOKING** (ns8_2015_self_completion) **W9SMOKING** (ns9_2022_main_interview)	**W1cignowYP; W2cignowYP; W3cignowYP** 1= Yes 2= No **W1cigfreqYP; W2cigfreqYP; W3cigfreqYP** 1= I have never smoked 2= I have only ever tried smoking once 3= I used to smoke sometimes but I never smoke a cigarette now 4= I sometimes smoke cigarettes now but I don't smoke as many a 5= I usually smoke between one and six cigarettes a week 6= I usually smoke more than six cigarettes a week **W8SMOKING; W9SMOKING** 1= never 2= used to, don't at all now 3= smoke cigs occasionally – not every day 4= smoke cigs every day

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Alcohol	**alcfst** **"First time mentioned having alcohol (age 14-32y)"** 14= 14y 15= 15y 16= 16y 17= 17y 19= 19y 20= 20y 25= 25y 32= 32y 99= never **alcfreq14; alcfreq15; alcfreq16; alcfreq17; alcfreq19; alcfreq20** **"Alcohol frequency (age yy)"** 0= Less often/not at all 1= Once every couple of months 2= Once to three times a month 3= Once or twice a week 4= Most days **audita25; audita32;** **"AUDIT-C A (age yy)"** 0= Never 1= Monthly or less 2= 2-4 times a month 3= 2-3 times a week 4= 4 or more times a week	We included three sets of variables related to alcohol consumption. The first captures the sweep at which a cohort member first reported having an alcoholic drink; this may not reflect the actual age of first use. The second set measures the frequency of alcohol consumption in the past 12 months, and the response categories were integrated into five levels to ensure consistency across sweeps (Sweeps 1, 2, 3, 4, 6, and 7). In Sweeps 8 and 9, alcohol consumption was assessed using the standardised Alcohol Use Disorders Identification Test – Consumption (AUDIT-C). For these sweeps, the original values were retained.	**W1alceverYP; W1alcmonYP; W1alcfreqYP** (wave_one_lsype_young_person_2020) **W2alceverYP; W2alcfreqYP** (wave_two_lsype_young_person_2020) **W3alceverYP; W3alcfreqYP** (wave_three_lsype_young_person_2020) **W4AlcEverYP; W4AlcFreqYP** (wave_four_lsype_young_person_2020) **W6AlcEverYP; W6AlcFreqYP** (wave_six_lsype_young_person_2020) **W7AlcEverYP; W7AlcFreqYP** (wave_seven_lsype_young_person_2020) **W8AUDIT1; W8AUDIT2; W8AUDIT6** (ns8_2015_self_completion) **W9AUDIT1; W9AUDIT2; W9AUDIT3** (ns9_2022_main_interview)	See Table 8.

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	auditb25; auditb32; **"AUDIT-C B (age yy)"** 0= 0 1= 1-2 drinks 2= 3-4 drinks 3= 5-6 drinks 4= 7-9 drinks 5= 10+ **auditc25; auditc32;** **"AUDIT-C C (age yy)"** 0= Never 1= Less than monthly 2= Monthly 3= Weekly 4= Daily, or almost daily			
Drug	**drgcnbevr** **"Cannabis ever used (age 14-32y)"** 0= No 1= Yes **drgcnbfst** **"First time reported ever used cannabis (age 14-32y)"** 14= 14y 15= 15y 16= 16y 17= 17y	We included two types of drug use in the core dataset: cannabis and other drugs. Although more specific substances were recorded after Sweep 7, we grouped all non-cannabis substances into a single "other drugs" category to enable cross-sweep comparisons and due to the small number of cases for individual drugs. For each type, three sets of variables were derived. The first indicates whether the cohort member had ever used the drug across any	**W1canntryYP** (wave_one_lsype_young_person_2020) **W2canntryYP** (wave_two_lsype_young_person_2020) **W3canntryYP** (wave_three_lsype_young_person_2020) **W4CannTryYP** (wave_four_lsype_young_person_2020) **W6DrugYP0a; W6DrugYP0b; W6DrugOftenYP0a; W6DrugOftenYP0b** (wave_six_lsype_young_person_2020)	See Table 8.

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	19= 19y 20= 20y 25= 25y 32= 32y 99= never **drgcnbnw19; drgcnbnw20; drgcnbnw25; drgcnbnw32** **"Cannabis current use (age yy)"** 0= No 1= Yes **drgothevr** **"Other drugs ever used (age 19-32y)"** 0= No 1= Yes **drgothfst** **"First time reported ever used other drugs (age 19-32y)"** 19= 19y 20= 20y 25= 25y 32= 32y 99= never **drgothnw19; drgothnw20; drgothnw25; drgothnw32** **"Other drugs current use (age yy)"** 0= No 1= Yes	sweep. The second captures the sweep at which the drug was first reported, though this may not reflect the actual age of first use. It is important to note that the age coverage differs between cannabis and other drugs due to variations in the original data collection. The final set of variables, available from age 19 onwards, captures whether the drug was used in the past four weeks.	**W7DrugYP1YP0a; W7DrugYP1YP0b; W7DrugYP2YP0a; W7DrugYP2YP0b; W7DrugOftenYP0a; W7DrugOftenYP0b; W7DrugOftenYP0c; W7DrugOftenYP0d; W7DrugOftenYP0e; W7DrugOftenYP0f; W7DrugOftenYP0g; W7DrugOftenYP0h; W7DrugOftenYP0i** (wave_seven_lsype_young_person_2020) **W8DRUGYP10A:I; W8DRUGYP20A:I; W8DRUGOFTEN0A-I** (ns8_2015_self_completion) **W9DRUGYP1J; W9DRUGYP10A:I; W9DRUGYP20A:J; W9DRUGOFTEN0A:J** (ns9_2022_main_interview)	

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Exercise	**spt14; spt15; spt17; spt19; spt20; spt25; spt32** **"Exercise frequency (age yy)"** 0= Most days 1= More than once a week 2= Once a week 3= Less than once a week/hardly ever/never	To allow cross-sweep comparison, the response categories were integrated into four levels.	**W1sportYP** (wave_one_lsype_young_person_2020) **W2sportYP** (wave_two_lsype_young_person_2020) **W4SportYP** (wave_four_lsype_young_person_2020) **W6SportYP** (wave_six_lsype_young_person_2020) **W7SportYP** (wave_seven_lsype_young_person_2020) **W8EXERCISE** (ns8_2015_main_interview) **W9EXERCISEH** (ns9_2022_main_interview)	**W1sportYP; W2sportYP; W4SportYP; W6SportYP; W7SportYP;** 1= most days 2= more than once a week 3= once a week 4= less than once a week 5= hardly ever 6= never **W8EXERCISE; W9EXERCISEH** (Number of days per week do exercise for 30 mins or more)
Absences	**abs1m14; abs1m15; abs1m16** **"Absence from school over 1 month in the past 12 months (age yy)"** 0= No 1= Yes	The original values were reversely recoded.	**W1abs1myMP** (wave_one_lsype_young_person_2020) **W2abs1myMP** (wave_two_lsype_young_person_2020) **W3abs1myMP** (wave_three_lsype_young_person_2020)	**W1abs1myMP; W2abs1myMP; W3abs1myMP** 1= yes 2= no

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Suspended	**susp14; susp15; susp16; susp17** **"Suspended from school (age yy)"** 0= No 1= Yes **expl14; expl15; expl16; expl17** **"Expelled from school (age yy)"** 0= No 1= Yes	The original values were reversely recoded. At age 17, suspended and expelled questions were only administered to the boost sample.	**W1suspendMP; W1expelMP** (wave_one_lsype_young_person_2020) **W2SuspendMP; W2ExpelMP** (wave_two_lsype_young_person_2020) **W3suspendMP; W3expelMP** (wave_three_lsype_young_person_2020) **W4Expel1YP; W4Expel2YP** (wave_four_lsype_young_person_2020)	**W1suspendMP; W1expelMP; W2SuspendMP; W2ExpelMP; W3suspendMP; W3expelMP; W4Expel1YP; W4Expel2YP** 1= yes 2= no
Truancy	**trua14** **"Truancy within past three years (age 14)"** **trua15; trua16; trua17** **"Truancy in the past year (age yy)"** 0= Never played truant 1= For weeks at a time 2= Several days at a time 3= Particular days or lessons 4= Odd day or lesson	To allow cross-sweep comparison, the response categories were integrated into five levels. Notably, in Sweep 4 (age 17), only the boost sample was asked about their truancy behaviour.	**W1truantYP; W1truant1YP** (wave_one_lsype_young_person_2020) **W2truantYP; W2truant1YP** (wave_two_lsype_young_person_2020) **W3truantYP; W3truant1YP** (wave_three_lsype_young_person_2020) **W4TruantYP** (wave_four_lsype_young_person_2020)	**W1truantYP; W2truantYP; W3truantYP** 1= yes 2= no **W1truant1YP; W2truant1YP; W3truant1YP** 1= for weeks at a time 2= several days at a time 3= particular days or lessons 4= odd day or lesson **W4TruantYP (only boost)** 1= for weeks at a time 2= several days at a time 3= particular days or lessons 4= odd day or lesson 5= never played truant

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Police contact	**pol14; pol15; pol16; pol17** **"Police contact binary (age yy)"** 0= never 1= yes/not in last 3 years **polcnt14; polcnt15; polcnt16; polcnt17** **"Police contact count (age yy)"** (continuous, counts) **polwrn25; polwrn32;** **"Police warning (age yy)"** **polars25; polars32;** **"Police arrest (age yy)"** **polcau25; polcau32;** **"Police caution (age yy)"** **polglt25; polglt32;** **"Police found guilty (age yy)"** **polpnd25; polpnd32** **"Given a Penalty Notice for Disorder (age yy)"** 0= No 1= Yes	The measures of police contact differed between Sweeps 1–4 and Sweeps 8–9, reflecting both changes in question wording and the cohort's life stage. We combined the measures from Sweeps 1–4 into a consistent format to enable comparisons during adolescence, while retaining the original measures from Sweeps 8 and 9 to reflect experiences in adulthood.	**W1Police1MP; W1police2MP** (wave_one_lsype_young_person_2020) **W2police1MP; W2Police2MP** (wave_two_lsype_young_person_2020) **W3police1MP; W3police2MP** (wave_three_lsype_young_person_2020) **W4Police1MP; W4Police2MP** (wave_four_lsype_young_person_2020) **W8CJSCONTACT0A:E** (ns8_2015_self_completion) **W9CJSCONTACT0A:E** (ns9_2022_main_interview)	**W1Police1MP; W4Police1MP** 1= yes, in last three years 2= no 3= not in the last three years **W2police1MP; W3police1MP** 1= yes 2= no **W1police2MP; W2Police2MP; W3police2MP; W4Police2MP** (times of policy contact) **W8CJSCONTACT0A:E; W9CJSCONTACT0A:E** 1= yes 2= no

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Bullying	**bul14; bul15; bul16; bul17; bul20; bul25** **"Bullying victimisation (age yy)"** 0= No 1= Yes	The original surveys captured several types of bullying victimisation, with variations in question wording and categories across sweeps. To enable cross-sweep comparison, align with common measures in the bullying literature, and address small sample sizes for some specific types, we created binary variables indicating whether the cohort member experienced any bullying victimisation at each available sweep.	**W1bulrc** (wave_one_lsype_young_person_2020) **W2bulrc** (wave_two_lsype_young_person_2020) **W3bulrc** (wave_three_lsype_young_person_2020) **W4V1perSYP; W4ViolentYP; W4HurtYP; W4ThreatsYP; W4MadegiveYP; W4NamesYP** (wave_four_lsype_young_person_2020) **W7BullyTypeYP0a:f** (wave_seven_lsype_young_person_2020) **W8BULLYTYPE0A:G** (ns8_2015_self_completion)	See Table 8.

Table 8. Selected variables with detailed source variables' response category

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Current qualification studied, own	**educaim17; educaim19; educaim20; educaim25; educaim32** **"Current qualification studied level 3-category (age yy)"** 0= NVQ 4-5 1= NVQ 1-3 2= None/entry 3= Other 4= None of these qualifications 5= Not studying **Key differences between 2, 3, and 4 categories: ▪ None/entry (2): No qualifications or only entry-level. ▪ Other (3): Matches the "Other" category in the original survey; respondent has qualifications outside the predefined list. ▪ None of these qualifications (4): Matches the "None of these qualifications" category in the original survey; respondent may have qualifications but explicitly indicated none fit the listed options.	To allow cross-sweep comparison, the new response categories followed the categorisation of the CLS COVID-19 cross-cohort survey (Centre for Longitudinal Studies, 2024). For ages 25 and 32, activity variables were also used to identify respondents who were not studying.	**w4saim** (wave_four_lsype_young_person_2020) **W6Saim** (wave_six_lsype_young_person_2020) **W7SAim** (wave_seven_lsype_young_person_2020) **W8ACTIVITY05; W8ACQUC0A-0Q; W8VCQUC0A-0R** (ns8_2015_main_interview) **W9ECONACT2; W9ACQUC0A-0V; W9VCQUC0A-AI** (ns9_2022_main_interview)	**w4saim** 1= NVQ3 2= AVCE 3= A/AS 4= Other level 3 5= NVQ2 6= Intermediate GNVQ 7= Other level 2 8= GCSE 9= NVQ1 10= foundation 11= other level 1 12= other 13= no detail 14= not studying **W6Saim** 1= NVQ5 2= First/other degree 3= NVQ4 4= other HE 5= NVQ3 6= AVCE 7= A/AS 8= Other level 3 9= NVQ2 10= Other level 2

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				11= GCSE 12= NVQ1 13= other level 1 14= other (level unknown) 15= no detail 16= not studying **W7SAim** 1= NVQ1 2= other level 1 3= NVQ2 4= GCSE 5= Other level 2 6= NVQ3 7= A/AS 8= AVCE 9= Other level 3 10= NVQ4 11= First/other degree 12= Other HE 13= NVQ5 14= other (level unknown) **W8ACQUC0A-0Q;** 1. University Higher Degree (e.g. MSc, PhD) 2. First degree level qualification including foundation degrees, graduate membership of a professional Institute, PGCE 3. Diploma in higher education 4. Teaching qualification (excluding PGCE)

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				5. Nursing or other medical qualification not yet mentioned 6. A Level 7. Welsh Baccalaureate 8. International Baccalaureate 9. AS Level 10. Higher Grade/Advanced Higher (Scotland) 11. Certificate of sixth year studies 12. GCSE 13. Standard Grade / Lower (Scotland) 14. Other school 15. None of the above **W8VCQUC0A-0R** 1. Youth training certificate 2. Key Skills 3. Basic skills 4. Entry level qualifications (Wales) 5. Modern apprenticeship/trade apprenticeship 6. RSA/OCR/Clerical and commercial qualifications (e.g. typing/shorthand/book-keeping/commerce) 7. City and Guilds Certificate 8. GNVQ/GSVQ 9. NVQ/SVQ - Level 1 - 2 10. NVQ/SVQ - Level 3 - 5 11. HNC/HND 12. ONC/OND

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				13. BTEC/BEC/TEC/EdExcel/LQL 14. SCOTVEC, SCOTEC or SCOTBEC 15. Other vocational, technical or professional qualification **W9ACQUC0A-0V** 1. Doctorate or equivalent 2. Masters or equivalent 3. Undergraduate or equivalent 4. Post-graduate Diplomas and Certificates (e.g. PGDip, PGCert, PGCE) 5. Diplomas in higher education and other higher education qualifications (e.g. DipHE, Foundation Degree) 6. Teaching qualifications for schools or further education (below degree level) 7. A/AS Levels or equivalent 8. GCSE a. Grade A-C, Level 4-9 b. Grade D-G, Level 1-3 9. Scottish Qualifications a. SCE Higher b. Scottish Certificate Sixth Year Studies c. SCE Standard d. National 4 and 5 e. National 2 and 3 10. Irish Qualifications a. Leaving Certificate b. Junior Certificate grade A-C c. Junior Certificate grade D and below

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				11. Other academic qualifications (including overseas) 12. None of these qualifications **W9VCQUC0A-AI** 1. Professional qualifications at degree level e.g. graduate member of professional institute, chartered accountant or surveyor 2. Nursing or other medical qualifications (below degree level) 3. NVQ or SVQ a. Level 4 or 5 b. Level 3 c. Level 2 d. Level 1 4. GNVQ a. GNVQ Advanced b. GNVQ Intermediate c. Level 3 d. Level 2 e. Level Foundation 5. City & Guilds a. advanced craft, Part III b. Craft, Part II c. Craft, Part I d. Level 3 e. Level 2 f. Level 1 6. RSA

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				a. Advanced Diploma b. Higher Diploma c. RSA Diploma d. RSA Stage I, II,III 7. BTEC a. Higher Level BTEC b. BTEC National c. BTEC First 8. SCOTVEC a. SCOTVEC National Certificate b. SCOTVEC first or general diploma c. SCOTVEC general diploma d. SCOTVEC modules 9. HND or HNC 10. OND or ONCM 11. Junior certificate 12. Other vocational qualifications (including some overseas) 13. None of these qualifications
Education, own	**educ25; educ32** **"Highest education level 3-category (yy)"** 0= NVQ 4-5 1= NVQ 1-3 2= None/entry 3= other 4= None of these qualifications	To allow cross-sweep comparison, the new response categories followed the categorisation of the CLS COVID-19 cross-cohort survey (Centre for Longitudinal Studies, 2024).	**W8VCQU0A-0R** (ns8_2015_main_interview) **W8DHANVQH** (ns8_2015_derived) **W9DANVQH; W9DVNVQH** (ns9_2022_derived_variables)	**W8VCQU0A-0R** 1. Youth training certificate 2. Key Skills 3. Basic skills 4. Entry level qualifications (Wales) 5. Modern apprenticeship/trade apprenticeship

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	Key differences between 2, 3, and 4 categories: ▪ None/entry (2): No qualifications or only entry-level. ▪ Other (3): Matches the "Other" category in the original survey; respondent has qualifications outside the predefined list. ▪ None of these qualifications (4): Matches the "None of these qualifications" category in the original survey; respondent may have qualifications but explicitly indicated none fit the listed options. **educadtl32 **"Highest education detail - academic (age 32)"** 1= Doctorate or equivalent 2= Masters or equivalent 3= Undergraduate or equivalent 4= Post-graduate Diplomas and Certificates 5= Diplomas in higher education and other higher education qualifications 6= Teaching qualifications for schools or further education 7= A/AS Levels or equivalent 8= GCSE - Grade A-C, Level 4-9 9= GCSE - Grade D-G, Level 1-3 10= Scottish Qualifications - SCE Higher 11= Scottish Qualifications - Scottish Certificate Sixth Year Studies		**W9ACQU0A-0V; W9VCQU0A-AI** (ns9_2022_main_interview)	6. RSA/OCR/Clerical and commercial qualifications (e.g. typing/shorthand/book-keeping/commerce) 7. City and Guilds Certificate 8. GNVQ/GSVQ 9. NVQ/SVQ - Level 1 - 2 10. NVQ/SVQ - Level 3 - 5 11. HNC/HND 12. ONC/OND 13. BTEC/BEC/TEC/EdExcel/LQL 14. SCOTVEC, SCOTEC or SCOTBEC 15. Other vocational, technical or professional qualification **W8DHANVQH** 1. NVQ Level1 2. NVQ Level2 3. NVQ Level3 4. NVQ Level4 5. NVQ Level5 95. Other academic qualification 96. None of these qualifications **W9DANVQH** 1. NVQ Level1 2. NVQ Level2 3. NVQ Level3 4. NVQ Level4 5. NVQ Level5 95. Other academic qualification 96. None of these qualifications

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	12= Scottish Qualifications - SCE Standard 13= Scottish Qualifications - National 4 and 5 14= Scottish Qualifications - National 2 and 3 15= Irish Qualifications - Leaving Certificate 16= Irish Qualifications - Junior Certificate grade A-C 17= Irish Qualifications - Junior Certificate grade D and below 18= Other academic qualifications (including overseas) 19= None of these qualifications **educvdtl32** **"Highest education detail - vocational (age 32)"** 1= Professional qualifications at degree level 2= Nursing or other medical qualifications 3= NVQ or SVQ - Level 4 or 5 4= NVQ or SVQ - Level 3 5= NVQ or SVQ - Level 2 6= NVQ or SVQ - Level 1 7= GNVQ - GNVQ Advanced 8= GNVQ - GNVQ Intermediate 9= GNVQ - Level 3 10= GNVQ - Level 2 11= GNVQ - Level Foundation 12= City & Guilds - advanced craft, Part III 13= City & Guilds - Craft, Part II 14= City & Guilds - Craft, Part I 15= City & Guilds - Level 3			**W9DVNVQH** 1. NVQ Level1 2. NVQ Level2 3. NVQ Level3 4. NVQ Level4 5. NVQ Level5 95. Other academic qualification 96. None of these qualifications **W9ACQU0A-0V** 1. Doctorate or equivalent 2. Masters or equivalent 3. Undergraduate or equivalent 4. Post-graduate Diplomas and Certificates (e.g. PGDip, PGCert, PGCE) 5. Diplomas in higher education and other higher education qualifications (e.g. DipHE, Foundation Degree) 6. Teaching qualifications for schools or further education (below degree level) 7. A/AS Levels or equivalent 8. GCSE a. Grade A-C, Level 4-9 b. Grade D-G, Level 1-3 9. Scottish Qualifications a. SCE Higher b. Scottish Certificate Sixth Year Studies c. SCE Standard d. National 4 and 5

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	16= City & Guilds - Level 2 17= City & Guilds - Level 1 18= RSA - Advanced Diploma 19= RSA - Higher Diploma 20= RSA - RSA Diploma 21= RSA - RSA Stage I, II,III 22= BTEC - Higher Level BTEC 23= BTEC - BTEC National 24= BTEC - BTEC First 25= SCOTVEC - SCOTVEC National Certificate 26= SCOTVEC - SCOTVEC first or general diploma 27= SCOTVEC - SCOTVEC general diploma 28= SCOTVEC - SCOTVEC modules 29= HND or HNC 30= OND or ONCM 31= Junior certificate 32= Other vocational qualifications (including some overseas) 33= None of these qualifications			e. National 2 and 3 10. Irish Qualifications a. Leaving Certificate b. Junior Certificate grade A-C c. Junior Certificate grade D and below 11. Other academic qualifications (including overseas) 12. None of these qualifications **W9VCQU0A-AI** 1. Professional qualifications at degree level e.g. graduate member of professional institute, chartered accountant or surveyor 2. Nursing or other medical qualifications (below degree level) 3. NVQ or SVQ a. Level 4 or 5 b. Level 3 c. Level 2 d. Level 1 4. GNVQ a. GNVQ Advanced b. GNVQ Intermediate c. Level 3 d. Level 2 e. Level Foundation 5. City & Guilds a. advanced craft, Part III b. Craft, Part II c. Craft, Part I

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				d. Level 3 e. Level 2 f. Level 1 6. RSA a. Advanced Diploma b. Higher Diploma c. RSA Diploma d. RSA Stage I, II,III 7. BTEC a. Higher Level BTEC b. BTEC National c. BTEC First 8. SCOTVEC a. SCOTVEC National Certificate b. SCOTVEC first or general diploma c. SCOTVEC general diploma d. SCOTVEC modules 9. HND or HNC 10. OND or ONCM 11. Junior certificate 12. Other vocational qualifications (including some overseas) 13. None of these qualifications

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Education, parents	**educpa** **"Highest education level father (CM age 14-17y)"** **educma** **"Highest education level mother (CM age 14-17y)"** 0= NVQ 4-5 1= NVQ 1-3 2= None/entry 3= other 4= None of these qualifications **Key differences between 2, 3, and 4 categories: ▪ None/entry (2): No qualifications or only entry-level. ▪ Other (3): Matches the "Other" category in the original survey; respondent has qualifications outside the predefined list. ▪ None of these qualifications (4): Matches the "None of these qualifications" category in the original survey; respondent may have qualifications but explicitly indicated none fit the listed options.	Parental education was selected from the earliest valid parents' reported response across the available sweeps (1, 2, and 4) for the original sample and Sweep 4 for the boost sample. To allow cross-sweep and parent-children comparison, the new response categories followed the categorisation of the CLS COVID-19 cross-cohort survey (Centre for Longitudinal Studies, 2024).	**W1hiqualmum; W1hiqualdad** (wave_one_lsype_family_background_2020) **W2hiqualmum; W2hiqualdad** (wave_two_lsype_family_background_2020) **w4hiqualmum; w4hiqualdad** (wave_four_lsype_family_background_2020)	**W1hiqualmum; W1hiqualdad; W2hiqualmum; W2hiqualdad; w4hiqualmum; w4hiqualdad** 1= Higher Degree 2= First Degree 3= HE Diploma 4= HNC/HND/NVQ4 5= Teaching qualification, non-degree 6= Nursing qualification, non-degree 7= A Levels 8= OND/ONC 9= City and guilds part III, NVQ3 10= CSYS 11= Scottish Higher Grade 12= AS Level 13= Trade apprenticeship 14= City and guilds part II, NVQ2 15= GCSE grade A-C and equivalent 16= GCSE grade D-E and equivalent 17= City and guilds part I, NVQ1 18= Youth training, skill seekers 19= Qualification, level unspecified 20= No qualification mentioned

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	educdtlpa **"Highest education detail father (CM age 14-17y)"** **educdtlma** **"Highest education detail mother (CM age 14-17y)"** 1= Higher Degree 2= First Degree 3= HE Diploma 4= HNC/HND/NVQ4 5= Teaching qualification, non-degree 6= Nursing qualification, non-degree 7= A Levels 8= OND/ONC 9= City and guilds part III, NVQ3 10= CSYS 11= Scottish Higher Grade 12= AS Level 13= Trade apprenticeship 14= City and guilds part II, NVQ2 15= GCSE grade A-C and equivalent 16= GCSE grade D-E and equivalent 17= City and guilds part I, NVQ1 18= Youth training, skill seekers 19= Qualification, level unspecified 20= No qualification mentioned			

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Economic activity, own	**ecoact17; ecoact18; ecoact19; ecoact20; ecoact25; ecoact32** **"Economic activity simple (age yy)"** 1= In paid work 2= Apprenticeship/government training scheme/training 3= Education 4= Unemployed 5= Looking after home 6= Other **ecoactadu25; ecoactadu32** **"Economic activity detailed (age yy)"** 1= employee – in paid work 2= self employed 3= voluntary work 4= unemployed 5= education 6= apprenticeship 7= Government employment scheme 8= sick/disabled 9= looking after home 10= something else	Two types of economic activity variables were created. A simple version for all sweeps allows comparison across sweeps. A more detailed version for Sweeps 8 and 9 where additional information is available, ensuring no loss of detail.	**W4empsYP** (wave_four_lsype_young_person_2020) **W5mainactYP** (wave_five_lsype_young_person_2020) **W6TCurrentAct** (wave_six_lsype_young_person_2020) **W7TCurrentAct** (wave_seven_lsype_young_person_2020) **W8DACTIVITYC** (ns8_2015_derived) **W9DACTIVITYC** (ns9_2022_derived_variables)	**W4empsYP** 1= Doing paid work for 30 or more hours 2= Doing paid work for fewer than 30 hours a week 3= Unemployed/looking for a job 4= On a training course or scheme 5= In full-time education/at school 6= looking after the family/household 7= Retired from work altogether 8= Sick/disabled 9= other **W5mainactYP** 1= Apprenticeship 2= PT work, PT education 3= in paid work 4= in education 5= on a training course/scheme 6= on the entry to employment scheme 7= unemployed/looking for work 8= looking after the family 9= waiting for a course or job to start 10= waiting for an exam result 11= waiting for the result of a job application **W6TCurrentAct** 1= doing a course at a university 2= in education 3= in paid work

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				4= on a training course or scheme 5= doing an apprenticeship 6= waiting for a course or job to start 7= looking after the family 8= unemployed 9= waiting for exam results or results of job application 10= PT work PT education 11= voluntary work **W7TCurrentAct** 1= University 2= School/college education 3= paid work 4= training course/scheme 5= apprenticeship 6= waiting for a course or job to start 7= looking after home/family 8= unemployed and looking for work 9= PT work PT education 10= voluntary work 11= Government employment scheme 12= travelling 13= break from work/college 14= ill/disabled 15= not defined

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				W8DACTIVITYC; W9DACTIVITYC 1= employee – in paid work 2= self employed 3= voluntary work 4= unemployed 5= education 6= apprenticeship 7= Government employment scheme 8= sick/disabled 9= looking after home 10= something else
Social class, Own	**nssec17; nssec18; nssec19; nssec20; nssec25; nssec32** **"NS-SEC 17-category (age yy)"** 1= Employers in large organisations 2= Higher managerial and administrative occupations 3= Higher professional occupations 4= Lower professional and higher technical occupations 5= Lower managerial and administrative occupations 6= Higher supervisory occupations 7= Intermediate occupations 8= Employers in small establishments 9= Own account workers	To allow comparison across sweeps, 17 categories were used. At ages 25 and 32, activity variables were also used to identify full-time students.	**W4nsseccatYP** (wave_four_lsype_young_person_2020) **W5nsseccatYP** (wave_five_lsype_young_person_2020) **w6nsseccatYP** (wave_six_lsype_young_person_2020) **W7NSSECCat** (wave_seven_lsype_young_person_2020) **W8DNSSEC17; W8DACTIVITYC** (ns8_2015_derived) **W9NSSEC** (ns9_2022_main_interview)	**W4nsseccatYP; W5nsseccatYP; w6nsseccatYP; W7NSSECCat** 1.0= Employers in large organisations 2.0= Higher managerial occupations 3.1= Higher professional traditional employee 3.2= Higher professional new employee 3.3= Higher professional traditional self emp 3.4= Higher professional new self emp 4.1= Lower professional traditional employee 4.2= Lower professional new employee 4.3= Lower professional traditional self emp 4.4= Lower professional new self emp 5.0= Lower managerial occupations 6.0= Higher supervisory occupations

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	10= Lower supervisory occupations 11= Lower technical occupations 12= Semi-routine occupations 13= Routine occupations 14= Never worked and Long-term unemployed 15= Full-time students 16= Not classified or inadequately stated 17= Not classifiable for other reasons		**W9DACTIVITYC** (ns9_2022_derived_variables)	7.1= Intermediate clerical and administrative 7.2= Intermediate sales and service 7.3= Intermediate technical and auxiliary 7.4= Intermediate engineering 8.1= Employers in small orgs non-professional 8.2= Employers in small orgs agriculture 9.1= Own account workers non professional 9.2= Own account workers agriculture 10.0= Lower supervisory occupations 11.1= Lower technical craft 11.2= Lower technical process operative 12.1= Semi routine sales 12.2= Semi routine services 12.3= Semi routine technical 12.4= Semi routine operative 12.5= Semi routine agricultural 12.6= Semi routine clerical 12.7= Semi routine childcare 13.1= Routine sales and service 13.2= Routine production 13.3= Routine technical 13.4= Routine operative 13.5= Routine agricultural 14.1= Never worked 14.2= Long-term unemployed 15.0= Full-time students 16.0= Not classified or inadequately stated

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				17.0= Not classifiable for other reasons **W8DNSSEC17** 1.0= Employers in large organisations 2.0= Higher managerial occupations 3.1= Higher professional traditional employee 3.2= Higher professional new employee 3.3= Higher professional traditional self emp 3.4= Higher professional new self emp 4.1= Lower professional traditional employee 4.2= Lower professional new employee 4.3= Lower professional traditional self emp 4.4= Lower professional new self emp 5.0= Lower managerial occupations 6.0= Higher supervisory occupations 7.1= Intermediate clerical and administrative 7.2= Intermediate sales and service 7.3= Intermediate technical and auxiliary 7.4= Intermediate engineering 8.1= Employers in small orgs non-professional 8.2= Employers in small orgs agriculture 9.1= Own account workers non professional 9.2= Own account workers agriculture 10.0= Lower supervisory occupations 11.1= Lower technical craft 11.2= Lower technical process operative

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				12.1= Semi routine sales 12.2= Semi routine services 12.3= Semi routine technical 12.4= Semi routine operative 12.5= Semi routine agricultural 12.6= Semi routine clerical 12.7= Semi routine childcare 13.1= Routine sales and service 13.2= Routine production 13.3= Routine technical 13.4= Routine operative 13.5= Routine agricultural **W9NSSEC** 1= Employers in large organisations 2= Higher managerial and administrative occupations 3= Higher professional occupations 4= Lower professional and higher technical occupations 5= Lower managerial and administrative occupations 6= Higher supervisory occupations 7= Intermediate occupations 8= Employers in small establishments 9= Own account workers 10= Lower supervisory occupations 11= Lower technical occupations 12= Semi-routine occupations 13= Routine occupations

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
				14= Never worked and Long-term unemployed 15= Full-time students 16= Not classified or inadequately stated 17= Not classifiable for other reasons
Social class of parents	**nssecpa14; nssecpa15; nssecpa16; nssecpa17; nssecpa18** **"NS-SEC father 17-category (CM age yy)"** **nssecma14; nssecma15; nssecma16; nssecma17; nssecma18** **"NS-SEC mother 17-category (CM age yy)"** 1= Employers in large organisations 2= Higher managerial and administrative occupations 3= Higher professional occupations 4= Lower professional and higher technical occupations 5= Lower managerial and administrative occupations 6= Higher supervisory occupations 7= Intermediate occupations 8= Employers in small establishments 9= Own account workers 10= Lower supervisory occupations 11= Lower technical occupations 12= Semi-routine occupations 13= Routine occupations 14= Never worked and Long-term unemployed	To allow comparison across sweeps and between parents and cohort members, 17 categories were used.	**W1nsseccatmum; W1nsseccatdad** (wave_one_lsype_family_background_2020) **W2nsseccatmum; W2nsseccatdad** (wave_two_lsype_family_background_2020) **W3cnsseccatmum; W3cnsseccatdad** (wave_three_lsype_family_background_2020) **w4cnsseccatmum; w4cnsseccatdad** (wave_four_lsype_family_background_2020) **w5Cnsseccatmum; w5Cnsseccatdad** (wave_five_lsype_family_background_2020)	**W1nsseccatmum; W1nsseccatdad; W2nsseccatmum; W2nsseccatdad; W3cnsseccatmum; W3cnsseccatdad; w4cnsseccatmum; w4cnsseccatdad; w5Cnsseccatmum; w5Cnsseccatdad** 1.0= Employers in large organisations 2.0= Higher managerial occupations 3.1= Higher professional traditional employee 3.2= Higher professional new employee 3.3= Higher professional traditional self emp 3.4= Higher professional new self emp 4.1= Lower professional traditional employee 4.2= Lower professional new employee 4.3= Lower professional traditional self emp 4.4= Lower professional new self emp 5.0= Lower managerial occupations 6.0= Higher supervisory occupations 7.1= Intermediate clerical and administrative 7.2= Intermediate sales and service 7.3= Intermediate technical and auxiliary 7.4= Intermediate engineering

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	15= Full-time students 16= Not classified or inadequately stated 17= Not classifiable for other reasons			8.1= Employers in small orgs non-professional 8.2= Employers in small orgs agriculture 9.1= Own account workers non professional 9.2= Own account workers agriculture 10.0= Lower supervisory occupations 11.1= Lower technical craft 11.2= Lower technical process operative 12.1= Semi routine sales 12.2= Semi routine services 12.3= Semi routine technical 12.4= Semi routine operative 12.5= Semi routine agricultural 12.6= Semi routine clerical 12.7= Semi routine childcare 13.1= Routine sales and service 13.2= Routine production 13.3= Routine technical 13.4= Routine operative 13.5= Routine agricultural 14.1= Never worked 14.2= Long-term unemployed 15.0= Full-time students 16.0= Not classified or inadequately stated 17.0= Not classifiable for other reasons

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Home ownership	**hown14; hown15; hown16; hown17; hown18; hown19; hown20; hown25; hown32** **"Household ownership (age yy)"** 1= own outright 2= own, buying with help of mortgage/loan 3= part rent, part mortgage 4= rent it 5= live rent-free 6= other **hownteen14; hownteen15; hownteen16; hownteen17; hownteen18; hownteen19; hownteen20** **"Household ownership detailed (age yy)"** 1= Owned outright 2= Being bought on a mortgage/bank loan 3= Shared ownership (owns & rents property) 4= Rented from a Council or New Town 5= Rented from a Housing Association, 6= Rented privately, 7= Rent free, 8= Some other arrangement.	Two types of variables were created. A simple version for all sweeps allows comparison across sweeps. A more detailed version for Sweeps 1-7 where additional information is available, ensuring no loss of detail.	**W1hous12HH** (wave_one_lsype_family_background) **W2Hous12HH** (wave_two_lsype_family_background_2020) **W3hous12HH** (wave_three_lsype_family_background_2020) **W4Hous12HH** (wave_four_lsype_family_background_2020) **W5Hous12HH; W5Hous12BHH; W5Hous12CHH** (wave_five_lsype_family_background_2020) **W6Hous12YP; W6Hous12bYP; W6Hous12cYP** (wave_six_lsype_young_person_2020) **W7Hous12YP; W7Hous12bYP; W7Hous12cYP** (wave_seven_lsype_young_person_2020) **W8TENURE** (ns8_2015_main_interview)	**W1hous12HH; W2Hous12HH; W3hous12HH; W4Hous12HH** 1= Owned outright, 2= Being bought on a mortgage/bank loan, 3= Shared ownership (owns & rents property), 4= Rented from a Council or New Town, 5= Rented from a Housing Association, 6= Rented privately, 7= Rent free, 8= Some other arrangement. **W5Hous12HH; W6Hous12YP; W7Hous12YP** 1= owned 2= rented 3= something else 6= not to be asked HH grid **W5Hous12BHH; W6Hous12bYP; W7Hous12bYP** 1= owned outright 2= being bought on a mortgage/bank loan 3= shared ownership (owns and rents the property) 4= some other arrangement **W5Hous12CHH; W6Hous12cYP; W7Hous12cYP** 1= Rented from a council or new town 2= rented from a housing association

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
			W9DTENURE (ns9_2022_derived_variables)	3= rented privately 4= rent free 5= some other arrangement **W8TENURE; W9DTENURE** 1= own outright 2= own, buying with help of mortgage/loan 3= part rent, part mortgage 4= rent it 5= live rent-free 6= squatting 7= other
General Health Questionnaire (GHQ) (Goldberg & Williams, 1988)	**ghq15; ghq17; ghq25; ghq32** **"General Health Questionnaire (GHQ12) symptom count (age yy)"** (continuous, higher scores indicate poorer mental health) **ghqtl15; ghqtl17; ghqtl25; ghqtl32** **"General Health Questionnaire (GHQ12) score (age yy)"** (continuous, higher scores indicate poorer mental health)	Two types of variables were created. First, each item was recoded into 0 (no to seldom having the symptom) or 1 (usually to always having the symptom) and then was summed up to reflect the number of psychological distress symptoms. The second set is the GHQ Likert score (12–48), capturing the severity of symptoms. Both scoring methods are commonly used in the literature and provide complementary insights into mental health status.	**W2concenYP; W2nosleepYP; W2usefulYP; W2decideYP; W2strainYP; W2difficYP; W2activYP; W2probsYP; W2depressYP; W2noconfYP; W2wthlessYP; W2happyYP; W2ghq12scr** (wave_two_lsype_young_person_2020) **W4ConcenYP; W4NoSleepYP; W4UsefulYP; W4DecideYP; W4StrainYP; W4DifficYP; W4ActivYP; W4ProbsYP; W4DepressYP; W4NoConfYP; W4WthlessYP; W4HappyYP; W4ghq12scr**	GHQ12 questions: Concentrate on what doing* Lost sleep over worry Playing a useful part in things* Capable of making decisions* Felt constantly under strain Felt couldn't overcome difficulties Enjoy day to day activities* Face up to problems* Felt unhappy or depressed Losing confidence is self Thinking of self as worthless Felt reasonably happy* *Responses of positive items were reversed. 1= not at all present 2= no more than usual

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
			(wave_four_lsype_young_person_2020) **W8GHQ12_1:12** (ns8_2015_self_completion) **W9GHQ12_1:12** (ns9_2022_main_interview) **W8DGHQSC** (ns8_2015_derived) **W9DGHQSC** (ns9_2022_derived_variables)	3= rather more than usual 4= much more than usual
Generalized Anxiety Disorder 2-item (GAD-2) (Kroenke et al., 2007)	**gad32** **"GAD 2-item score (age 32y)"** (continuous, higher scores indicate more anxious)	GAD-2 was only measured at age 32. Original values were retained.	**W9DGAD2** (ns9_2022_derived_variables)	GAD2 questions Feeling nervous, anxious or on edge Not being able to stop or control worrying 1= not at all 2= several days 3= more than half the days 4= nearly every day
Patient Health Questionnaire-2 (PHQ-2) (Kroenke et al., 2003)	**phq32** **"PHQ 2-item score (age 32y)"** (continuous, higher scores indicate more depressed)	PHQ-2 was only measured at age 32. Original values were retained.	**W9DPHQ2** (ns9_2022_derived_variables)	PHQ2 questions Little interest or pleasure in doing things Feeling down, depressed or hopeless 1= not at all 2= several days 3= more than half the days 4= nearly every day

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
UCLA loneliness (Russell, 1996)	**lon32** **"UCLA Loneliness 3-item (age 32y)"** (continuous, higher scores indicate lonelier)	UCLA Loneliness scale was only measured at age 32. Original values were retained.	**W9DLONELINESS** (ns9_2022_derived_variables)	UCLA Loneliness 3 item questions: How often feels lack companionship How often feels left out How often feels isolated from others 1= hardly ever 2= some of the time 3= often
Alcohol	**alcfst** **"First time mentioned having alcohol (age 14-32y)"** **alcfreq14; alcfreq15; alcfreq16; alcfreq17; alcfreq19; alcfreq20** **"Alcohol frequency (age yy)"** 0= less often/not at all 1= once every couple of months 2= once to three times a month 3= once or twice a week 4= most days **audita25; audita32;** **"AUDIT-C A (age yy)"** 0= never 1= monthly or less 2= 2-4 times a month 3= 2-3 times a week 4= 4 or more times a week	We included three sets of variables related to alcohol consumption. The first captures the sweep at which a cohort member first reported having an alcoholic drink; this may not reflect the actual age of first use. The second set measures the frequency of alcohol consumption in the past 12 months, and the response categories were integrated into five levels to ensure consistency across sweeps (Sweeps 1, 2, 3, 4, 6, and 7). In Sweeps 8 and 9, alcohol consumption was assessed using the standardised Alcohol Use Disorders Identification Test – Consumption (AUDIT-C). For these sweeps, the original values were retained.	**W1alceverYP; W1alcmonYP; W1alcfreqYP** (wave_one_lsype_young_person_2020) **W2alceverYP; W2alcfreqYP** (wave_two_lsype_young_person_2020) **W3alceverYP; W3alcfreqYP** (wave_three_lsype_young_person_2020) **W4AlcEverYP; W4AlcFreqYP** (wave_four_lsype_young_person_2020) **W6AlcEverYP; W6AlcFreqYP** (wave_six_lsype_young_person_2020) **W7AlcEverYP; W7AlcFreqYP**	**W1alceverYP; W2alceverYP; W3alceverYP; W4AlcEverYP; W6AlcEverYP; W7AlcEverYP** 1= Yes 2= No **W1alcmonYP** 1= Yes 2= No **W1alcfreqYP; W2alcfreqYP; W3alcfreqYP; W4AlcFreqYP;** 1= most days 2= once or twice a week 3= two or three times a month 4= once a month 5= once every couple of month 6= less often **W6AlcFreqYP; W7AlcFreqYP** 1= almost every day 2= five or six days a week 3= three or four days a week

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
	auditb25; auditb32; **"AUDIT-C B (age yy)"** 0= 0 1= 1-2 drinks 2= 3-4 drinks 3= 5-6 drinks 4= 7-9 drinks 5= 10+ **auditc25; auditc32;** **"AUDIT-C C (age yy)"** 0= Never 1= less than monthly 2= monthly 3= weekly 4= daily, or almost daily		(wave_seven_lsype_young_person_2020) **W8AUDIT1; W8AUDIT2; W8AUDIT6** (ns8_2015_self_completion) **W9AUDIT1; W9AUDIT2; W9AUDIT3** (ns9_2022_main_interview)	4= once or twice a week 5= once or twice a month 6= once every couple of months 7= once or twice a year 8= not at all **W8AUDIT1; W9AUDIT1** 1= never 2= monthly or less 3= 2-4 times a month 4= 2-3 times a week 5= 4 or more times a week **W8AUDIT2; W9AUDIT2** 1= 1-2 drinks 2= 3-4 drinks 3= 5-6 drinks 4= 7-9 drinks 5= 10+ **W8AUDIT6; W9AUDIT3** 1= Never 2= less than monthly 3= monthly 4= weekly 5= daily, or almost daily

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
Drug	**drgcnbevr** **"Cannabis ever used (age 14-32y)"** 0= no 1= yes **drgcnbfst** **"First time reported ever used cannabis (age 14-32y)"** **drgcnbnw19; drgcnbnw20; drgcnbnw25; drgcnbnw32** **"Cannabis current use (age yy)"** 0= no 1= yes **drgothevr** **"Other drugs ever used (age 19-32y)"** 0= no 1= yes **drgothfst** **"First time reported ever used other drugs (age 19-32y)"** **drgothnw19; drgothnw20; drgothnw25; drgothnw32** **"Other drugs current use (age yy)"** 0= no 1= yes	We included two types of drug use in the core dataset: cannabis and other drugs. Although more specific substances were recorded after Sweep 7, we grouped all non-cannabis substances into a single "other drugs" category to enable cross-sweep comparisons and due to the small number of cases for individual drugs. For each type, three sets of variables were derived. The first indicates whether the cohort member had ever used the drug across any sweep. The second captures the sweep at which the drug was first reported, though this may not reflect the actual age of first use. It is important to note that the age coverage differs between cannabis and other drugs due to variations in the original data collection. The final set of variables, available from age 19 onwards, captures whether the drug was used in the past four weeks.	**W1canntryYP** (wave_one_lsype_young_person_2020) **W2canntryYP** (wave_two_lsype_young_person_2020) **W3canntryYP** (wave_three_lsype_young_person_2020) **W4CannTryYP** (wave_four_lsype_young_person_2020) **W6DrugYP0a; W6DrugYP0b; W6DrugOftenYP0a; W6DrugOftenYP0b** (wave_six_lsype_young_person_2020) **W7DrugYP1YP0a; W7DrugYP1YP0b; W7DrugYP2YP0a; W7DrugYP2YP0b; W7DrugOftenYP0a; W7DrugOftenYP0b; W7DrugOftenYP0c; W7DrugOftenYP0d; W7DrugOftenYP0e; W7DrugOftenYP0f; W7DrugOftenYP0g; W7DrugOftenYP0h; W7DrugOftenYP0i**	**W1canntryYP; W2canntryYP; W3canntryYP; W4CannTryYP** 1= yes 2= no **W6DrugYP0a; W6DrugYP0b; W7DrugYP1YP0a; W7DrugYP1YP0b** 1= mentioned 0= not mentioned **W6DrugOftenYP0a; W6DrugOftenYP0b** 1= not at all 2= once 3= twice 4= three or more **W7DrugYP2YP0a; W7DrugYP2YP0b** 1= mentioned 0= not mentioned **W7DrugOftenYP0a; W7DrugOftenYP0b; W7DrugOftenYP0c; W7DrugOftenYP0d; W7DrugOftenYP0e; W7DrugOftenYP0f; W7DrugOftenYP0g; W7DrugOftenYP0h; W7DrugOftenYP0i** 1= not at all 2= once 3= twice 4= three or more

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
			(wave_seven_lsype_young_person_2020) **W8DRUGYP10A:I; W8DRUGYP20A:I; W8DRUGOFTEN0A-I** (ns8_2015_self_completion) **W9DRUGYP1J; W9DRUGYP10A:I; W9DRUGYP20A:J; W9DRUGOFTEN0A:J** (ns9_2022_main_interview)	**W8DRUGYP10A:I; W9DRUGYP1J; W9DRUGYP10A:I** 1= yes 2= no **W8DRUGYP20A:I; W9DRUGYP20A:J** 1= yes 2= no **W8DRUGOFTEN0A-I; W9DRUGOFTEN0A:J** 1= yes 2= no
Bullying	**bul14; bul15; bul16; bul17; bul20; bul25** **"Bullying victimisation (age yy)"** 0= no 1= yes	The original surveys captured several types of bullying victimisation, with variations in question wording and categories across sweeps. To enable cross-sweep comparison, align with common measures in the bullying literature, and address small sample sizes for some specific types, we created binary variables indicating whether the cohort member experienced any bullying victimisation at each available sweep.	**W1bulrc** (wave_one_lsype_young_person_2020) **W2bulrc** (wave_two_lsype_young_person_2020) **W3bulrc** (wave_three_lsype_young_person_2020) **W4V1perSYP; W4ViolentYP; W4HurtYP; W4ThreatsYP; W4MadegiveYP; W4NamesYP** (wave_four_lsype_young_person_2020) **W7BullyTypeYP0a:f** (wave_seven_lsype_young_person_2020)	**W1bulrc; W2bulrc; W3bulrc** 1= yes 2= no **W4V1perSYP; W4ViolentYP; W4HurtYP; W4ThreatsYP; W4MadegiveYP; W4NamesYP** 1= yes 2= no **W7BullyTypeYP0a:f** 1= every day 2= a few times a week 3= once or twice a week 4= once every two weeks 5= once a month 6= less often than this 7= it varies 8= never

Topic	Core data variable names, labels and response categories	Derivation logic	Source variables / datasets	Source variables' response category
			W8BULLYTYPE0A:G (ns8_2015_self_completion)	**W8BULLYTYPE0A:G** 1= yes 2= no